A close-up on Mobile TV: The effect of low resolutions on shot types

Hendrik Knoche, Angela Sasse, John McCarthy
University College London
h.knoche@cs.ucl.ac.uk

Abstract

The advent of mobile TV with small screens and low resolutions made content producers think about refraining from using shots that depict subjects from a great distance. Shot types that fill the screen with the subject are deemed to be more appropriate for mobile devices. This paper explores how shot types used in regular broadcast television are affected when shown on mobile devices at reduced resolutions. The data were taken from a study with 128 participants who judged the acceptability of four different content types at four resolutions. The results show that acceptability of shot types depends on the content. Contrary to the belief of mobile content producers extreme long shots were not most sensitive to low resolutions in football content. On the contrary, the medium shot which focuses on the upper half of a subject's body was the least acceptable in football. In news content, the same shot was the most acceptable.

1 Introduction

There are many services that aim to provide users with a TV-like experience while on the move. For service providers the simplest and cheapest solution is to deliver TV material without additional editing, but little is known about the technical requirements to deliver an acceptable Quality of Experience (QoE) (Jain, 2004) for this type of service. The QoE of the service depends on the perceived visual, audio and textual quality of the consumed content (Knoche & Sasse, 2006), and the interaction through which the user has to go to access it (e.g. the delay between selecting content and start of play). In this paper, we focus on the former.

The content distributed to mobile devices ranges from highly interactive content specifically created for the mobile to relayed material that is produced for standard TV consumption.

Original TV material may undergo an additional editing process to prepare it for mobile consumption or it maybe be directly encoded. The sports network ESPN, for example, is considering resizing graphics for the small screen and minimizing the use of long shots in their coverage (Gwinn & Hughlett, 2005).

One of the central factors of the visual quality of mobile TV is the resolution of the image. Research has shown that lowering the resolution of TV clips affects the acceptability of the perceived video quality non-uniformly and depends on the kind of content depicted (Knoche, McCarthy, & Sasse, 2005). The content also influences the directors' decision as to which shot types to use while shooting the footage.

Image resolution is a central concern for all actors involved in the field of mobile TV.

Users: Results from focus groups comprised of people unfamiliar with mobile TV show that concerns about screen size (both in terms of watchability and portability) may inhibit uptake (Knoche & McCarthy, 2004). Mobile devices are operated at 'arm's length'; continued viewing at distances closer than the resting point of vergence – approx. 90cm, with a 30º downward gaze – can contribute to eyestrain (Owens & Wolfe-Kelly, 1987). When viewing distances come close to 15cm, people experience discomfort (Ankrum, 1996). Paper, keyboard and display objects are typically operated at distances ranging from 30cm to 70cm.

Device manufacturers: Mobile devices displays come in a range of shapes, sizes and resolutions, from VGA PDAs (480x640 pixels) and high end 3G phones (320x240) to more compact models, e.g., Nokia 6230 (128x128).

Content distributors: If the resolution of TV images can be lowered without affecting the perceived visual quality, less bandwidth is required and more content can be distributed at lower prices.

Content producers: The camera shots used in television range from very wide shots (VWS) to extreme close-ups (XCU) and consider image size and resolution of typical TV setups (Weiner, 1996). Image size and resolution cannot be reduced indefinitely as important detail will be lost. The used shot types might be affected differently by reduced resolutions. Producers of tailor-made mobile content use a mix of shot types to optimize the viewing experience on small low resolution screens.

So far, production of bespoke mobile content is not based on research but on experience, intuition and aesthetics. To date it is unknown, however, if and how shot types affect the perceived visual quality of

mobile TV content at low resolutions and encoding bitrates used in current mobile TV services.
To shed some light on this topic we classified all the shots used in video clips of a previous study (Knoche et al., 2005) according to their shot types. This paper presents the results from the analysis of the previous data set extended by the shot type classification.
Section 2 describes the different shot types and reviews the previous literature on the effects of image size and resolution. Section 3 describes the original study on image resolution and presents the results of our shot type classification, which are discussed in Section 4. The main findings are summarized in Section 5 and the conclusions presented in Section 6.

