Quality of Experience based Investigation of Interactive Video Streaming Applications

Alhad Kuwadekar and Khalid Al-Begain
Integrated Communications Research Centre
University of Glamorgan, UK
{akuwadek, kbegain}@glam.ac.uk

***Abstract* - Video call capabilities in 3G mobile handsets and higher data rate in HSPA networks are fuelling resurgence in video telephony. This enables the delivery of exciting video applications to subscribers. The addition of Interactive Voice and Video Response (IVVR) opens up a wide stream of interactive video applications. In this paper we present a comparative study on the encoding tools and the techniques used to store and stream video on demand to subscribers. The paper presents the performance results obtained on a IP Multimedia Subsystem (IMS) based Next Generation Network test-setup connected to a live operator network. These results are based on the Quality of Experience (QoE) that end user perceive. These results can help service operators to deliver content rich applications to the end users as per their expectations.**

Quality of Experience, Next Generation Networks, Video Streaming Applications

I. INTRODUCTION

The convergence of telecommunication technologies is currently driving the notion of Next Generation Networks (NGN). Most service providers are moving towards triple play, a service which enables the delivery of video, voice and broadband facility over NGN. In this context, the 3rd Generation partnership Project (3GPP) has standardised the IP Multimedia Subsystem (IMS) [1] to act as platform that will enable the provision of services over fixed and mobile IP networks. The complexity of delivering video streams across different devices and different access networks is relatively high. This is due to the different device types (HDTV, PC, mobile phone, etc.), different screen sizes (3.5 inches for a mobile phone to 103” for a plasma screen), different access technologies and varying wireless channel conditions. Most current operator video streaming implementations use a single encoding technique that is provided to all users irrespective of the access technology or the channel conditions they receive.

The outline of this paper is as follows: section 2 gives a brief overview of the important technologies considered in our approach and the quality of experience aspect associated with them. Section 3 describes the IMS Testbed architecture at the University of Glamorgan, UK. Section 4 describes various application scenarios and end users quality of experience on a standard IMS architecture. The last section presents conclusion and future work.

II. RELATED WORK

One of the primary problems with multimedia streaming in IMS is adaptability to device capability, codec supported and network conditions. Camarillo et all [2] have described various scenarios for transcoding in RFC 4117. They have discussed about invoking a third party transcoding service for an end user to end user communication. For real time communication such as a voice call and video call, such a service could add significant delay and reduce the overall user experience. Another research performed by Park et all [3] presents a feasible solution of dynamic session control for scalable video coding over IMS. However, this solution works at the session layer and does not describe techniques to adapt to end user’s client device. Similarly a multimedia adaptation plug-in was proposed in the past which suggests the addition of new data fields in the SDP [4]. Adding new fields to the SDP will not make it compliant with all service providers that would most probably follow the specification as defined in RFC 4566 [5]. A similar approach has been followed by the authors in GenXfone [6] wherein the camera capture quality is adjusted in accordance with changing network conditions and device capabilities. This would enhance the overall end user experience for a video call only as there is no mention of the operator side implementation for enhanced video streaming experience. This paper proposes an operator side plug-in that works on the standardized SIP methods and adapts to device capabilities and network condition changes.

The next step is to perform testing. For years together communication engineers have focused on the Quality of Service (QoS) paradigm and what technology provides the best QoS [7]. Most end users do not care about what technology goes into a product. They care about the best possible experience that they can get from that product [8]. This test is called as Quality of Experience (QoE) test. Measuring QoE needs a huge human resource and is quite expensive. Hence a QoS/QoE measurement implementation in NGN was proposed by Satoshi et all [9]. This architecture suggests an automated system to measure the end users QoE to be implemented on mobile nodes. This is not an ideal implementation as different humans perceive things differently. Based on an automated system it is difficult to guess the acceptability rate of end users. So far none of the research works have taken into consideration the end users QoE for varying access technologies, encoding bit rates and changing device capabilities. This paper also presents test results obtained with end users based on their QoE across varying network conditions and device capabilities on a real operator's network and a standard IMS implementation.

