How Low Can You Go?
The Effect of Low Resolutions on Shot Types in Mobile TV

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

Abstract

The advent of mobile TV which is often viewed on small screens with low resolution has made TV content producers think about refraining from using shots that depict subjects from a great distance. Shot types where the object of interest fills the screen are deemed to be more appropriate for mobile devices. This paper reports a study on how shot types used in regular broadcast television are affected when shown on mobile devices at reduced levels of resolution. 72 native speakers judged the acceptability of four different content types at four resolutions (240x180, 208x156, 168x126, 120x90) across seven encoding bitrates. The results show that acceptability of shot types depends on the content and the resolution. Extreme long shots of football content were only less acceptable than other shot types at resolutions smaller than 240x180. The medium shot which portrays the upper half of a subject's body was the most acceptable for news content but for football content was judged worse than shot types that showed less detail. Our results suggest that for a young audience extreme long shots may be used with no detrimental effect for resolutions of 240x180 and higher. At lower resolutions and for content with a high degree of dynamism both the medium shot and the extreme long shot might render poorly for the audience. Service providers are well advised to include the results at hand to customise content in terms of shot type use for their audience that will watch the content at very low resolutions. Further research should assess older audiences and the effectiveness of cropping schemes that zoom in on part of the content for low target resolutions.

1 Introduction

There are many services that aim to provide users with a TV-like experience while on the move. The Quality of Experience (QoE) of mobile TV depends on the perceived audio-visual quality of the consumed content 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, specifically created for the mobile, to material that is produced for standard TV or cinema consumption. Original TV material may undergo an additional editing process to prepare it for mobile consumption. Producers of tailor-made mobile content are trying to come up with a mix of shot types to optimize the viewing experience on small low resolution screens. The sports network ESPN, for example, is considering resizing graphics for the small screen and minimizing the use of [extreme] long shots in their coverage (Gwinn & Hughlett, 2005). However, manual editing is costly and it is faster and cheaper for service providers to directly encode and deliver existing broadcast material without additional editing.

One of the central factors of the visual quality of mobile TV content is the spatial resolution of the image which matters to all actors involved in the field of mobile TV:

Device manufacturers: Mobile device displays come in a range of shapes, sizes and resolutions, from VGA PDAs (480x640 pixels) and high end 3G or DVB-H enabled phones (320x240) to more compact models with QCIF size (176x144).

Users: Previous research has shown 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'. On a display of 8cm height mobile users could actually perceive the difference between TV content at standard resolution and high definition (HD) if the device could display HD resolution. 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), (Song, Won, & Song, 2004).

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 content influences the directors' decision as to which shot types to use while shooting the footage. The camera shots used in television range from very wide shots (VWS) to extreme close-ups (XCU) and consider image size, resolution and possible maladjustments of typical TV setups (Weiner, 1996). Image size and resolution cannot be reduced indefinitely as important detail will be lost and shot types might be affected differently.

So far, if and how shot types affect the perceived visual quality of mobile TV content at the low resolutions and encoding bitrates used in current mobile TV services has been unknown. However, previous research reported that participants complain about the lack of detail in certain shot types (e.g. extreme long shots in football) or that they cannot identify people or objects when presented on small displays (Knoche et al., 2005). To shed some light on this topic we classified all the 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 data set extended by the shot type classification.

Section 2 presents the background on human visual acuity and the effects and interdependencies of viewing distance, image size and image resolution based on previous research in the field as well as a common classification scheme for shot types. We describe the original study on image resolution in Section 4 and present the results of our shot type classification, which are discussed in Section 5. 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. Therefore we reviewed the previous research on human visual perception with respect to viewing distance, picture size and resolution and how the size and resolution of video content influences the audience's perception which in turn is limited and influenced by their visual acuity.

Even though resolution, viewing distance and picture size are not independent of one another and should all be considered during analysis previous research has identified a number of limiting factors for each of them which are presented in the following subsections.

2.1 Visual acuity

The ability to resolve detail at different distances is determined by people's visual acuity. Ophthalmologists distinguish between three types of visual acuity: minimum visible acuity, minimum resolvable (ordinary) acuity, and minimum discriminable acuity (hyperacuity) (Westheimer, 1992). Most frequently used within the engineering literature is minimum resolvable (ordinary) acuity. This is determined by peoples' ability to identify a target – such as whether a letter is a C or an O. – and depends on identifying the presence of a gap or feature in the letter. By varying the object size one can determine the minimum resolvable threshold. Normal 20/20 vision is classified as the ability to resolve 1 minute of arc.