2 Background

We were unable to find any published reports on the influence of low resolutions on the different shot types used in television content and how these would come across on small mobile devices.
We will therefore review the previous research that examined how the various sizes and resolutions of moving picture content influenced the audience's experience.
Most of the research on the effect of screen sizes has examined the impact of increasing the image size in the viewer's visual field by means of large physical displays or projection areas. Typically these studies have compared very large screens (e.g. 46") to standard sized TV screens (15"-20") (Reeves & Nass, 1998), (Lombard, Grabe, Reich, Campanella, & Ditton, 1996). The results show that larger image sizes are more arousing, better remembered, and better liked than smaller ones. Other studies also show that users generally prefer bigger image sizes – ideally depicting people and objects up to life-size (Okada, Maeda, Ichikawaa, & Matsushita, 1994). However, in another study Reeves et al. found no difference in arousal and attention between users watching 2" and 13" screens, although arousal and attention were greater when watching content on a very large screen (56") (Reeves, Lang, Kim, & Tartar, 1999).
Where TV images are concerned, the general message from these studies is, '*the bigger the better*'. This clearly presents a challenge to mobile TV where there is a trade-off between the screen size and the portability of the device. These concerns have been noted in focus groups assessing the potential uptake of mobile TV services (Knoche et al., 2004). Users want a screen as large as possible for viewing, but they do not want their phones to be too big. Moreover, it is not clear whether users will want higher arousal and immersion in a mobile context, because of the increased risk of errors and accidents.
In a study on resolution requirements on mobile TV Knoche et al. found that in general content shown on mobile devices at higher resolutions are more acceptable than when compared to lower resolutions at identical encoding bitrates. However, the differences were not uniform across content types (Knoche et al., 2005).
Other studies have even shown that smaller image resolutions can improve task performance. For example, (Horn, 2002) showed that lie detection was better with a small (53x40) than a medium (106x80) video image resolution. In another study, however, smaller video resolutions (160x120) had no effect on task performance but did reduce satisfaction when compared to 320x240 image resolutions (Kies, Williges, & Rosson, 1996). 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 (Barber & Laws, 1994).

2.1 Shot types

The language of film represents a cultural technique. The way in which objects are shot, edited, presented and decoded by the audience follows established conventions (Thompson, 1998). The different shot types used in film making help the audience to "read" the message the director wants to convey. Faced with the more constrained visual real estate content producers are considering using a different mix of shot types for mobile TV.
In Asia content creators have started to produce specially made soap operas for mobile devices that are very short and rely heavily on close-up shots with very little dialogue. Most emotions have to be conveyed by means of facial expressions and *"there is very little dialogue and a lot of close-ups of characters striking exaggerated poses"* (Guardian, 2005). In sports coverage for mobile devices ESPN is minimizing the use of long shots in their coverage (Gwinn et al., 2005) and instead using more high-lights with close-up shots.
Unfortunately, the terms used to classify shot types can differ and popular usage of the terms deviates further. In order to have a consistent terminology within this paper we will from now on use a classification based on (Thompson, 1998) which is presented below. We limit our account of shot types to those that were most common in the footage used in this study.

In an extreme long shot (XLS) the subject is barely visible and the recognition of the environment and/or the scene is more important (see Figure 1).

Figure 1: Extreme long shot (XLS)

In a very long shot (VLS) the majority of the frame is still concerned with the environment the subject is in. However, some details of the subject such as clothing and gender are recognizable (see Figure 2).

Figure 2: Very long shot (VLS)

The subject almost covers the frame from top to bottom in a long shot (LS) (Figure 3).

Figure 3: Long shot (LS)

The subject attracts more attention now but not too much detail of the subject can be seen.

In the medium shot (MS) the entire subject does not fit into the frame anymore (Figure 4). The eyes of the subject can be clearly seen.

Figure 4: Medium shot (MS)

The facial expression becomes predominant in the medium close-up (Figure 5). The attention is drawn to the face and the background is not important anymore.

Figure 5: Medium close-up (MCU)

On the close-up the attention is completely drawn to the subject's eyes and mouth (Figure 6).

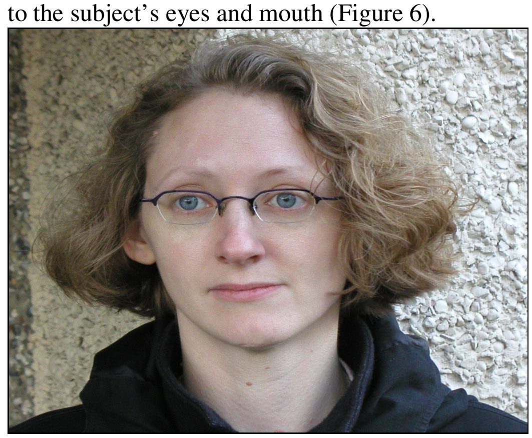

Figure 6: Close-up (CU)

3 Image Resolution Study

The aim of the study (Knoche et al., 2005) was to evaluate the effects of varying image resolution and encoding bitrate on the acceptability of video quality. The logic of the method was to gradually change encoding parameters to find the critical point at which quality became unacceptable.