The testbed for video streaming architecture is described in section 3. Section 4 describes different scenarios and the results obtained from testing carried out on the testbed.

III. IMS TESTBED ARCHITECTURE

IMS is a standardised architecture that employs voice and video over IP technology as described in the 3GPP2 standards. Figure 1 describes the IMS testbed architecture available at the University of Glamorgan used for testing video streaming and IVVR applications. For simplicity, the testbed has been divided into three parts: the service layer, the control layer and the access network.

The IMS applications are hosted in the service layer. This layer consists of the Ubiquity Avaya SIP Application Server (A/S) which executes IMS applications and services.

The control layer consists of nodes that perform call establishment, management and release. The heart of the IMS network consists of a Call Session Control Function (CSCF). It manages SIP sessions and coordinates with other network entities for session control, service control and resource allocation. The Home Subscriber Server (HSS) acts as a central repository of user related information. The core components of this IMS network are developed by Orange – France Telecom. When the CSCF or AS requires media capabilities, it routes the call to the Voxpilot Media Server which acts as a Media Resource Function (MRF). Finally the NMS Vision Server acts as an interface between the circuit switched network and the packet switched network.

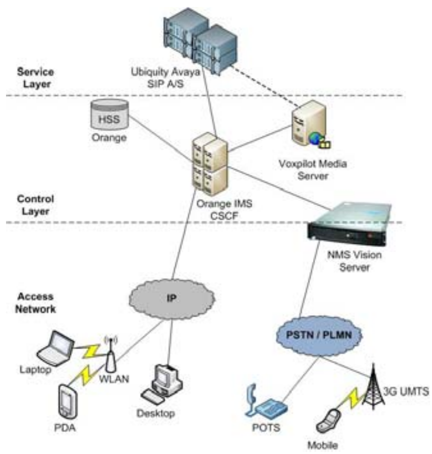

Figure 1: IMS Testbed Architecture

Though the NMS Vision Server can also be used as a MRF as it runs Voxpilot Media Server at its heart, a separate Media Server is used to reduce the amount traffic flowing to the Vision Server.

The access network in the testbed consists of wired LAN, wireless LAN and an indoor Nokia 3G Node B antenna supported by Orange. The 3G Node B antenna and the NMS Vision Server are connected to British Telecommunications ISDN 30 network. The next section describes the scenario and the call flow used for deployment and testing streaming IVVR application on the above mentioned testbed.

IV. SCENARIO, CALL FLOW AND TEST RESULTS

A. Test Scenario:

The IMS application scenario described in this paper is the provision of video on demand service to a subscriber. The subscriber can select a video of his choice based on an IVVR application that displays the selection number of the movie along with its trailer. The selection of any one of these videos is based on the SIP INFO DTMF tone collection as described in RFC2833. Once the user presses the appropriate number, the movie associated with it is streamed in 3gpp format. Wireshark was used to measure the time interval after the user presses the keypad and the new video stream begins to playback. The IVVR application along with the media files is stored on the Voxpilot Media Server. The SIP A/S acts as a Back to Back User Agent (B2BUA) to connect the IMS User Agent (UA) and the Media Server. Throughout the experiment, Helix Mobile Producer and FFMPEG encoders were used. A comparative study of both these encoders from end user point of view has been described in the later section of this paper.

B. Subjective Measurement:

Quality of Experience (QoE) is a subjective measure of a customer's experience with the system. QoE describes the system-level activities focusing on the joint optimization of experienced multimedia quality in wireless multimedia systems. The quality of an end user's experience is the true litmus test of a proper video and voice deployment. QoE is concerned with how the video is perceived by a viewer and designates his or her opinion on a particular video sequence. A video sequence is presented to the viewer, encoded in variety of frame rates, bit rates, using different access technologies, using different encoders and on different end point devices. The opinion of each viewer is collected and the best encoding technique is selected. To make sure that the best opinion of each user is obtained, a questionnaire consisting of ten questions was provided to them. These questions were concerned with factors like video quality, audio quality, start-up time, DTMF tone collection time and overall acceptability rate. The questionnaire was filled by ten users. The overall acceptability rate was calculated in as follows:

$$\text{Acceptability Rate} = \frac{\sum_{n=1}^{10} \text{Accepted by no. of users}}{\sum_{n=1}^{10} \text{Total no. of users}}$$