Research on human resolving power on TV display is often determined using sets of alternating black and white lines of equal width. One black/white line pair represents one cycle which in pixel based displays would require two pixels i.e. two columns with a width of one pixel. The number of cycles that can be resolved across one degree of the eye's viewing field is typically used as a measure of human visual acuity, and is stated in cycles (line pairs) per degree. Campbell and Green found that the maximum resolution of the retina is about 60 cycles per degree (Campbell & Green, 1965). In practice, research in TV imaging has shown that approximately 22 cycles (44 pixels) per degree is perceived as a sharp image (Silbergleid & Pescatore, 2000).

2.2 Resolution

Decisions about resolution occur at several times in the process of creation, editing, delivery and presentation of visual content. At the content creation stage the producers have to decide which resolution should be used. The delivery of high resolution content demands more resources and therefore service providers need to find a trade off between the added visual quality and the additional cost or reduction in the amount of content that can be delivered.

For example, we can predict that reducing the image resolution can have two opposing effects:

1. A smaller image resolution will give bitrate savings as there is less information to be coded. Thus, for a fixed encoding bitrate, it is possible that the perceived quality is increased as the bandwidth budget per pixel is increased when the image resolution is reduced. This is of course dependent on the efficiency and overhead of the codec used to encode the content. So far research has not provided an answer to this question. A study by Knoche et al. did not reveal any interaction of encoding bandwidth and picture resolution (Knoche et al., 2005) within the parameter range used in the study at hand.

2. As image resolution is reduced, there are fewer pixels to represent information of importance to the user. This may cause problems with some content types – such as sport – as there are very few screen pixels available to display important details such as the location of the ball. Research in face recognition has shown that human observers require at least 15 pixels per face (in vertical resolution) in order to be able to identify faces (Bachmann, 1991; Bathia, Lakshminarayanan, Samal, & Welland, 1995). If identifying people is of concern to viewers violating this requirement might affect the

perceived visual quality. Thus, for a fixed bitrate it is possible that perceived quality is decreased when image resolution is reduced. These problems have been noted in previous research where participants complained about their inability to identify A (Knoche et al., 2005).

At the presentation stage the capabilities of the end user equipment determines the resolution at which content can be presented.

2.3 Viewing distance

Mobile devices are operated at 'arm's length'; continued viewing at distances closer than the resting point of vergence – approx. 90cm, with a 30° 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.

Viewing distance is often expressed relatively to the picture height. A viewing distance of 5H, for example denotes that the distance between the viewer and the screen is five times the height of the screen. The size of the display in the viewer's visual field depends on both the viewing distance and the size of the screen. The viewing ratio (VR) is defined as the viewing distance divided by the picture height (H).

For a given combination of picture height and resolution of a presented picture increasing the viewing distance has two opposing effects with respect to the perceived picture quality. The negative effect on the perceived quality is due to the fact that the picture angle becomes smaller in the eye of the observer. At the same time, however, the angular resolution of the pictures increases and thus improves the quality, as long as the observers are not at their visual acuity threshold.

Jesty found evidence for an optimal viewing distance. When faced with the decision of placing a chair to view projected pictures with a fixed size, observers chose their viewing distance in a way that depended only on the resolution of the picture. The quotient of picture height and optimal viewing distance was constant for a given resolution (Jesty, 1958).

Findings by Westerink et al. confirmed the existence of an optimal viewing distance and showed that at constant viewing distance subjective picture quality of still pictures was influenced both by the resolution of the pictures and their width (Westerink & Roufs, 1989). The optimal viewing distance of still pictures was chosen such that the resolution equalled 16 cycles per degree independent of the picture width. For pixel based displays this would translate to 32 pixels per degree. This indicates that the gains in perceived visual quality from achieving a higher visual resolution beyond 16 cycles per degree are not big enough to compensate for the reduction in picture angle.

2.4 Image size

Research on the effects of image size was grounded to a large degree on detection tasks dating back to studies by e.g. Steedman and Baker (1960) which were based on still pictures. Presenting moving images in constrained spatial settings has been an area of research for a long time.

Most of the research on the effect of screen sizes in the field of consumer electronics 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 generally preferred to smaller ones. In tele-presence e.g. video-conferencing setting participants prefer big displays depicting up to life-size pictures (Okada, Maeda, Ichikawaa, Matsushita 1994).

Other studies involving moving images as part of a video conferencing systems showed 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 as large a screen 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 content shown on mobile devices at higher resolutions is generally more acceptable than lower resolutions at identical encoding bitrates. However, the differences were not uniform across content types (Knoche et al., 2005). All content types received poor results when presented at resolutions smaller than 168x126.