We extended this study by categorising each shot of all the used clips according to the classification described by Thompson (Thompson, 1998).

Four different image sizes were examined to encompass a range typical of current mobile phones (see Table 1). The four image sizes were also chosen to represent roughly equal increments of pixel estate. We chose not to directly control for viewing distance. As with normal use, participants were free to adjust the viewing ratio (VR) of the different image resolutions to their individual preferences. The VR indicated in Table 1, however, is based on an average viewing distance of 40cm.

Table 1: Image sizes used on PDA

Screen area (mm^2)	Pixels (P)	P/mm^2	VR
(53 x 40) 2,120	(240 x 180) 43,200	20	10H
(46 x 34.5) 1,587	(208 x 156) 32,448	20	12H
(37 x 28) 1,036	(168 x 126) 21,268	20	14H
(26.5 x 20) 530	(120 x 90) 10,800	20	20H

The encoding bitrate is an important factor as the effect of image size/resolution might be different at different encoding bitrates. For example, when the bitrate available for the video content is scarce, reducing the image resolution could free up valuable encoding bitrate to improve the perceived quality. Similarly, when bitrate is abundant there may be less loss of detail as the image resolution is reduced.

Encoding bitrate was manipulated in two ways. Within a particular TV clip the bitrate allocated to video was degraded every 20 seconds by 32 kbps from a maximum of 224kbps down to 32kbps. These intervals are illustrated in

Table 2.

Table 2: 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

The boundaries of the intervals were not pointed out to the participants. They were simply presented with a continuous clip that gradually decreased in quality. In addition to changing the video bitrate within a clip, two duplicate sets of clips were produced with different bitrates allocated to the audio channel.

The *Low Audio* clips were coded at at 16kbps (Windows Media Audio V9) whereas the *High Audio* clips were coded at 32 kbps. Theses values were selected based on the results of previous studies on mobile devices in which participants' acceptability of 32bps audio compared to 16kbps audio had declined from 95% to 80% (McCarthy, Miras, & Knoche, 2004).

Material

Test material used for quality evaluation is usually selected from a video or audio test set. For example, VQEG uses a test set of 20 8-second clips (VQEG, 2000) to represent a range of difference types of motions, content and camera position. While such test sets are suitable for comparing performance differences between codecs, they are less useful in evaluating the perceived quality of service. In addition the clips are without audio and therefore not representative of the experience users would have with mobile TV. Mobile TV viewing is typically considerably longer than 8-10 seconds, and composed of a mixture of different motion, content and camera shots.

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. Bespoke editing takes time (which means access to topical content such as news is delayed) and is 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 or DVD material without any special editing steps. Clips of this type have been successfully used to examine quality tradeoffs for football coverage on mobile TV [9].

To understand the type and length of program people are likely to watch, we drew on two studies of mobile TV services (Knoche et al., 2004), (Södergård, 2003). These indicated that watching time was likely to be between 2 and 5 minutes, and that news was the most highly demanded content class by all user groups. Other content of interest to two different subgroups were *sports highlights* and *music videos*. As an additional category we included stop-frame animation (claymation) as a category. Animation can be very bandwidth efficient and is representative of the type of content delivered over low bandwidth networks (GPRS).

In total, four clips for each of the four content types were produced, giving us a total of 16 source clips. A summary of the clips is presented in Table 3.

Table 3: Used content types overview

Clip	Content Type	Description

N1-N4	News	BBC News 24 clips
S1-S4	Sport	Football World Cup 2002: Goal Highlights
M1-M4	Music	Clips directed by M. Gondry
A1-A4	Animation	Clips from “Creature Comforts”

The video clips were prepared as follows: We recorded footage from TV (BBC24 News) and from DVDs (2002 Fifa World Cup football, Creature Comforts animation, and Michael Gondry music videos). All extracted clips were chosen such that after 2:20min (or shortly thereafter), a story line would end. We used *Virtualdub* to segment these source clips into seven 20 second long clips at the different resolutions with a nominal frame rate of 12.5fps. These segments were encoded with Windows Media Encoder (WME) using the Microsoft Windows Media Video V8 codec with the different bitrates for the different segments as shown in
Table 2. Each group of seven WMV segment files were then converted and concatenated to one AVI file using TMPGEnc Express. Finally, these files were encoded using WME again to alter the audio encoding to either 32 or 16kpbs using 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 to the video quality of any of the segments.

