Breaking the News on Mobile TV:
User requirements of a popular mobile content

Hendrik O. Knoche[*], M. Angela Sasse

Dept. of Computer Science, University College London, Gower St, London WC1E 6BT, UK

ABSTRACT

This paper presents the results from three lab-based studies that investigated different ways of delivering Mobile TV News by measuring user responses to different encoding bitrates, image resolutions and text quality. All studies were carried out with participants watching News content on mobile devices, with a total of 216 participants rating the acceptability of the viewing experience. Study 1 compared the acceptability of a 15-second video clip at different video and audio encoding bit rates on a 3G phone at a resolution of 176x144 and an iPAQ PDA (240x180). Study 2 measured the acceptability of video quality of full feature news clips of 2.5 minutes which were recorded from broadcast TV, encoded at resolutions ranging from 120x90 to 240x180, and combined with different encoding bit rates and audio qualities presented on an iPAQ. Study 3 improved the legibility of the text included in the video simulating a separate text delivery.

The acceptability of News' video quality was greatly reduced at a resolution of 120x90. The legibility of text was a decisive factor in the participants' assessment of the video quality. Resolutions of 168x126 and higher were substantially more acceptable when they were accompanied by optimized high quality text compared to proportionally scaled inline text. When accompanied by high quality text TV news clips were acceptable to the vast majority of participants at resolutions as small as 168x126 for video encoding bitrates of 160kbps and higher.

Service designers and operators can apply this knowledge to design a cost-effective mobile TV experience.

Keywords: Mobile TV, mobile devices, mobile multimedia, video quality, audio quality, image resolution, image size, text in video

1. INTRODUCTION

Recent interest in mobile TV services illustrates the potential demand for services that deliver TV without geographical or temporal constraints.

A variety of services are currently offered, aiming to provide a TV-like experience. Some services have started relaying standard TV channels to mobile phones. Video iPods and other devices allow the downloading of video content for consumption on the move. New broadcast standards like DVB-H[1] and SDMB[2] are employing power saving schemes to maximize the time that users can get exposure to content. To realize that potential, content providers and network operators need to deliver economically viable multimedia services at quality levels that meet and satisfy their customers' expectations and demands. But the overall experience of consuming multimedia on the move is currently not well understood.

What are decisive dimensions that determine the Quality of Experience (QoE) of mobile multimedia services? The quality of experience of the content depends on the perceived visual, audio and textual quality of the content consumed, and the interaction through which the user has to go to access it (e.g. the delay between selecting content and start of play)[3]. In this paper, we focus on the former. Resolution and viewing ratio (picture height/viewing distance) are two of the central factors that influence video quality perception. They are central to a discussion of mobile devices, which have small screens and resolutions in order to be energy efficient and portable and are operated within arm's length. The visual resolution of the content affects the required bandwidth for transmission and the amount of visual detail that can be displayed, and consequently limits the amount of content that can be presented through a mobile TV service.

News was identified by users in focus groups and surveys as the most interesting content type for mobile consumption[4]. The typical length of a News bulletin fits well with observed use of mobile TV in bursts of less than 10 minutes[5]. Because is not clear whether the audience of mobile TV will support the cost involved in editing content especially for mobile TV, the simplest and cheapest solution would be to deliver TV material without additional editing. However, text

---

[*] h.knoche@cs.ucl.ac.uk; phone +44 207 679 3642; fax +44 207 387 1397

is an important tool in the presentation of news, and text sizes that work on standard television sets might pose a problem for smaller resolutions and screen sizes.

To ensure an optimal News service the following questions need to be addressed:

1. How do different encoding bitrates of both audio and video affect the perceived quality on mobile devices?
2. Is it possible to repurpose standard TV news without reediting and down to what resolutions?
3. How big is the influence of text?

In Sec. 2 we provide more background on the impact of size and resolution on perceived video quality. The three following sections present the three studies that address the above three questions. Section 6 summarizes the findings from all three studies.

2. BACKGROUND

From an economic point of view, videos with lower resolutions require less bandwidth and hence reduce the price of delivery. However, it is also clear that image resolution and the image size of a viewing device cannot be reduced indefinitely, as important detail will be lost.

The size of the display in the viewer's visual field depends on both the size of the screen and the distance between viewer and the screen. The viewing ratio (VR) is defined as the viewing distance divided by the picture height (H). When planning the production of their programs, television producers consider the typical resolutions of TV sets, their sizes and the seating distance at which the audience is located[6]. In a mobile context, however, some factors restrict the range of possible combinations. Mobile devices are typically operated at 'arm's length' and continued viewing at distances closer than the resting point of vergence – approx. 89cm, with a 30º downward gaze – can contribute to eyestrain[7].

Previous research has examined the impact of increasing the image size in the viewer's visual field by means of large physical displays or projection areas. These studies have typically compared very large size screens (e.g. 46") to standard sized TV screens (15-20")[8,9]. The results show that larger image sizes are more arousing than smaller ones, better remembered, and better liked. Other studies also show that users generally prefer bigger image sizes – ideally depicting people and objects up to life-size[10]. With reduced screen sizes, screen clutter could become an issue for mobile users at a much earlier point than it does at traditional TV resolutions, where screen clutter has been shown to impede attention and comprehension[11]. The general message from these studies is '*the bigger the better*'. This clearly presents a challenge to mobile TV, which faces a tradeoff between the screen size and the portability of the device. These concerns were noted in a study involving focus groups that assessed the potential uptake of mobile TV services[4]. Users wanted as large a screen as possible for viewing, but they did not want their phones to be too big. Some stated that they would opt for a dedicated mobile TV device to circumvent this dilemma. It was also unclear whether users would want higher arousal and immersion in a mobile context because some feared the increased risk of errors and accidents.

