Reading the Fine Print: The Effect of Text Legibility on Perceived Video Quality in Mobile TV

Hendrik O. Knoche
University College London
Gower St
London, WC1E 6 BT, UK
+ 44 207 679 3642
h.knoche@cs.ucl.ac.uk

John D. McCarthy
University College London
Gower St
London, WC1E 6 BT, UK
+ 44 207 679 3644
j.mccarthy@cs.ucl.ac.uk

M. Angela Sasse
University College London
Gower St
London, WC1E 6 BT, UK
+ 44 207 679 7212
a.sasse@cs.ucl.ac.uk

ABSTRACT
Mobile TV services are available in an increasing number of countries. For cost reasons, most of these services offer material directly recoded for mobile consumption (i.e. without additional editing). This paper reports the findings of a study on the influence of text legibility and quality on the perceived video quality of mobile TV content. The study, with 64 participants, examined responses to news footage presented at four image resolutions and seven video encoding bitrates. The results showed that a simulated separate delivery of a news ticker and other textual information significantly increased the perceived video quality of the entire screen for native speakers. In addition, some automatable changes to the layout of news content resulted in substantial increases in perceived video quality. The results can be used to quantify the perceived quality gains when considering text delivery separately from the video stream and in the development of more accurate multimedia quality models.

Categories and Subject Descriptors
H5.m. Information interfaces and presentation: Miscellaneous

General Terms
Design, Experimentation, Human Factors

Keywords
Video quality assessment, text quality, mobile TV

1. INTRODUCTION
Mobile TV services have become available in an increasing number of countries. The deployment of these services has been driven mainly by technical concerns. With uptake falling short of expectations, service providers are examining ways to improve what the user experiences and values in mobile TV, i.e., the Quality of Experience (QoE). To assist service providers we need to understand how people experience multimedia content on mobile devices. This knowledge can aid in the configuration of hard- and software of user terminals and the delivery of mobile TV services. News has been identified as one of the most desired mobile TV content types. Text appears in many TV programs and especially frequently in news programs because producers often use subtitles, headlines, ticker lines, and logos. If not transmitted separately, text represents a medium within the medium of video. Small resolutions and low encoding bit rates render it illegible. However, the degree to which imperfectly rendered text affects the overall perceived video quality is unknown. Therefore, we cannot compare the cost of sending text and video separately to gains in experienced video quality. To address these trade-offs, we designed a study to assess the impact of the visual quality of text on the overall perceived video quality and to measure the respective gain by using special protocols or multimedia formats, e.g. SMIL or QuickTime, which can include text separately. Because of the diversity of user equipment and bandwidth limitations we tested different image resolutions of mobile TV news at a range of encoding bitrates.

In this paper we first review the effects of image size, reduced resolution and text size requirements. In Sec. 3 we describe the study on the effect of text quality at different resolutions on video quality and present the results. A discussion of the results follows in Sec. 4. The conclusions in Sec. 5 include recommendations for mobile TV news delivery.

2. BACKGROUND
Normal 20/20 vision is classified as the ability to resolve one minute of arc (1/60º). According to research on TV quality approximately 22 cycles (44 pixels) per degree is perceived as a sharp image [9]. The iPAQ used in our study had a resolution of 320x240 (111ppi), which equates to approximately 15cycles/º at a typical viewing distance of 40cm – classified as low to normal resolution in TV terms. For people with 20/20 visual acuity, the minimum readable text size is five minutes of arc [2]. The ANSI recommends a minimum size of 16 minutes of arc [1], while the US military standard is 15 minutes of arc for the principal viewer and 10 minutes at the maximum viewing distance [8]. Text in regular TV content quickly falls below the minimum of five minutes of arc when resized, e.g. from 720x576 pixels down to 120x90 pixels. 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.

Previous studies on text legibility in the HCI domain have looked at the various dimensions, e.g. contrast & color, formatting, size, and dynamism, all of which influence reading performance on computer screens (see [4] for an overview). To the best of our knowledge, however, no work has been conducted on the influence of text quality and its legibility on the perceived quality of video that includes text. Until now, only the study by Knoche et al. has shown that text quality might greatly influence peoples' video quality perception in mobile TV content and found that displays of news at resolutions of 168x126 and less received low levels of acceptability [6]. However, it is not clear from the results of the study to what extent text quality and legibility influenced the overall video quality perception because

1. it employed illegible text (at 120x90 and 168x126) which was smaller than five pixels in height and
2. participants were not tested for their visual acuity.