Similarly users were asked their opinion on the scenario that comprised of pressing a number and the new stream begins to playback. The users were asked to mark their opinion on a rating sheet running from 1 to 5. Based on the user ratings, the Mean Opinion Score (MOS) was calculated. The MOS is the arithmetic mean of all individual scores and ranges from 1 (worst) to 5 (Best). The following rating scheme was defined:

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not Annoying
3	Fair	Slightly Annoying
2	Poor	Annoying
1	Bad	Very Annoying

Table 1: Ratings for MOS Calculation

The acceptability rate for different access technologies is described in different scenarios below.

C. Scenario 1:

The first test was carried out using X-Lite as an IMS enabled video client over the LAN network. X-Lite client was installed on a desktop computer which was connected to the IMS network through LAN connection.

Figure 2 describes the simplified signalling for video enabled media streaming.

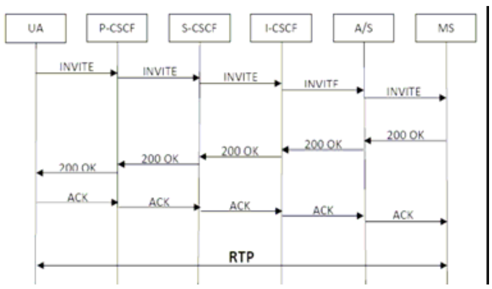

Figure 2: Simplified IMS Signalling

The encoding of files was performed using both Helix Mobile Producer and FFMPEG. The frame rate was set to 15 frames per second (fps) and encoded in H.263 format as it is supported by both X-Lite and Voxpilot Media Server. The frame size was set to 176 pixels by 144 pixels. Out of the total encoding bit rate 12.2 kbps was used for audio encoding. For example if the total encoding bit rate is 384 kbps, the encoding audio bit rate is set to

12.2 kbps and the video bit rate is set to 371.8 kbps. The encoding bit rate of the video was varied and shown to the viewers. Initially the video was encoded at a bit rate of 384 kbps. After the encoding was completed, each video was played back in Quicktime player to detect any playback errors. This video was then streamed to ten viewers over their desktop machine. The acceptability rate was zero as the picture started displaying video errors on X-Lite as shown in Figure 3. Having said that, the audio stream was error free from user point of view.

Encoded Image Received Image

Figure 3: Video Errors

This meant that receiving bit rate for X-Lite had to be reduced. So the bit rate was reduced by 32 kb and again streamed to the users. This process was carried out until the encoding bit rate was set to 32 kbps. An acceptability rate graph is shown below.

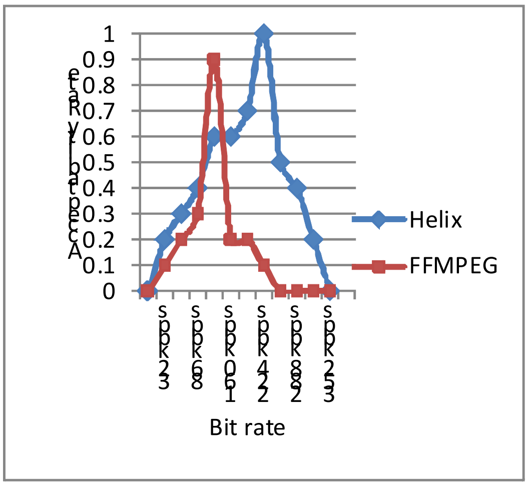

Figure 4: Acceptability Rate over LAN

With Helix Mobile Producer, the overall acceptability rate was the highest at 256 kbps. With FFMPEG, the overall acceptability rate was obtained at 160 kbps. It declined on the higher encoding bit rate due to video errors leading to poor video quality. On the lower encoding bit rate side, the acceptability rate reduced due to low bit rate leading to pixelation and poor video quality.