Design

As shown in Table 4 we ran four groups, each comprising 32 participants. Each group was presented with 16 clips in total in groups of four clips at each of the four image resolutions. The groups differed in whether they experienced *Increasing* or *Decreasing* image resolutions and whether the audio quality was *High* or *Low*. Within each group, we also ran four variations to control for content using a Latin squares design such that the different content clips (e.g. N1-N4) were tested at each of the different image resolutions across participants.

Table 4: Experimental design

Group	**Audio**	**Res. Order**	**Image Resolution**	**Content Clip**			
A (32)	Good (32kbps)	Decreasing	240x180	N1	S1	M1	A1
			208x156	N2	S2	M2	A2
			168x126	N3	S3	M3	A3
			120x90	N4	S4	M4	A4
B (32)	Good (32kbps)	Increasing	120x90	N1	S1	M1	A1
			168x126	N2	S2	M2	A2
			208x156	N3	S3	M3	A3
			240x180	N4	S4	M4	A4
C (32)	Poor (16kbps)	Decreasing	240x180	N1	S1	M1	A1
			208x156	N2	S2	M2	A2
			168x126	N3	S3	M3	A3
			120x90	N4	S4	M4	A4
D (32)	Poor (16kbps)	Increasing	120x90	N1	S1	M1	A1
			168x126	N2	S2	M2	A2
			208x156	N3	S3	M3	A3
			240x180	N4	S4	M4	A4

The dependent variable was *Video Acceptability*. Independent variables were *Image Resolution*, *Content Types*, *Video Bitrate*, *Audio Bitrate.* Control variables were *Resolution Order*, *Sex*, *Native Speaker* and *Corrected Vision*. The variable *Corrected Vision* coded whether participants had uncorrected vision or wore contact lenses or glasses.

Equipment

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 colours and a resolution of 240x320. 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 software (2003) to embed the Windows Media Player. It presented the clips along with a volume control and two response buttons that allowed for toggling between acceptable and unacceptable quality. When the acceptable button was clicked the background of the application was green. In the unacceptable state the background was red.

Procedure

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 quality for watching different types of content. The instructions 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 short time to practice pressing the buttons on the display. When they were ready the experiment began and the participants watched 16 clips in succession.

During the session we recorded the participants' interactions with the devices on video. The video was later used to measure viewing distance at the different image resolutions. The participants' ratings, i.e. the taps on the 'Unacc.' and 'Acc.' buttons, were recorded on the device.

Participants

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 a total of 26 different countries. English was the first language for 72 of the participants.

4 RESULTS

The data were generated from the acceptability replies of the participants on a per second basis. For example, when a participant clicked unacceptable during the 35th second of a clip we marked all the previous seconds 1-34 acceptable and all the following seconds from 35 to 140 as unacceptable.

We analysed the obtained data using a binary logistic regression to test for main effects and interactions between the independent variables – *Image Resolution*, *Video Encoding Bitrate*, *Content Type, Shot Type* and *Audio Bitrate.* Control variables *Gender, Corrected Vision, Resolution Order* and *Native Speaker* were also included in this analysis.

The regression revealed significant effects of all of the control and independent variables. As expected, higher encoding bitrates resulted in higher acceptability. Increasing the resolution also increased the acceptability of the video quality. The acceptability of video quality of the different content types depended on both the resolution and encoding bitrates.

From hereon we limit our details to shot types. The results about the acceptability of the content types at the different resolutions and encoding bitrates can be found in (Knoche et al., 2005).

In a given content type we will only report the acceptability of shot types of which participants had watched at least a total of 40 seconds.

To illustrate the differences in shot type mixes we present the percentage at which a given shot type was used in the different content types in Figure 7. For example, roughly 50% of the football content was presented in extreme long shots, which were not used at all in the animation clips.

Shot type was a significant predictor of acceptability [$\chi^2(1)=330.5$, $P<0.001$]. Averaged across all content types, resolutions and encoding bitrates the acceptability of the long shot was the highest and the extreme long shot was the lowest.