3. TEXT QUALITY STUDY

The aim of our study was to evaluate the effects of text quality on the overall acceptability of image quality at different image sizes/resolutions and encoding bitrates. In the experiment we used a between subjects design, in which one half of the participants saw news footage with *inline* text that degraded with the rest of the image. The other half experienced a simulated *separate* text delivery and saw the same footage with unimpaired text at high quality. This setup should help to answer how much can be gained from transmitting text separately from the video and rendering the composed video at the user terminal. We will use the terms *inline* and *separate* text when we want to emphasize the implications for delivery and refer to the corresponding conditions as high quality text and degrading text in this study.

The devices had a fixed resolution. By varying the resolutions of the content the size it occupied on the screen varied proportionally. In other words, the smaller resolution videos were represented by fewer pixels on the user terminal. The participants were able to freely adjust the viewing distance to the device such that the pixels per degree could be changed according to their preferences. The study employed the binary assessment technique as introduced in [7] in which participants judge the video quality as acceptable or unacceptable and the effect of text quality was thereby measured indirectly.

3.1 Material

For the purposes of this study, we investigated the acceptability of directly recoded TV news without any manual editing steps. To allow for comparison we included four news clips that were used in [6], one of which included additional small text within the main window of the picture apart from the text components described in the following section. We recorded four additional news clips from the same digital TV channel in the UK (BBC24 news). These eight clips included a range of typical news coverage consisting of anchor person shots, stills, graphics and field reports. Each clip lasted approximately 2:20 minutes.

3.1.1 Preparation of the clips

The clips were cropped to 532x399 pixels from the original 720x576 to remove the letterboxing and to create a picture with a 4:3 aspect ratio. To control the influence of text quality the following alterations were made to the video material:

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 part below the ticker). The logo area, which contained both the logo and a clock, was overlaid by a bigger version containing only the logo. News headlines that appeared temporarily in the area to the right of the logo were overlaid with bigger font size versions that were still legible at the smallest resolution (120x90). The area below the logo that featured a word to contextualize the ticker text was used to extend the space for the ticker. 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. This allowed for comparisons of results with [6] since the amount of presented information was approximately equivalent (compare Figure 1 left and right).

Figure 1: Content before (l.) and after (r.) editing steps. Text inserts appeared in the hatched area.

To ensure comparability of results with [6] the audio was encoded at 32kbps in stereo (WMV V9). We encoded the video clips at four resolutions (240x180, 208x156, 168x126, 120x90) and manipulated them in two ways. Within each news clip the bitrate allocated to video was gracefully degraded every 20 seconds in steps of 32 kbps from a maximum of 224kbps down to 32kbps. The boundaries of the intervals were not pointed out to the participants. They were told only that the quality would change and were presented with 16 clips, each of which gradually decreased in quality. In addition to changing video bitrate within a clip, two duplicate sets of clips were produced with different text qualities.

We used *Virtualdub* to segment the source clips into seven 20 second-long clips at 12.5fps and at all resolutions using a bicubic resize. These segments were encoded using Windows Media Encoder (WME), which used the MS Media Video V8 codec with different bitrates for the segments. Each group of seven WMV segment files was then converted and concatenated to one AVI file using *TMPGEnc Express*. From these videos we produced a second set with high text quality in which the ticker line, the BBC logo, and text inserts above the ticker were replaced with the footage before the described degradation. Both the degrading and the high quality text versions were then subjected to a final encoding using WME. The video was encoded at a much higher bitrate than the maximum of the first WME encoding in order to

prevent significant alterations to the video quality. Two of the eight clips contained text in the main window that was rendered illegible by lower resolutions. For better comparison with [6] we chose to include these clips in the tested set and included a control variable for them in the analysis.

3.2 Design

The experimental design followed the one used in [6]. We ran four groups. Each group of 16 participants viewed eight clips in groups of two clips at each of the four resolutions. The groups differed in whether they experienced *increasing* or *decreasing* image resolutions and whether the text quality of the ticker, the headline inserts, and the news logo was *degrading* along with the video quality or of constant *high quality*. Within each group, we ran eight variations to control for content using a Latin squares design. This ensured that the different content clips were tested at each of the image resolutions across participants. The dependent variable was *Video Quality Acceptability*. The independent variables were *Image Resolution*, *Video Encoding Bitrate*, *Text quality.* Control variables were *Resolution Order*, *Sex*, *Native English Speaker, Text in Content*, and *Normal Vision*. We used the control variable *Text in Content* to identify the two aforementioned clips that contained small text in the main window. The variable *Normal Vision* coded whether participants had 100% visual acuity according to the administered Snellen test [3].

3.3 Equipment

The test material was presented on an iPAQ 2210 with a 400Mhz X-scale processor, 64MB of RAM and a 512MB SD card. The screen was a transflective TFT display with 64k colors and a resolution of 240x320. The iPAQ was equipped with a set of Sony MDR-Q66LW headphones to deliver the audio. We used the same interface as in [6] to present the clips.