Figure 5 shows the graph for DTMF Tone Collection and Playback start Time. This is the time interval after the user presses the button and the playback starts. The single user graph signifies that a single user was accessing the IMS network at a given point in time. Since a single user was on the network, there was no traffic for him. In such a situation the average time after the user presses the button and the new stream to start playing was 2.076 seconds.

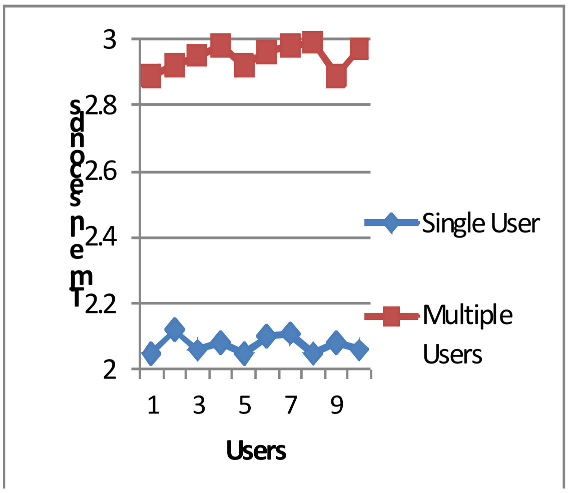

Figure 5: DTMF Tone Collection and Playback start Time

The next decision was to create some traffic on the network. The best way to do this was to ask multiple users to access the same IVVR application at the same time and select a video based on their preferences. The average time to playback the new stream was 2.945 seconds. Since the test was being carried out on the same network, this timeframe was very high from Quality of Service point of view. This would be higher when a user accesses a service from a different network. But from user point of view the experience was acceptable.

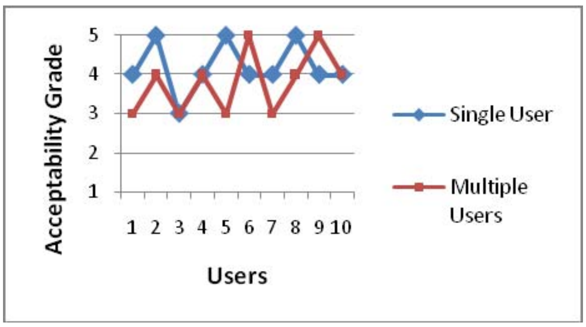

Figure 6: DTMF Tone Collection Acceptability Grade

Figure 6 shows the DTMF Tone Collection Acceptability Grade. Each viewer was asked to rate the time taken for his/her selected video to start as mentioned in Table 1. Based on the feedback of all the users the MOS was calculated. In case of single user the MOS was 4.2, while it was 3.8 for multiple users which meant it was acceptable.

D. Scenario 2:

In this scenario, the desktop computer was replaced by a Windows Mobile 6 device with Kapanga softphone installed on it. This was then connected to the IMS network through WLAN.

The previously encoded files were used for this test. In the previous scenario, the 256 kbps encoding rate with Helix Mobile Producer was given the highest rating. Hence this test was started using the same video. Unfortunately only two viewers thought that the quality was acceptable. This rate was quite low as compared to the ratings in the last scenario.