In one of the few studies that specifically examined smaller screens, Reeves et al. found no difference in arousal and attention between users watching 2" and 13" screens, although arousal and attention were larger with a very large screen (56")[12]. Other studies have even shown that smaller image resolutions can improve task performance. For example, Horn showed that lie detection was better with a small (53x40) than a medium (106x80) video image resolution[13]. In another study, however, smaller video resolutions (160x120) had no effect on task performance but did decrease satisfaction when compared to 320x240 image resolutions[14]. In a study by Barber et al., a reduction in image resolution (from 256x256 to 128x128) at constant image size led to a loss in accuracy of emotion detection especially in a full body view[15].

Video quality cannot be completely isolated from the accompanying audio component. As a byproduct in a study on TV viewing experience, Neuman et al. discovered that the perceived video quality was improved by better audio[16]. However, this only held true for one of the three content types. A similar study by Beerends and Caluwe, who used a 29cm monitor, found that video quality was rated slightly higher when accompanied by CD quality audio than when accompanied by no audio[17]. The two lower video quality levels (with bandwidth limited to 0.15 MHz and 0.025 MHz) were judged worse, however, when they were presented with audio than without audio. Furthermore, the legibility and quality of text contained in News clips might affect the perceived quality of the overall video picture. To the best of our knowledge no study has ever looked at the impact of text legibility and quality on perceived video quality.

In Study 1 we look at the acceptability of different audio and video encoding bitrates when maximizing the visual screen estate on two different mobile devices. It was designed to identify encoding bitrates of News that would constitute a good basis for QoE in mobile TV. Study 2 examines the effect of reduced resolution and possible interactions with video and audio encoding bitrates. In Study 3 we investigate the contribution of legible high quality text to the overall perceived quality of News footage. Together these studies should help us understand how all media involved in News might influence its experience in mobile TV.

3. STUDY 1

This study was designed to research encoding bitrates that would deliver acceptable video quality to mobile users watching news on two mobile devices with different resolutions. The aim of this study was to identify bandwidth requirements to deliver an acceptable service on the two different devices when all available screen estate was used for the presentation. We also wished to find out which audio encoding bitrates would be acceptable to mobile users.

3.1. Equipment

The experiments were carried out on two devices:

1. iPAQ 2210 PDA, with an additional 1GB CF storage card, and a pre-installed windows media player, a 3.5" (55x73mm) TFT liquid crystal display, 64K colors, (240x320 pixels)

2. NEC e616 3G phone, with 19MB internal memory, capable of decoding/playing 3GPP MPEG-4 video, a 2.2" (35x43mm) TFT liquid crystal display, 64K colors, (240x176 pixels)

3.2. Content and Encoding

Test material used for quality evaluation is usually selected from a video or audio test set. For example, the video quality experts group (VQEG) uses a test set of 20 8-second clips[18] to represent a range of different types of motions, content and camera position. While such test sets are suitable for comparing technical performance differences between codecs, they are less useful in evaluating the perceived quality of service as the clips do not have audio and are not necessarily representative of the experience users would have with mobile TV. Realistic clip lengths were used in Study 2.

This study used a 15 second clip from an online CBS news bulletin on Iraq, similar in length to those used in the VQEG test set. This clip used big Latin text for headlines apart from the ticker and other small text all of which was in Arabic.

The material was encoded for display on both devices. The encoding bit-rates ranged from 28Kbps to 256 Kbps depending on the device, as presented in Table 1 and Table 2.

Table 1: Parameters for 3G phone (176x144)

Total bitrate	Encoding bitrate: Video (PV MP4)	fps	Encoding bitrate: Audio (GSM AMR)
28 Kbps	16 Kbps	6	6.7 Kbps
64 Kbps	45.4 Kbps	10	12.2 Kbps
128 Kbps	111 Kbps	15	12.2 Kbps
256 Kbps	243 Kbps	25	12.2 Kbps

Table 2: Parameters for iPAQ (320x240)

Total bitrate	Encoding bitrate video (WMV v8)	fps	Encoding bitrate audio (WMA v9)
64 Kbps	52 Kbps	10	12 Kbps
128 Kbps	112 Kbps	15	16 Kbps
256 Kbps	224 Kbps	25	32 Kbps
350 Kbps	285 Kbps	25	64 Kbps
512 Kbps	448 Kbps	25	64 Kbps

For the 3G phone, clips were encoded using a trial version of the 3GPP MPEG-4 encoder (*PV Author*) available online from Packet Video[19]. The iPAQ clips were encoded with *MS Media Encoder* in Microsoft Video and Audio format.

3.3. Participants and Procedure

Twenty paid participants were tested viewing the described material. There were equal numbers of men and women and the mean age of participants was 26. The participants were told that we were investigating acceptable quality levels for mobile TV.

On the iPAQ, the video the clips were arranged in a play list in the Windows Media Player. On the phone each clip was selected separately from a folder. The participants watched all clips first on one device and then on the other. The order of devices on which the participants watched the videos first was randomized. Each clip was shown to the participants at all of the different encoding bitrates starting with the highest encoding bitrate and descending to the lowest encoding bitrate. The participants alternately watched a clip and then rated the acceptability of the video and the audio quality on a rating sheet. These were binary decisions[20] as participants had to tick either acceptable or unacceptable as a reply.

3.4. Results

The acceptable/unacceptable responses on the video and audio quality from the participants were averaged for each encoding bit rate, and are presented for both devices in Figure 1 and Figure 2, respectively.