3.4 Procedure

After completion of a two-eyed Snellen test for 20/20 vision 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 news. The instructions stated: *"If you are watching the coverage and you find that the video 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'… you can hold the PDA at any distance that is comfortable for you."* The participants watched eight clips in succession. Each clip started with the interface in the 'Acc.' state.

3.5 Participants

Most of the 64 paid participants (31 women and 33 men) were university students. The age of the participants ranged from 19 to 67 with a median of 25 years. The majority came from the UK (25) and China (18). English was the first language for 36 of the participants. Visual acuity was 100% for 48, 95% for seven, 85% for seven, and 80% or below for two of the participants.

3.6 Results

Before analyzing the results, we conservatively coded each 20 second interval of a clip as *unacceptable* if the video quality had been unacceptable at any point during that period. The resulting data was analysed using a binary logistic regression to test for main effects and interactions between the independent variables. The regression revealed significant effects for the following control variables. *Resolution Order* was a predictor of acceptability [$\chi^2(1) = 4.2$, $P < 0.05$]. The participants who started with large image resolutions that decreased during the experiment were generally more likely to rate the quality unacceptable than those who saw clips increasing in image resolution. *Sex* was a significant predictor of acceptability with men being less likely to rate a clip as unacceptable than women [$\chi^2(1) = 45.5$, $P < 0.001$]. The variable *Native English Speaker* was also a significant predictor for the experienced acceptability of the video quality [$\chi^2(1) = 14.7$, $P < 0.001$]. Native English speakers were less likely to rate the quality as unacceptable compared to non-native speakers (see Sec. 3.6.1).

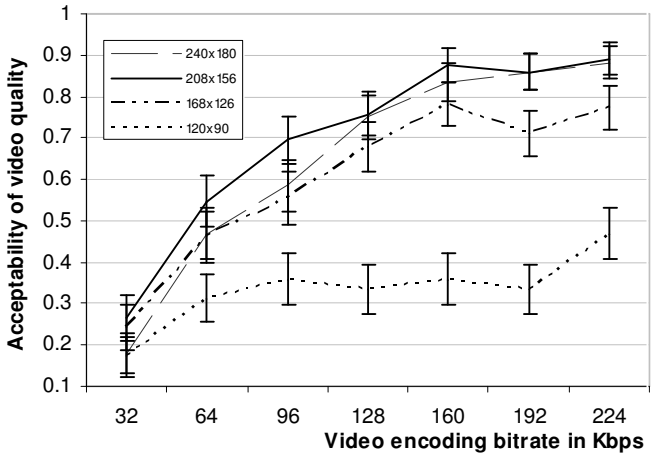

Figure 2: Acceptability of news at different resolutions and encoding bitrates of all participants with standard error bars

Text in main window, which was used to distinguish the three clips with text in the main window from the other clips, was also a significant predictor of acceptability [$\chi^2(1)=7.5$, $P = 0.01$]. Videos without text in the main window were less likely to be rated unacceptable than those that did. The control variable *Visual Acuity* was not a significant predictor of acceptability [$\chi^2(1)=2.2$, n.s.]. As expected and in accordance with [6], *Image Resolution* [$\chi^2(3)=270.7$, $P<0.001$] and *Video Bitrate* [$\chi^2(6)=414.6$, $P<0.001$] were significant predictors of acceptable video quality and are summarized in Fig. 2 averaged across the two text qualities. Despite the legibility of the text in terms of size in this study compared to [6] the acceptability of video quality still dropped dramatically when image size was reduced to 120x90 pixels.

3.6.1 The effects of Text Quality

Across all participants text quality was not a significant predictor of the acceptability of video quality [$\chi^2(1) =2.4$, n.s.]. This is due to the fact that the opposing ratings of the non-native and native speakers cancelled each other out. Post-hoc tests revealed an interaction between *Text Quality* and *Native Speaker* [$\chi^2(1) = 40.1$, $P < 0.001$]. This effect came as a surprise. Native speakers who watched clips supported by high text quality rated them higher in terms of acceptability than the non-native speakers. The non-native speakers rated video quality higher when video was accompanied by text that degraded with the video. We partitioned the data set and looked separately at the two groups. Two non-

parametric Mann-Whitney tests showed significant differences for *Text Quality* for both the native speakers [Z=-2.1, P<0.05] and the non-native speakers [Z=-5.3, P<0.001]. We ran the original binary logistic regression without the variable *Native English Speaker* on the partitioned data sets. Along with all the previously described variables *Text Quality* turned out to be a significant predictor of acceptability in the analysis of the native speakers [$\chi^2(1)=8.2$, P<0.01] and the non-native speakers [$\chi^2(1)=21.7$, P<0.001] - but in opposing directions as described above. Similarly, the control variable *Text in main window* was a significant predictor of acceptability [$\chi^2(1)=17.4$, P<0.001] for the native speakers but not for the non-native speakers [$\chi^2(1)=0.01$, n.s.]. Considering the impact of the non-native speakers we will limit the presentation of results to the 36 native speakers. Averaged across all encoding bitrates and resolutions the acceptability of news content increased from 57% with degrading text to 62% when presented with high text quality.