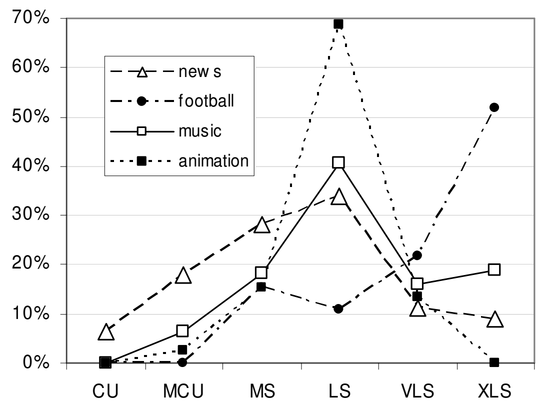

Figure 7: Distribution of shot type usage in experimental clips by content types

The regression also exposed an interaction of *Shot Type* with *Image Resolution.* Averaged across all content types all shot types became more acceptable with increasing resolution as depicted in Figure 8. Interestingly, it was not the two shot types showing the most detail (MCU and MS) that were the most acceptable across content types. At the three highest resolutions the long shot received the highest score in acceptability of video quality. The extreme long shot was the least acceptable shot at all resolutions when averaged across the content types.

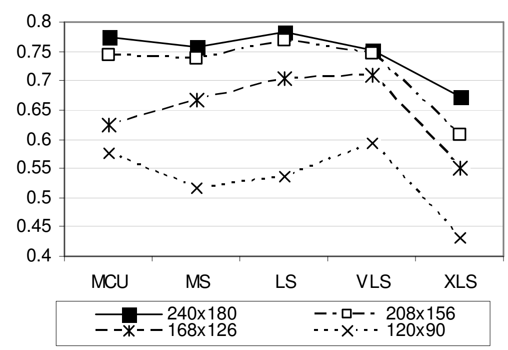

Figure 8: Acceptability of shot types at different resolutions

Furthermore, the regression revealed an interaction of *Shot Type* and *Content Type* [$\chi^2(1)=1337.1$, $P<0.001$]. We will address each content type in turn.

4.1 News

News had the biggest range of shot types in our experiment as can be derived from Figure 7.

Overall the shot type that yielded the highest acceptability of video quality across all resolutions was the medium shot. The long shot was the least acceptable across all sizes.

The differences in video acceptability for the different shot types were not overly pronounced apart from the fact that the long shot was by far the least acceptable at the two lower resolutions and the medium shot was the most acceptable at the higher resolutions.

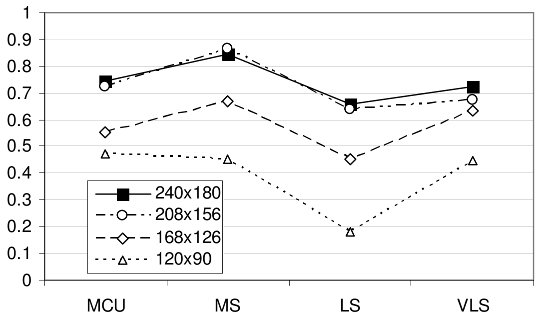

Figure 9: Acceptability of shot types of news content

4.2 Football

Shot types closer than a medium shot are not very common in football coverage. It is hard to zoom in on and follow players further because they often move in unpredictable ways.

Surprisingly, it was not the very wide shot with the least detail that was the least acceptable in the football clips. To the contrary, the shot type with the highest amount of detail - the mid shot - was the least acceptable in terms of the video quality across all resolutions (see Figure 10). At the highest resolution 240x180 the difference between the shot types was least pronounced but the extreme long shots were the most acceptable. Non-parametric tests showed that at 240x180 extreme long shots were significantly more acceptable than the medium shots.

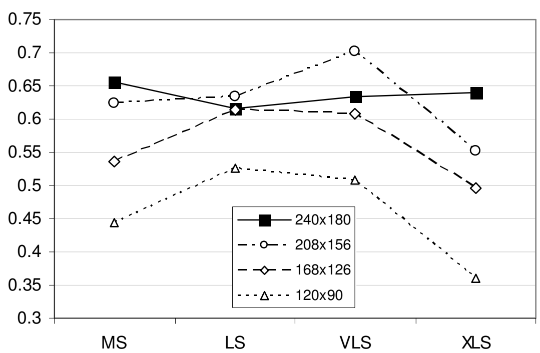

Figure 10: Acceptability of shot types of football content

4.3 Music

The differences in acceptability of the different shot types were immense and the most pronounced compared to the other content types.