The video quality of News presented on the iPAQ was acceptable to all participants when clips were encoded at 285kbps and higher. On both devices, News quality was acceptable to the vast majority of participants at encoding bit rates above 100kbps.
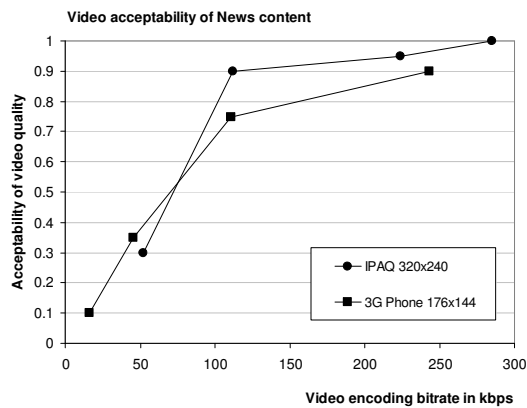

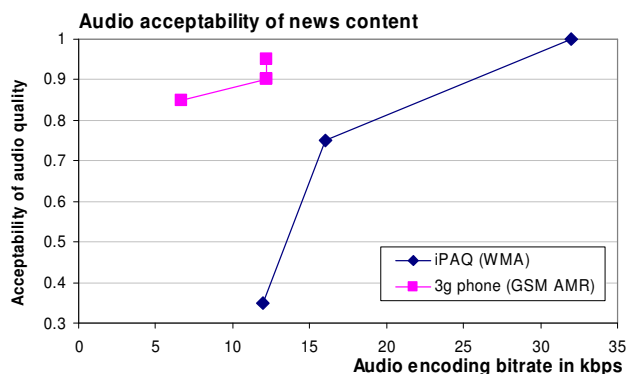**Figure 1: News video quality acceptability**

Figure 2: News audio quality acceptability

However, we did not find a statistical difference between acceptability at any comparable bitrates: [52 vs. 45; Z = -0.33, ns: 112 vs. 111; Z=-1.11, ns: 224 vs. 243; Z=-0.57, ns]. Thus, at any particular bitrate there was no evidence that the iPAQ has a quality advantage because of the larger screen and higher resolution, and there was no evidence that the phone has a quality advantage because of a higher bitrate per pixel ratio.

The audio part of the news clips on the iPAQ was acceptable to all participants when the encoding bitrate was 32kbps or higher. Below 16kbps audio encoding bitrate the acceptability dropped off sharply. Only 35% of the participant found the audio acceptable when encoded at 12kbps with WMA v9 on an iPAQ. In comparison, audio on the 3G phone was acceptable to 90% of the participants when encoded at 12.2kbps with 64kbps and 128 kbps video. When played alongside video encoded at 243kbps the same audio quality on the 3G phone was acceptable to 95% of the participants.

3.5. Discussion

The results from Study 1 indicated that:
1. A video encoding bitrate of at least 100kbps was required to deliver an acceptable News service to the majority of the audience.
2. Screen resolution and size had little effect on the boundaries of on acceptable service.

The substantial differences in audio acceptability between the two devices and the codecs used could be due to a range of factors, e.g. codec efficiency. The quality expectations on the phone might be much lower if participants took the typical low voice quality as a baseline.

However, there are a number of limitations to this study:
1. The sample was small, so it is possible that the tests were not sensitive to small differences in service acceptability.
2. Even though we showed the videos on both devices at the same frame rates for comparable encoding bitrates, the comparison was based on different video and audio codecs, and the devices had different hardware implementations (such as screen brightness and contrast).
3. The News clips used were much shorter than a standard News broadcast, and were also repeated many times in the test. It is not clear that this procedure was representative of the experience users would have from watching an actual News bulletin on a mobile device.
4. The clips only had text sizes that were legible at the tested resolutions.

4. STUDY 2

The second study was designed to look more closely at the effects of reduced image resolution on video quality acceptability. The study addresses the limitations of Study 1 by testing a much larger sample, controlling for hardware and codec differences, and using representative News material. An additional factor examined was the effect of varying

audio quality. News content relies heavily on the intelligibility of spoken words, and some loss of visual detail might be compensated for by having high audio quality. We decided to include two audio encoding bitrates (32 and 16kbps) that had been acceptable to 100% and 75% of the participants in the previous study.

4.1. Equipment

The experiments were carried out on the same iPAQs that were used for Study 1. The content was stored on a 512MB SD card. The iPAQ was equipped with a set of Sony MDR-Q66LW headphones to deliver the audio. A customized application was programmed in C# using the Odyssey CFCOM middleware to integrate the Windows Media Player for Windows CE. It presented the clips along with a volume control and two response buttons to indicate acceptable and unacceptable quality as depicted in Figure 3.

Content and Encoding

For current mobile TV services, there is usually an additional editing process to prepare the material for mobile consumption. This involves removing certain shots that would not render or compress well for a mobile device. Editing processes take time (which means access to topical content such as news is delayed) and are expensive; thus, many service providers favor immediate re-use of TV material. For the purposes of this study, we investigated the acceptability of directly recorded TV material without any special editing steps. Clips of this type have been successfully used to examine quality tradeoffs for football coverage on mobile TV[20].

Another consideration was the length of the test clips. Mobile TV viewing will typically be considerably longer than 8-15 seconds, and composed of a mixture of different foreground and background motion of the objects depicted in a scene, content and camera shot types. Whereas the clips used in Study 1 were 15 seconds long we introduced a more realistic setup in Study 2 by using clips that lasted two minutes and 20 seconds each.