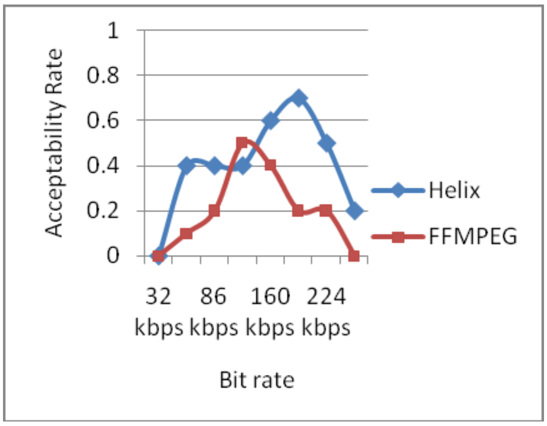

Figure 7: Acceptability Rate over WLAN

Therefore the encoding bit rate was dropped by 32 kbps. Dropping the encoding bit rate by 32 kbps made the acceptability rate to improve dramatically. The encoding bit rate was dropped further. As seen in Figure 7, the highest acceptability rate was 0.7 at 192 kbps with Helix Mobile Producer. Similarly the test was carried out for FFMPEG. The best acceptance rate of 0.5 was obtained at 128 kbps.

Since packet capture could not be done windows mobile, the time duration for the DTMF tone collection could not be calculated. The best test to understand the quality of the DTMF tone collection was to ask the viewer to rate the time frame for his/her selected video to start.

Figure 8 shows the DTMF Tone Collection Acceptability Grade. Based on the Table 1, the viewers were asked to rate their experience.

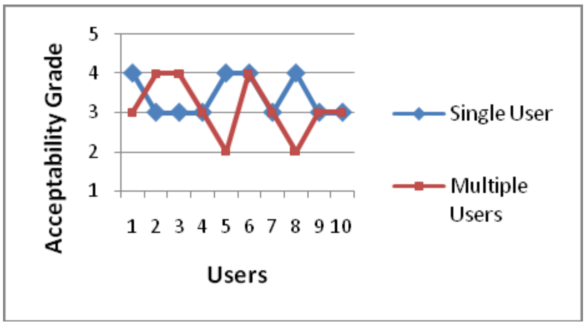

Figure 8: DTMF Tone Collection Acceptability Grade

Based on the results the MOS was calculated. The MOS was 3.4 for a single user while it was 3.1 for multiple users which meant it was slightly annoying but acceptable.

One of the primary reasons for the test to yield such poor results was that this scenario came straight after scenario. After scenario 1, the viewer expectations were high and they expected the same quality for Scenario 2. Since the test was carried out on WLAN on the same network, the DTMF collection shouldn't have been very slow. To investigate this problem, a WLAN enabled laptop with Kapanga PC softphone installed on it. The results obtained were similar to Scenario 1. This led to the conclusion that Kapanga on mobile runs slower than on desktop machines.

The next scenario was to move the access technology from WLAN to 3G.

E. Scenario 3:

In this scenario, the access technology was changed from WLAN to 3G. The setup was similar to Scenario 2 and Kapanga Mobile softphone connected through the 3G network was used for this test. Since 3G works at a lower data rate than WLAN, it was decided to start the test at a lower encoding data rate. The starting data rate was chosen to be 128 kbps at 15 fps. At the suggested frame rate and bit rate, the video at seldom displayed video errors as shown in Figure 4. On reducing the bit rate dramatically, the audio stream was streaming perfectly but the video would stall for a while and then start streaming again with video errors. Due to these streaming problems, the acceptability rate was very low.

Hence the frame rate was reduced to 12 fps as shown in Figure 9.