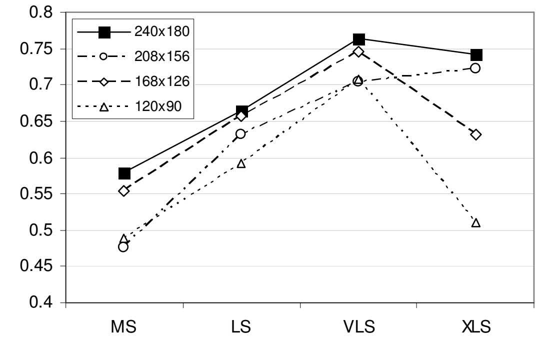

Figure 11: Acceptability of shot types of music content

As with football content the medium shot was least acceptable across all sizes in the music clips. The acceptability of the extreme long shot changed dramatically with the image resolution. At the smallest resolution its acceptability was the same as the least acceptable mid shot. Shown at the highest resolution, however, this shot was the most acceptable. Image resolution seemed to have little effect on the acceptability of the very long and mid shots. On average the very long shot was the most acceptable of the shot types for the music clips.

4.4 Animation

The animation content relied mainly on three shot types: VLS, LS and MS. Shots with more detail than MS are not desirable as the imperfections of the claymation process, e.g. fingerprints, might become visible.

In the fairly static animation content the medium shot presenting the most visual detail (MS) was the most acceptable. The long shot was slightly but significantly more acceptable than the very long shot across the resolutions [$Z=-5.2$, $P<.001$] (see Figure 12).

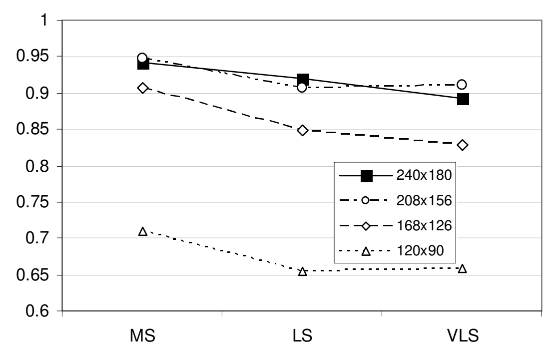

Figure 12: Acceptability of shot types of animation content

5 DISCUSSION

In football the extreme long shots were not the least acceptable. This is contrary to what content producers seem to believe. A possible explanation is that these shots are very central in understanding what is going on in a game. A large number of indicators of what was happening, e.g. the direction in which the camera pans, the direction in which the players are moving was available. Even the location of the ball could often be reconstructed from these cues. Visual detail might not be that important to derive this information. It could also be argued that the increased effort required on part of the participants to follow the game in these shots due to the low resolutions reduced their attention to rate the acceptability of the video quality. Whatever the reason for this might be, it should be kept in mind that people were watching only relatively short clips of highlights that have a different shot mix than live games. In longer footage, e.g. live games, people might rate the shot worse over time due to fatigue.

The mid shot received the worst ratings of all shot types both in football and music clips. Compared to the animation and news clips both of the former had many camera pans with moving background. For example, a football player is usually not static in this shot type. But camera pans were also used in other shot types both in football and music clips. One possible explanation is that in the medium shots the lack of detail due to the low resolutions is most apparent. The unmet expectations of what should be visible in this kind of shot might also be responsible for low acceptability ratings. The importance of visual detail had also been shown in a study by McCarthy et al., which found that visual detail was more important in football coverage than a smooth frame rate (McCarthy, Sasse, & Miras, 2004).

If we consider the visual content of the news clips to be the closest to soap operas we also find no reason to use close-up or medium-close up shots instead of, for example, medium shots. The latter allows for more body language to be presented in a frame.

6 CONCLUSIONS

From the results at hand we cannot back the mentioned quick fixes for low resolution and small screen mobile content. Tailor-made content for mobile TV might be more enjoyable as a whole when prepared without extreme long shots for football and with heavy use of close-ups for mobile soaps. However, these shot types did not affect the acceptability of the video quality in the expected manner.

Medium shots were more acceptable than the medium close-ups in news coverage. Extreme long shots that are deemed to be problematic for football coverage were much more acceptable than the medium shots that are used more often in football highlights.