We prepared the video clips as follows: We recorded the News footage from DVB-T TV (BBC24 news). All extracted clips were chosen such that after 2:20 minutes (or shortly thereafter) a story line would end. In addition, the news material chosen for this study was recorded from TV and included a variety of text information in the form of news ticker, clock, logos, graphs and captions. We cropped the original 16:9 footage to obtain clips with an aspect ratio of 4:3. We used VirtualDub to segment these source clips into seven 20-second long clips at the different resolutions at a nominal 12.5fps. These segments were encoded with Windows Media Encoder (WME) by means of the Microsoft Windows Media Video V8 codec with the different bitrates for the different segments as shown in Table 3. Each group of seven WMV segment files was then converted and concatenated to one AVI file with TMPGEnc Express. Finally, these files were encoded with WME again to alter the audio encoding to either 32 or 16kpbs using the Windows Media Audio v9 codec. The video was encoded at a higher bitrate than the maximum of the first WME encoding in order to prevent significant alterations in the video quality in any of the segments.

It should be noted that the setting of the frames per seconds as shown in Table 3 only represents the target for the encoder. The actual frame rates of the final videos with the lower encoding bit rates deviated from this value.

Table 3: Encoding bitrates for video segments

Interval	Time (secs)	Encoding bitrate video	Encoding bitrate audio
1	1-20	224 kbps	16 / 32 kbps
2	21-40	192 kbps	16 / 32 kbps
3	41-60	160 kbps	16 / 32 kbps
4	61-80	128 kbps	16 / 32 kbps
5	81-100	96 kbps	16 / 32 kbps
6	101-120	64 kbps	16 / 32 kbps
7	121-140	32 kbps	16 / 32 kbps

Figure 3: Application with volume control and feedback buttons 'Acc.' and 'Unacc.'

4.2. Participants and Procedure

Most of the 128 paid participants (83 women and 45 men) were university students. The age of the participants ranged from 18 to 67 with an average of 24 years. They came from 26 different countries. English was the first language for 72 of the participants.

The participants were told that a technology consortium was investigating ways to deliver TV content to mobile devices, and that they wanted to find out the minimum acceptable video quality for watching different types of content. The instructions further stated: *"If you are watching the coverage and you find that the quality becomes unacceptable at any time please click the button labelled 'Unacc'* […] *When you continue watching the clips and you find that the quality has become acceptable again then please click the button labelled 'Acc'.*

Once it was clear that they understood the instructions, participants were provided with headphones and an iPAQ and given a minute to practice pressing the buttons on the display. When they were ready the experiment began and the participants watched 16 clips in succession, four of which were news clips. Sixty-four participants saw the clips increasing in size and the other 64 saw the clips decreasing in size. The participants' ratings, i.e. the taps on the 'Unacc.' and 'Acc.' buttons, were recorded on the device. At the end of the video rating session, we interviewed the participants to find out what aspects of the video quality they found unacceptable.

4.3. Results

Before analyzing the results, we conservatively coded each 20-second interval of a clip as *unacceptable* if the participant had the application in the unacceptable state at any point during that period. Nonparametric Mann-Whitney tests revealed a significant difference in the way native and non-native speakers judged the acceptability of the news clips. Non-native speakers found the quality of the news clips more acceptable than the native speakers [Z=-4.243, p<.001]. Because of this we excluded the non-native speakers from the following analysis. From here on, 'participants' only refers to the 72 native speakers (48 women, 24 men, average age 24) unless stated otherwise.

The resulting data were analyzed using a binary logistic regression to test for main effects and interactions between the independent variables – *Image Resolution*, *Video Bitrate* and *Audio Bitrate*. The control variables *Gender*, *Corrected Vision* and *Resolution Order* were also included in this analysis. Tests between resolutions were performed using nonparametric Mann-Whitney tests.

The binary logistic regression revealed significant effects for the control variables *Corrected Vision* and *Resolution Order*. Those participants who had corrected vision found the video quality of the News clips more acceptable than the participants who had non-corrected vision [$\chi^2(1) = 5.37$, P=0.021]. In line with the influence of expectations on perceived quality[21] we found that participants who saw the clips decreasing in resolution found the video less acceptable than the participants who saw the clips increasing in resolution [$\chi^2(1) = 75.88$, P<0.001]. Gender was not a significant predictor for acceptability [$\chi^2(1) = 0.097$, n.s.].

The analysis revealed a significant difference in video quality acceptability between the four resolutions [$\chi^2(1)=141.08$, P<0.001]. However, there was no significant difference between the two largest sizes 240x180 and 208x168 [Z=−0.257 n.s.]. As expected, there were also significant differences between the seven video bitrates tested [$\chi^2(6) = 202.01$, P<0.001], with higher video bitrates leading to higher service acceptability. However, when supported with the higher audio bitrate (32kbps) the video quality was judged significantly less acceptable than when it was supported by 16kbps audio [$\chi^2(1)=12.41$, P<0.001]. When averaged across all resolutions and video encoding bitrates the acceptability of the clips presented with the higher audio encoding bitrate was 12% less in relative terms when compared to their presentation with 16kbps audio. This phenomenon is discussed in more detail elsewhere[22]. We found no interactions between the independent variables. The acceptability of the four resolutions at the seven encoding bitrates for the clips shown with 32kbps audio are presented in Figure 4 along with the grayed graphs from Study 1 for comparison.

4.4. Qualitative Results

When asked why they rated the video quality of news as unacceptable, participants mentioned a number of factors. Across the 72 participants, a total of 171 comments described the reasons for the unacceptability of news coverage. Of these comments, 34% related to *text detail*, i.e., the legibility of the news ticker, the headline text, the clock, the logo, or the captions for the people being interviewed by the newscaster. Other problems people reported were *facial details* and expressions from anchor person and field reporter. A summary of these problems and the frequency with which they were mentioned is presented in Figure 5.