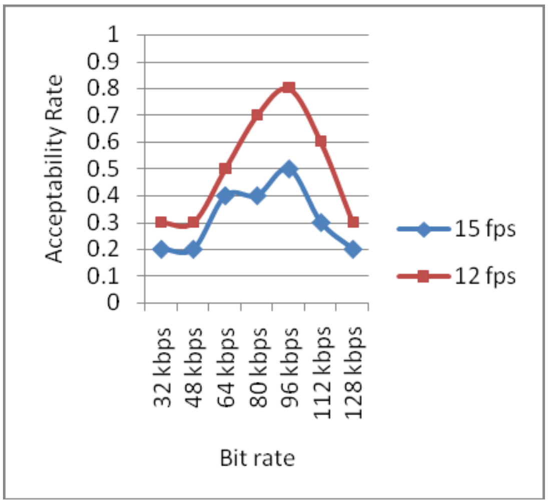

Figure 9: Acceptability Rate over 3G with Helix Mobile Producer

At 12 fps, the first testing file was encoded at 128 kbps. The acceptability rate was quite low due to video errors. Hence the bit rate was eventually reduced by 16 kbps. At 112 kbps, the acceptability rate improved. The bit rate was further dropped. At 96 kbps, the acceptability rate was 0.7 which was the highest. The same test was carried out using FFMPEG encoder. The test was started at 128 kbps encoding rate and 15fps. At this bit rate and frame rate, the video was not clearly visible and would stall for some time. The next frame would come in after some video errors. While on the other hand the audio stream was flowing in smoothly. Since viewers are accustomed to Digital TV, YouTube and BBC iPlayer, this kind of streaming was totally unacceptable to them as shown in Figure 9. The bit rate was further reduced by 16 kbps to see if the viewers changed their opinion. But there was no drastic change in the acceptance rate as shown in Figure 10. The best acceptance rate was 0.2 based at 80 kbps and 64 kbps encoding bit rate. Since this acceptance rate was quite low, the frame rate was dropped to 12 fps. The acceptance rate improved at 64 kbps.

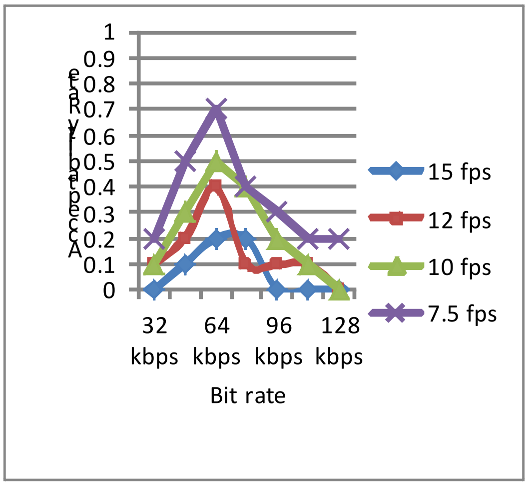

Figure 10: Acceptability Rate over 3G with FFMPEG

The frame rate was further dropped to 10 fps. A slight improvement in the acceptance rate can be seen in Figure 10. The frame rate was further reduced to 7.5 fps. At this frame rate, from the viewer's perspective the video clarity was much better than the previous frame rates but the motion picture was slower than the way it is perceived at 15 fps. The acceptability rate was the highest i.e. 0.7 at this frame rate and 64 kbps encoding bit rate. The reason for this was that users felt that the video and the audio were clear and did not mind that the picture was slower.

The next step was to ask the viewers about their opinion on the DTMF tone collection. On average viewers were not so happy on the time taken by the DTMF tone collection over 3G. They felt it was too slow.

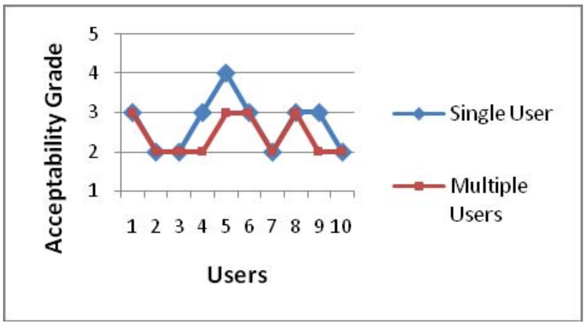

Figure 11: DTMF Tone Collection Acceptability Grade

Figure 11 shows the DTMF Tone Collection Acceptability Grade. Based on the Table 1, the viewers were asked to rate their experience. Based on the results the MOS was calculated. The MOS was 2.7 for a single user while it was 2.4 for multiple users which meant it was annoying.

The next question was how would viewers react to a video call over the circuit switched network? Scenario 4 describes this issue.