3.6.2 Gains in perceived quality

We compared the results from this study with [6] and measured the acceptability gains of the layout changes described in Sec. 3.1.1 and of inline and separate text delivery for the resolutions 120x90 and 168x126. We only included ratings from native speakers for those four clips that had been used in both studies. In Figure 3 we plotted the acceptability values averaged across all encoding bitrates for the three clips without and the one clip with small text in the main window. The video acceptability of the two groups of clips benefited from both the layout changes and the high quality text. Averaged across the two groups the layout changes improved acceptability from 38% to 50%. From the distance between the two curves we can see that the influence of text on the overall video quality is especially large when the text was presented in the middle of the screen. The acceptability of video clips that had no small text in the main window increased from 42% for degrading text quality to 57% for high text quality.

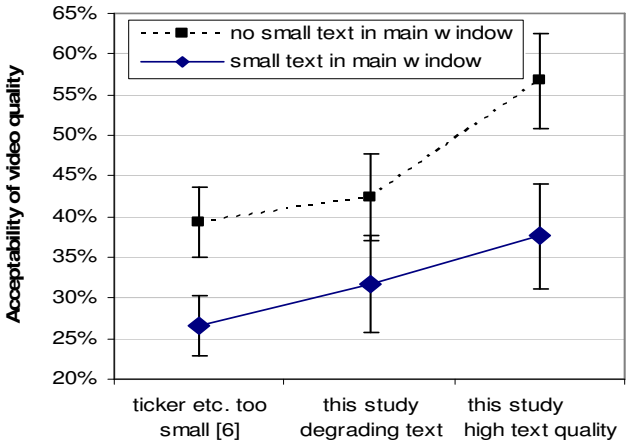

Figure 3: Acceptability of different layouts and deliveries

4. DISCUSSION

From the results we can deduce that text quality has a strong influence on the overall perceived video quality. In terms of acceptability we could save on the encoding bitrate of mobile TV news content to native speakers by roughly 32kbps if the included textual data were rendered in high quality at the receiver. This approach would not work, however, for an international audience. Non-native speakers did not seem to include textual quality considerations into their video quality ratings, which we know because the *Text in main window* variable was not a significant predictor in their quality ratings. Native speakers rated video quality lower when text was shown in the main window and perceived an overall increase in quality when the text quality was high.

5. Conclusion

We cannot recommend reusing unedited TV news for mobile consumption as important text quickly drops below a viewing angle of five minutes of arc or below five pixels and drastically reduces the perceived video quality. This is especially true for text that is presented in the main window in the center of attention. The visual quality of text that appears in the periphery of the screen also influences the perceived overall video quality. The changes we made to our clips ensured text legibility, could be automated and reaped substantial benefits in perceived video quality. Service providers are advised to use resolutions of at least 168x126 or QCIF format (176x144) for mobile news content regardless whether the included text is big enough to read or is delivered separately from the video. Models of audio-visual multimedia quality, e.g. [5] usually consider video quality as a single entity. They could make much better predictions if they incorporated a measure of text legibility into the term encompassing the video quality.

Acknowledgements

Hendrik Knoche is funded by the EU IST-project UNIC.

6. REFERENCES

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

[2] Bailey, I. L. & Lovie, J. E. *New Design Principles for Visual Acuity Letter Charts* American Journal of Optometry & Physiological Optics, 53 (11), p. 740-5, 1976.

[3] Bennett, A. G. *Ophthalmic test types* Br.J.Physiol.Opt., 22, p. 238-71, 1965.

[4] Bergfeld Mills, C. & Weldon, L. J. *Reading Text from Computer Screens* ACM Computing Surveys, 19 (4), p. 329-58, 1987.

[5] Hands, D. S. *A Basic Multimedia Quality Model* IEEE Transactions on Multimedia, 6 (6), p. 806-16, Dec. 2004.

[6] Knoche, H., McCarthy, J., Sasse, M. A. *Can Small Be Beautiful? Assessing Image Resolution Requirements for Mobile TV* in Proc.of ACM Multimedia 2005, p. 829-38, ACM, 2005

[7] McCarthy, J., Sasse, M. A., Miras, D. *Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video* in Proc.CHI, p. 535-42, 2004

[8] Musgrave, G., *Legibility of Projected Information.*, www.conceptron.com/articles/pdf/legibility_of_projected_information.pdf. 2001

[9] Silbergleid, M. & Pescatore, M. *The Guide To Digital Television* (3rd ed.) Miller Freeman Psn Inc, 2000