4.5. Discussion

Both quantitative and qualitative results indicate that the primary effect of reducing image resolution was a loss of visual detail. The effect of reducing image resolution was more pronounced when bandwidth was abundant. When bandwidth was scarce, there was little or no effect of reducing the image resolution, as visual detail was already low. At 128kbps and above, there was a sharp reduction in acceptability when image resolution was dropped from 168x126 to 120x90.
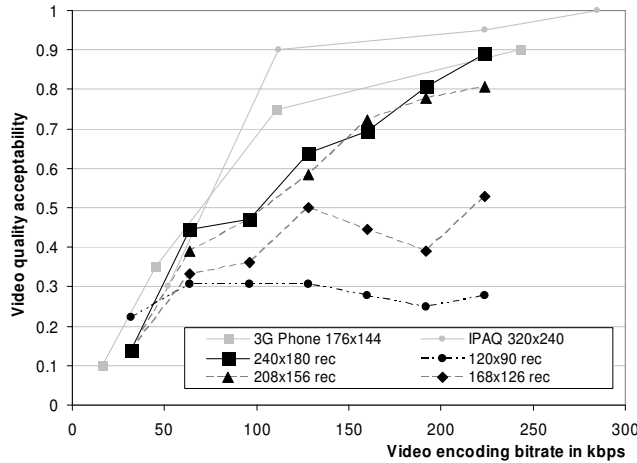
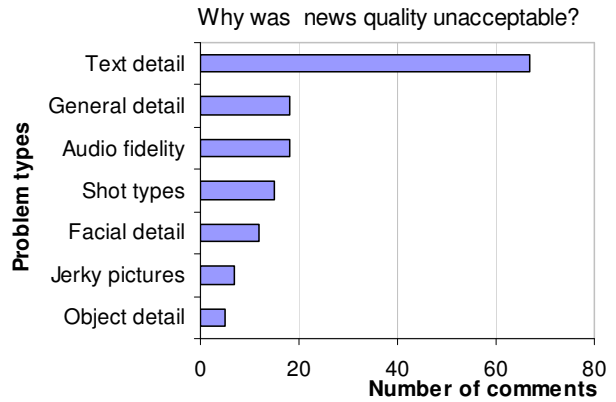**Figure 4: Acceptability on an iPAQ with 32kbps audio**

Figure 5: Reasons for unacceptable news quality

One possible explanation for why the values for Study 2 are much lower than the ones obtained in Study 1 is that the participants' experience was marred by the amount of screen space left unused on the mobile device. Many people dislike letterboxing widescreen content in regular television. Many owners of 16:9 TV sets do not mind the distortion involved in stretching 4:3 content to their 16:9 screens because they prefer to use the whole screen width. However, we only received two comments that hinted at this explanation. The qualitative comments, however, exposed another problem. Of the seven most frequently cited problems, four relate to identifying or distinguishing detail – such as text and facial features. The primary detail on which news video quality was judged was the ability to distinguish textual information – whether the news ticker, the clock, headline text or person names. It could be that the reduction in resolution from 240x180 to 208x156 was compensated by a perceived increase in the quality of the text. If text were coded and transmitted separately from the video we would expect clips encoded at an image resolution of 240x180 to be more acceptable than 208x156. The dominance of text legibility would also explain the differences in acceptability between Study 1 and Study 2. The clip used in Study 1 did not use small font text and at a video encoding bitrate of 128kbps reached an acceptability of 75% on the NEC phone at 176x144. The clips in Study 2 used a lot of small font text and the acceptability of the video at 128kbps was only 53% on the iPAQ at a comparable resolution of 168x126. In Study 2 the size of the text in the ticker line had ranged from three (at 120x90) to six pixels (at 240x180), rendering the text illegible for the videos at resolutions of 120x90 and 168x126. Comparably few comments were made about the audio quality. However, the unexpected effect of lower audio quality resulting in better video acceptability made these comments hard to interpret. Furthermore, since the participants were rating the video quality rather than the combined audio-visual quality no conclusion could be reached about which audio encoding bitrate would make a good match for the tested video encoding bitrates. The relatively few comments about low frame rates referred to picture quality as 'jerky'.

In short, the main insights from Study 2 are:
1. Reducing image resolution has the biggest effect on acceptability at higher bandwidths (>100kbps).
2. Text legibility is the biggest problem of recoded TV news content.

However, it was not clear from the results whether, and to what extent, the text quality influenced the overall video quality perception because Study 2:
1. employed text too small to be legible, and
2. did not test participants for their visual acuity.

In order to look into this more closely we extended Study 2 by a follow-up study in which we controlled for text legibility and the participants' visual acuity.

5. STUDY 3

Study 2 suggested that text quality significantly influenced peoples' video quality perception. Text can be more efficiently transmitted separately through special media formats such as QuickTime and SMIL, which allow for rendering and integration of textual information with the video information on the receiver side. In this study we simulated a separate text delivery by overlaying the textual parts of the degraded news clips with non-degraded text information.

5.1. Text Size Requirements

Previous studies on text legibility in the HCI domain have looked at the various dimensions, e.g. contrast and color, formatting, size, and dynamism, all of which influence reading performance on computer screens (see[23] for an overview). According to Sanders et al.[24] the ratio between text height (h) and reading distance (d) can be expressed by the Visual Angle (VA) in arc minutes with the following equation:

$$VA = 3438*h/d$$

For people with 20/20 visual acuity, the minimum readable text size is five minutes of arc[25]. The ANSI recommends a minimum size of 16 minutes of arc[26], while the US military standard is 15 minutes of arc for the principal viewer and 10 minutes at the maximum viewing distance[27]. Resized TV content that includes text viewable at regular TV setups, e.g. from 720x576 pixels down to 120x90 pixels, quickly falls below the minimum of five minutes of arc, which renders text illegible. Fonts also need to be at least five pixels in height to be legible. The letter 'E', for example, needs three rows for the strokes and 2 for the spaces in between. These values depend on appropriate contrast, which suffers when text is encoded as part of a low bitrate video stream.