F. Scenario 4:

This scenario deals with video calls made over the circuit switched network. The Media Gateway (MGW) was introduced to bridge the gap between the circuit switched network and the packet switched network as shown in Figure 12. The main functionality of the MGW was to perform circuit switched signalling with the BT ISDN line at one end and SIP signalling at the other. For this test, Kapanga softphone for mobile was replaced with a video call enabled windows mobile handset over circuit switched network. The viewer was asked to make a video call to the BT ISDN number. The MGW was the terminating point for the BT ISDN. The MGW would then forward this call to the IMS testbed which would then stream the IVVR application.

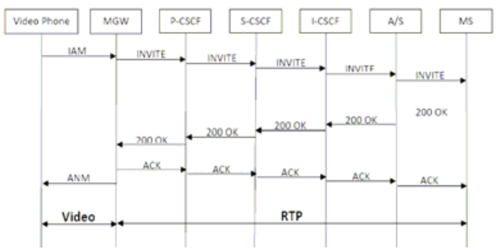

Figure 12: Simplified ISDN to IMS Signalling

A diagrammatic representation of the call flow model is shown in Figure 13 as defined in the RFC 3666, SIP to PSTN Call Flow.

Since the bandwidth of ISDN is 64 kbps per channel, the test was started with a file encoded at 64 kbps and 12 fps using Helix Mobile Producer. At this frame rate and bit rate, the streaming video showed video errors as shown in Figure 3. But still two users felt that it was an acceptable video as shown in Figure 13. The reason for video errors was that there was not enough bandwidth left for multiplexing the audio and video. Hence the encoding bit rate was dropped further to 56 kbps. The acceptability rate was 0.6 at 56 kbps encoding rate and remained the same at 48 kbps as shown in Figure 13.

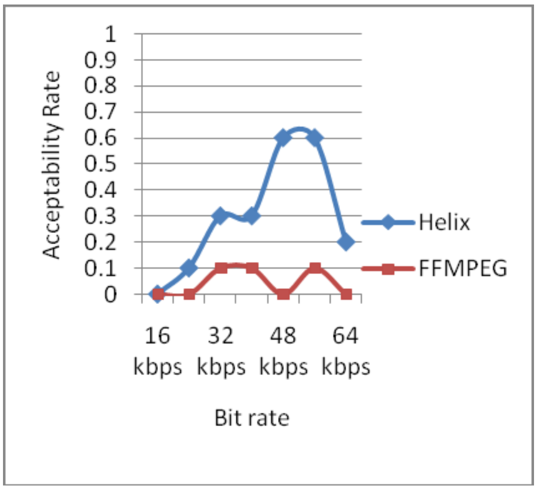

Figure 13: Acceptability Rate over ISDN with Helix Mobile Producer

At 64 kbps and 12 fps, with FFMPEG, the video errors were quite dominant. Secondly there wasn't enough bandwidth left for the multiplexing channel to work. Hence the acceptance rate was null. On reducing the bit rate, video errors could still be seen but were seldom.

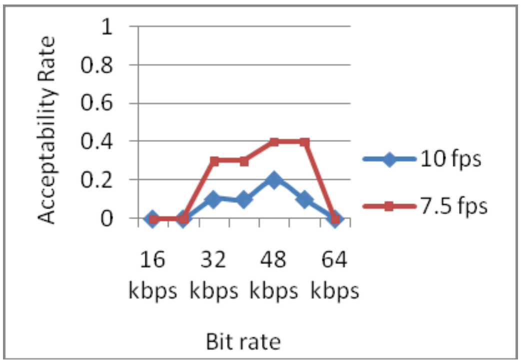

Figure 14: Acceptability Rate over ISDN with Helix Mobile Producer

Hence the acceptability rate improved slightly but not dramatically as shown in Figure 14. Hence a decision was taken to reduce the frame rate. The frame rate was reduced to 10 fps. At this frame rate the acceptance rate improved slightly. Since with 3G connection in scenario 3, FFMPEG provided excellent results, the frame rate in this case was dropped to 7.5 fps. The acceptability rate shot to 0.4 at an encoding rate of 56 kbps and 48 kbps making it the best with FFMPEG over ISDN line.