Clearly, the results at hand warrant more research in this area to aide mobile content producers in making informed choices in this novel area of multimedia consumptions.

There were a few limitations to this study. First, the experimental setup was not specifically designed for the analysis of shot types. Therefore, shot type occurrences were not counterbalanced and exposed to all encoding bitrates. Furthermore, the overall Quality of Experience of a mobile TV service might differ from the acceptability of the video quality. Older people might have different requirements. On average our study focused on a fairly young study population. Another recent study found that text legibility has a significant effect on video quality perception (Knoche et al., 2006) for which we did not control. Last but not least, data loss – a relevant problem for broadcast services – was not considered.

7 FUTURE WORK

We would like to compare the same content produced for regular TV with its counterpart tailor-made for mobile consumption measuring the level of enjoyment derived from the two competing formats.

Having a tailor-made mix of shot types might be more important in order to enjoy and/or understand content than optimizing for video quality.

REFERENCES

Odyssey software inc. CFCOM (2003). http://www.odysseysoftware.com/ [On-line].

Ankrum, D. R. (1996). Viewing Distance at Computer Workstations. *Work Place Ergonomics,* 10-12.

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

Guardian (2005, June 28). Romantic drama in China soap opera only for mobile phones. *Guardian Newspapers Limited.*.

Gwinn, E. and Hughlett, M. (2005, October 10). Mobile TV for your cell phone. *Chicago Tribune*.

Horn, D. B. (2002). The effects of spatial and temporal video distortion on lie detection performance. In *The Association for Computing Machinery conference on Human Factors in Computer Systems*.

Jain, R. (2004). Quality of Experience. *IEEE Multimedia, 11,* 95-96.

Kies, J. K., Williges, R. C., & Rosson, M. B. (1996). *Controlled Laboratory Experimentation and Field Study Evaluation of Video Conference for Distance Learning Applications* (Rep. No. HCIL 96-02). Virginia Tech.

Knoche, H. & McCarthy, J. (2004). Mobile Users' Needs and Expectations of Future Multimedia Services. In *WWRF12*.

Knoche, H., McCarthy, J., & Sasse, M. A. (2005). Can Small Be Beautiful? Assessing Image Resolution Requirements for Mobile TV. In *ACM Multimedia* (pp. 829-838). ACM.

Knoche, H. & Sasse, M. A. (2006). Breaking the News on Mobile TV: User requirements of a popular mobile content. In *IS&T/SPIE Symposium on Electronic Imaging*.

Lombard, M., Grabe, M. E., Reich, R. D., Campanella, C., & Ditton, T. B. (1996). Screen Size and viewer responses to television: A review of research. In *Annual conference of the Association for Education in Journalism and Mass Communication*.

McCarthy, J., Sasse, M. A., & Miras, D. (2004). Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video. In *Conference on Human Factors in Computing Systems CHI '04* (pp. 535-542).

McCarthy, J., Miras, D., & Knoche, H. (2004). *TN01-1.1.03_UCL_MAESTRO_bandwidth_study_V02*.

Mediacollege (2006). http://www.mediacollege.com/video/shots/ [On-line].

Okada, K.-I., Maeda, F., Ichikawaa, Y., & Matsushita, Y. (1994). Multiparty videoconferencing at virtual social distance: MAJIC design. In *Computer Supported Cooperative Work* (pp. 385-393).

Owens, D. A. & Wolfe-Kelly, K. (1987). Near Work, Visual Fatigue, and Variations of Oculomotor Tonus. *Investigative Ophthalmology and Visual Science, 28,* 743-749.

Reeves, B., Lang, A., Kim, E., & Tartar, D. (1999). The effects of screen size and message content on attention and arousal. *Media Psychology, 1,* 49-68.

Reeves, B. & Nass, C. (1998). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. University of Chicago Press.

Södergård, C. (2003). *Mobile television - technology and user experiences Report on the Mobile-TV project* (Rep. No. P506). VTT Information Technology.

Thompson, R. (1998). *Grammar of the shot*. Elsevier Focal Press.

VQEG (2000). *Final Report from the video quality experts group on the validation of objective models of video quality assessment* (Rep. No. http://www.vqeg.org).

Weiner, R. (1996). *Webster's New World Dictionary of Media and Communications*. (rev. and updated ed.) New York, NY: Macmillan.