5.2. Equipment and Material

We used the same equipment and encoding scheme (see Table 3) as in Study 2. We included the previous four News clips and added four more clips from the same source.

Before the encoding we made the following alterations to the footage to obtain legible text for the ticker line, the logo, and text inserts at all resolutions.

1. In order to obtain content with an aspect ratio of 4:3 and an enlarged ticker we cropped off 36 pixels from the left, 14 from the right and 14 from the bottom (the section below the ticker line).

2. The logo area in Figure 6, which contained both the logo and a clock, was overlaid by a bigger version containing only the logo.

3. News headlines that were inserted in the area right of the logo (highlighted in Figure 7) were overlaid with bigger font size versions that were still legible at the smallest resolution (120x90).

4. The area below the logo that featured a word to contextualize the ticker text was used to extend the space for the ticker as shown in Figure 7.

5. Varying lengths of the original ticker line were used such that in the final version the text ran across the whole horizontal length of the picture. The height of the ticker line ranged from 9 to 12 pixels for the four resolution sizes with the respective text height of the capitalized text ranging from approximately six to eight pixels. At a viewing distance of 40cm this resulted in a viewing angle of the ticker text of 11 arc minutes for the smallest image size. The rest of the picture was slightly condensed in the vertical dimension to accommodate the ticker while meeting the target pixel estate laid out in Table 1.

Figure 6: Video content before cropping and overlaying. Areas outside the section marked off by the dotted line were cropped or overlaid.

Figure 7: Cropped and altered video at 240x180 with the region for text inserts highlighted, which was also overlaid for the duration of the inserts.

The resulting clips were then subjected to the same coding scheme that was used in Study 2. After the encoding we overlaid the ticker line, the logo and possible text inserts with the high quality footage of the clips before the encoding. Two of the eight clips contained text in the main window (see Figure 6), which was rendered illegible by lower resolutions. We chose to include a control variable for these clips in the analysis. The audio that accompanied the videos was encoded at 32kbps in stereo (Windows Media Audio V9).

5.3. Participants and procedure

The majority of the 20 paid participants (9 women, 11 men) were university students. All of them were native English speakers. Their average age was 28. The participants completed a two-eyed Snellen test for 20/20 vision and an Ishihara test for color-blindness. Of the 20 participants 18 had a visual acuity of 100% or better and the remaining two had 90%. One of the participants did not pass the color-blindness test.

Apart from watching eight clips in succession the methodology and the experimental interface was identical to the one in Study 2. Half the participants watched the clips increasing in resolution, while the other half watched them decreasing.

5.4. Results

As in Study 2 we coded the results conservatively in the 20-second segments. We analyzed the resulting data using a binary logistic regression to test for main effects and interactions between the independent variables – *Image Resolution* and *Video Bitrate. Control variables Resolution Order, Gender, Visual Acuity* and *Text in main window* were also included.
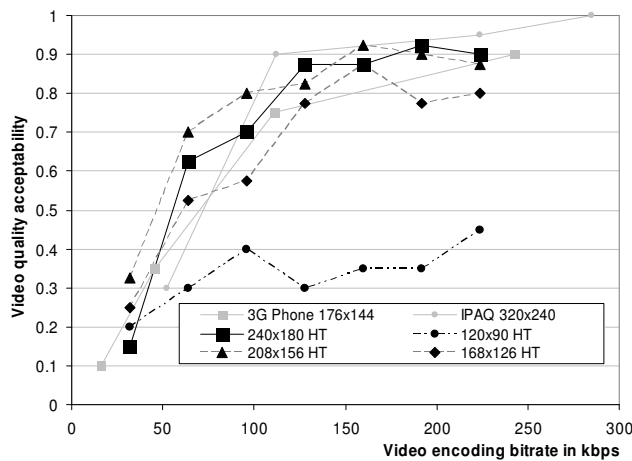

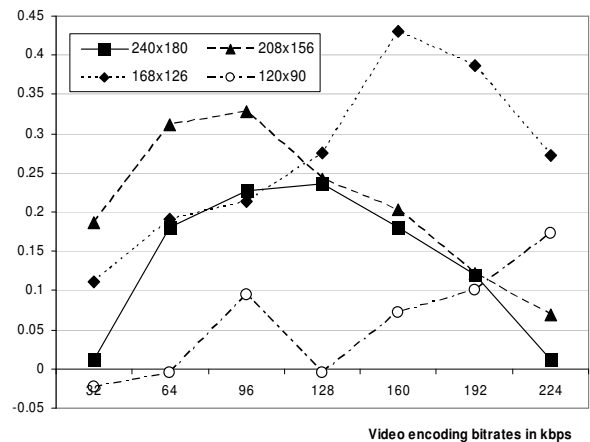Figure 8: News acceptability with high quality text (HT)

Figure 9: Acceptability gains seen when comparing corresponding clips in Study 3 and Study 2

The regression revealed significant effects for the following control variables. In this study *Resolution Order* was not a significant predictor of acceptability [$\chi^2(1) = 0.67$, n.s]. Gender was a significant predictor of acceptability with men less likely to rate a clip unacceptable than women [$\chi^2(1)=13.47$, P<0.001]. *Text in main window,* which was used to distinguish the two clips with small text in the main window from the other clips, was also a significant predictor of acceptability [$\chi^2(1)= 21.76$, P<0.001]. As could be expected from the results of Study 2, videos without text in the main window were less likely to be rated unacceptable than those that did. The control variable *Visual Acuity* was a significant predictor of acceptability [$\chi^2(1)=13.22$, P<0.001]. The two participants with less than 100% visual acuity found the video quality significantly less acceptable than the participants with 100% or better visual acuity.