The next step was to ask the viewers about their opinion on the DTMF tone collection.

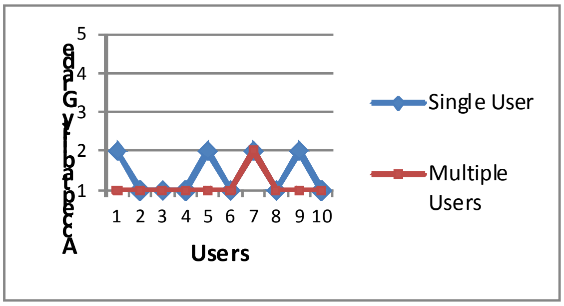

Figure 15: DTMF Tone Collection Acceptability Grade

Figure 15 shows the DTMF Tone Collection Acceptability Grade. Based on the Table 1, the viewers were asked to rate their experience. Based on the results the MOS was calculated. The MOS was 1.4 for a single user while it was 1.1 for multiple users which meant it was extremely annoying.

The above mentioned scenarios show that different encoding rates were needed for various access technologies and device capabilities. The encoding bit rate was also dependent on the encoders used. The best encoding technique was chosen based users Quality of Experience (QoE).

V. CONCLUSION

The aim of this work was to understand, with a series of experiment, the behaviour of video streaming across various devices and changing network conditions. Through passive measurements, we characterised the behaviour of video streaming and evaluated the users QoE. Based on the actual measurement results, the main conclusions are: (1) Video streaming over IMS can provide good quality video if streamed as per device capabilities and changing network conditions; (2) The choice of the right encoding tool is extremely important to get the best video quality. (3) Users do not experience much difference in the video quality if the encoding bit rate is lowered slightly. A noticeable change can be seen only if the encoding rate is severely changed and the frame rate is changed; (4) Streaming over ISDN suffers extensively as far video quality and DTMF tone collection is concerned.

REFERENCES

[1] 3GPP, TS 23.228. IP Multimedia Subsystem; (stage 2). www.3gpp.org
[2] Camarillo G., Burger E., Schulzrinne H., van Wijk A., Transcoding Service Invocation in the Session Initiaition protocol using third party call control, RFC 4117, June 2005.
[3] Park S., Lee S., Dynamic Session Control for Scalable Video Coding over IMS, IEEE Vehicular Technology Conference, May 2008, pp2779 – 2783.
[4] Ozcelebi, T. Radovanovic, I., Multimedia Adaptation with SIP Resource Availability Signalling in IMS Networks, IEEE Convergence Information Technology Conference, Nov 2007, pp1714 – 1719.
[5] Handley M., Jacobson V., Perkins C., Session Description Protocol, RFC 4566, July 2006.
[6] Kuwadekar A., Balakrishna C., Al-Begain K., GenXfone – Design and Implementation of Next Generation Ubiquitous SIP client, IEEE NGMAST'08 Conference, Sep 2008, pp118 – 124.
[7] Kim J., Um Tai., Lee Byung, Heterogeneous Networks and Terminal-Aware QoS/QoE Guaranteed Mobile IPTV Service, IEEE Communications Magazine, May 2008, pp110 – 117.
[8] Jain R., Quality of Experience, IEEE Computer Society Magazine, 2004, pp95 - 96.
[9] Uemura S., Fukumoto N., Nakamura H., QoS/QoE measurement system implemented on cellular phone for NGN, IEEE Consumer Communication and Networking Conference, Jan 2008, pp117 – 121.
[10] Dialogic IP Media Server, http://www.dialogic.com/products/ip_enabled/ip_media_server/IP_media_server.htm, website accessed on 5th October 2008.
[11] Voxpilot Media Server, http://www.voxpilot.com, website accessed on 5th October 2008.
[12] NMS Communications, http://www.nmscommunications.com, website accessed on 5th October 2008.