As in Study 2 *Image Resolution* [$\chi^2(3)=131.53$, P<0.001] and *Video Bitrate* [$\chi^2(6)=141.73$, P<0.001] were significant predictors of acceptable video quality and are summarized in Figure 8 along with the grayed graphs from Study 1. However, among the three biggest resolutions there was only a significant difference in acceptability between 208x156 and 168x126 [Z=-2.881, P=0.004].

At video encoding bitrates of 160kbps the acceptability reached asymptote for all resolutions. The acceptability of video quality of the three biggest resolutions dropped off for encoding bitrates smaller than 160kbps. The drop was most dramatic for encoding bitrates below 64kbps. The acceptability of the smallest resolution (120x90) only profited from increases in encoding bitrates up to 96kbps. Even at this point its acceptability was low compared to higher resolution clips at the same encoding bitrate and did not get beyond 40% even for higher encoding bitrates. The only exception was at 224kbps where the acceptability of video quality was 45% - possibly an artifact of the experimental design.

We computed the acceptability gains from legible high quality text, i.e. the difference in video quality acceptability, between Study 3 and 2 for the different resolution clips all of which were shown with 32kbps audio. We summarized these scores in Figure 9. Post-hoc tests between the studies were performed using nonparametric Mann-Whitney tests. Despite an apparent trend for an acceptability gain for 120x90 for encoding bitrates of 160kbps and higher there was no significant difference in the overall acceptability scores between Study 3 and 2 for the smallest resolution [Z=-1.444,n.s]. However, all of the three biggest resolutions saw significant gains in acceptability. The gains were most substantial when the video was encoded at bitrates between 64kbps and 196kbps. The acceptability gains from legible high quality text were less pronounced for resolutions 240x180 and 208x156 at very low (32kbps) and high (>196kbps) encoding bitrates. The resolution 168x126 benefited the most from the text enhancements especially at video encoding bitrates greater than 128kbps. At 160 and 192kbps encoding bitrates its gains in acceptability were bigger than 90% in relative terms.

5.5. Qualitative Results

With textual quality controlled in Study 3, the biggest issue identified by the qualitative feedback became the insufficient size of the video. Typical complaints were that it was "…just too small to figure out what *is* going on". This was most notable and irritating for facial features. When eyes, eyebrows, lips and mouth could not be identified or were not sufficiently clear in the footage, this was a prominent trigger for people to judge the quality as unacceptable. Participants who complained about inadequate text quality usually referred to text in the main window and the ticker at the smallest resolution that was just legible. Another area of complaint was the insufficient smoothness in terms of frames per second and the inherent audio-visual asynchrony resulting from it. This occurred especially at the smaller encoding bitrates where the encoder had to skip frames in order to meet the target bitrates. A majority of participants stated that the biggest picture was the best and that in general the bigger the picture became the better the quality got.

5.6. DISCUSSION

Despite the better legibility of the ticker, logo and inserts in this study compared to Study 2 the acceptability of video quality dramatically decreased when image resolution was reduced to 120x90 pixels.

The resolution 168x126 benefited the most from the simulated separate text delivery introduced in Study 3. In Study 2 this resolution had rendered text illegible with font sizes dropping below 5 pixels.

The two biggest resolutions that had not rendered the ticker text illegible in Study 2 also benefited from separate text delivery. However, we believe that at low video encoding bitrates the poor resolution of the visual content in the main window dominated and therefore the effect of the quality text was less pronounced. At the high encoding bitrate (224kbps) we assume that the text in Study 2 was rendered at a comparable level to what the separate text delivery in Study 3 presented to the participants. We did not see a statistically significant difference in the acceptability of clips presented at 208x156 and 240x180 when presented with high quality text as had been expected after Study 2. Surprisingly, there was not a significant difference between 240x180 and 168x126, either, when text quality was high. More research is required to understand the reasons behind this.

Overall, the results obtained in Study 3 are much closer to those obtained in Study 1 and the main reason for the differences between Study 2 and 1 were due to text legibility.

The fact that only part of the screen estate was used in Study 3 did not seem to have a big effect on the acceptability of the visual quality of the presented clips when compared to the clip used in Study 1.

In summary, Study 3 showed that:
1. Separate text delivery that secured legible text with high visual quality offered tremendous gains in acceptability of overall perceived video quality.
2. With high quality text
   a. the acceptability of news for all resolutions reached asymptote at an encoding bitrates of 160kbps, and
   b. a resolution of 168x126 was feasible for the presentation of recoded TV News on a mobile device.

Despite the large variety of shot types and kinds of footage presented in News programs we would like to caution against the generalization of the presented results to other content types. In many cases the visuals in News programs were not essential to understanding the information which was to a high degree in the audio track. Other content types might have different requirements as presented for example in Knoche et al[4].

6. SUMMARY

The popular mobile content News consists of three different media types: video, audio and text. The quality of all these are important to the perceived video quality on mobile devices. When following news content at low resolutions participants complained about the loss of detail especially for text legibility. If we do not simply recode TV content "as is" for mobile consumption but assure that the text used, e.g. in ticker, logos, inserts, etc., is legible, the perceived acceptability of video quality can be dramatically increased. This could be achieved by either changes in layout and/or separate delivery and rendering of text at the receiver. Assuming separate text delivery video encoding bitrates of 160kbps resulted in video quality acceptability approaching their asymptote for all resolutions from 120x90 to 320x240. When both separate text delivery and recoding of the content are not feasible solutions service providers should try to avoid using resolutions that render textual information of news content illegible. The maxim for resolution, then, is the higher the resolution the better.

Furthermore, participants felt strongly about details that help to identify objects as well as facial details. Lip-reading aids speech perception and the mouth, eyes, and eyebrows all contribute to the decoding of facial expressions. Mobile TV producers have to make sure that small screens, low resolutions, the employed shot types and the used encoding bitrates still adequately render these features of depicted people to the user.

The biggest limitation of the three studies at hand is that they assumed perfect delivery of the data. The impact of data corruption or loss on the perceived video quality of News content which might play an important role in mobile broadcast delivery was not addressed.

ACKNOWLEDGMENTS

REFERENCES

1. U. Reimers, "Digital Video Broadcasting", *IEEE Communications Magazine,* 36, 1998.
2. C. Selier and N. Chuberre, "Satellite Digital Multimedia Broadcasting (SDMB) System Presentation", *Proceedings of 14th IST Mobile & Wireless Communications Summit,* 2005.
3. H. Knoche and J. McCarthy, "Design Requirements for Mobile TV", *Proceedings of Mobile HCI,* pp. 69-76, 2005.
4. H. Knoche and J. McCarthy, "Mobile Users' Needs and Expectations of Future Multimedia Services", *Proceedings of the WWRF12,* 2004.
5. C. Södergård, "Mobile television - technology and user experiences Report on the Mobile-TV project", P506, VTT Information Technology, 2003.
6. R. Weiner, "Webster's New World Dictionary of Media and Communications", rev. and updated, Macmillan, New York, NY, 1996.
7. D. A. Owens and K. Wolfe-Kelly, "Near Work, Visual Fatigue, and Variations of Oculomotor Tonus", *Investigative Ophthalmology and Visual Science,* 28, pp. 743-749, 1987.
8. B. Reeves and C. Nass, "The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places", University of Chicago Press, 1998.
9. M. Lombard, M. E. Grabe, R. D. Reich, C. Campanella, and T. B. Ditton, "Screen Size and viewer responses to television: A review of research", *Annual Conf of the Assoc for Education in Journalism and Mass Communication,* 1996.
10. K.-I. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita, "Multiparty videoconferencing at virtual social distance: MAJIC design", *Proc ACM conf on Computer supported cooperative work,* pp. 385-393, 1994.
11. M. Edwardson, K. Kent, and M. McConnell, "Television news information gain: Videotext versus a talking head", *Journal of Broadcasting & Electronic Media,* 29, pp. 367-378, 1985.
12. B. Reeves, A. Lang, E. Kim, and D. Tartar, "The effects of screen size and message content on attention and arousal", *Media Psychology,* 1, pp. 49-68, 1999.
13. D. B. Horn, "The effects of spatial and temporal video distortion on lie detection performance", *Proceedings of CHI '02,* 2002.

14. J. K. Kies, R. C. Williges, and M. B. Rosson, "Controlled Laboratory Experimentation and Field Study Evaluation of Video Conference for Distance Learning Applications", HCIL 96-02, Virginia Tech, 1996.

15. P. J. Barber and J. V. Laws, "Image Quality and Video Communication", *Proceedings of IEEE International Symposium on Multimedia Technologies & their Future Applications,* R. Damper, W. Hall, and J. Richards, pp. 163-178, Pentech Press, London, UK, 1994.

16. W. R. Neumann, A. N. Crigler, and V. M. Bove, "Television Sound and Viewer Perceptions", *Proc Joint IEEE/Audio Eng Soc Meetings,* pp. 101-104, 1991.

17. J. G. Beerends and F. E. de Caluwe, "The influence of video quality on perceived audio quality and vice versa", *Journal Audio Eng Soc,* 47, pp. 355-362, 1999.

18. VQEG, "Final Report from the video quality experts group on the validation of objective models of video quality assessment", http://www.vqeg.org, 2000.

19. Packet Video, "PV Author", *http://packetvideo com,* 2005.

20. J. McCarthy, M. A. Sasse, and D. Miras, "Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video", *Proc CHI,* pp. 535-542, 2004.

21. A. Bouch and M. A. Sasse, "The case for predictable network service", *Proceedings of MMCN'00,* pp. 188-195, 2000.

22. H. Knoche, J. McCarthy, and M. A. Sasse, "Can Small Be Beautiful? Assessing Image Resolution Requirements for Mobile TV", *Proc of ACM Multimedia 2005,* pp. 829-838, ACM, 2005.

23. C. Bergfeld Mills and L. J. Weldon, "Reading Text from Computer Screens", *ACM Computing Surveys,* 19, pp. 329-358, 1987.

24. M. S. Sanders and E. J. McCormick, "Human Factors in Engineering and Design", 7th, McGraw-Hill Inc., New York, NY, 1993.

25. I. L. Bailey and J. E. Lovie, "New Design Principles for Visual Acuity Letter Charts", *American Journal of Optometry & Physiological Optics,* 53, pp. 740-745, American Academy of Optometry, 1976.

26. American National Standards Institute (ANSI), "American national standard for human factors engineering of visual display terminal workstations", ANSI/HFS Standard No.100-1988, The Human Factors Society Inc., Santa Monica, CA, 1988.

27. G. Musgrave, "Legibility of Projected Information", http://*www.conceptron.com/articles/pdf/legibility_of_projected_information.pdf,* 2001.