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). That shot type is equivalent to a long shot as defined in Sec. 2.6.

2.5 Possible enhancements

There are a number of content based pre-encodings that could improve the presentation of standard TV content to mobile users by:

- Cropping off the surrounding area of the footage that is outside the final safe area for action and titles and does not include essential information. The broadcast material includes this to compensate for maladjustment of TV receivers (Thompson, 1998).
- Zooming in on the area displaying the most important aspects (Dal Lago, 2006), (Holmstrom, 2003).
- Visually enhancing content, e.g. by sharpening the colour of the ball in football content (Nemethova, Zahumensky, & Rupp, 2004).

However, all of these possible improvements lack subjective testing on mobile devices. Furthermore, it is unknown how much zoom or cropping is advisable for which target resolutions and for which shot types these schemes would prove beneficial.

2.6 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.

Unfortunately, the terms used to classify shot types can differ and popular usage of the terms deviates further. For consistency we will use the classification from (Thompson, 1998) which is presented below (see Figure 1-6).

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.

2.6.1 Extreme long shot (XLS)

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).

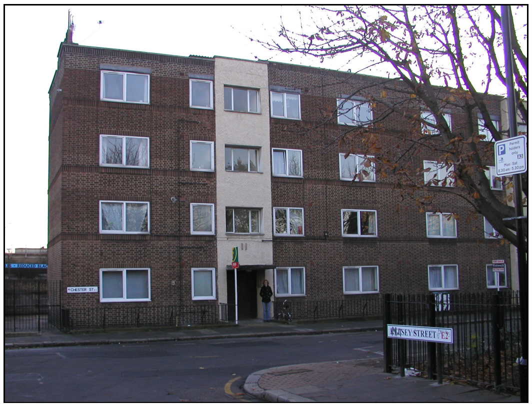

Figure 1: Extreme long shot (XLS)

2.6.2 Very long shot (VLS)

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).

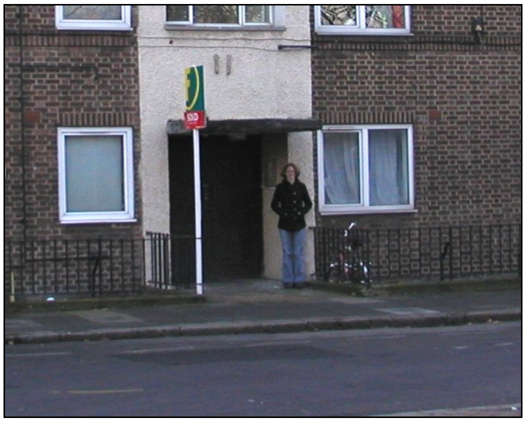

Figure 2: Very long shot (VLS)

2.6.3 Long shot (LS)

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

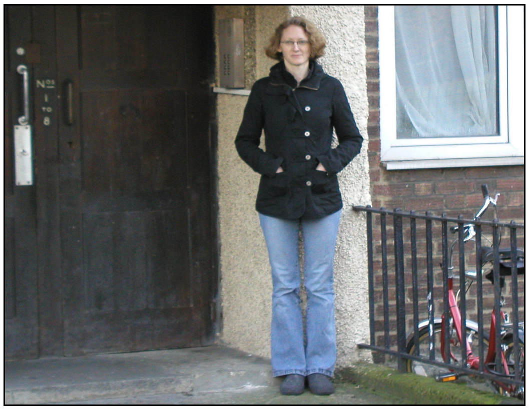

Figure 3: Long shot (LS)

2.6.4 Medium shot (MS)

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.

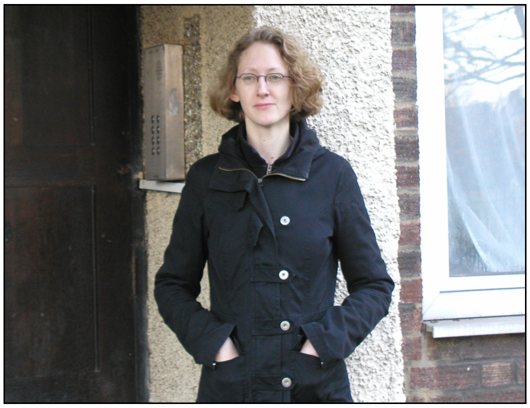

Figure 4: Medium shot (MS)

2.6.5 Medium close-up (MCU)

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

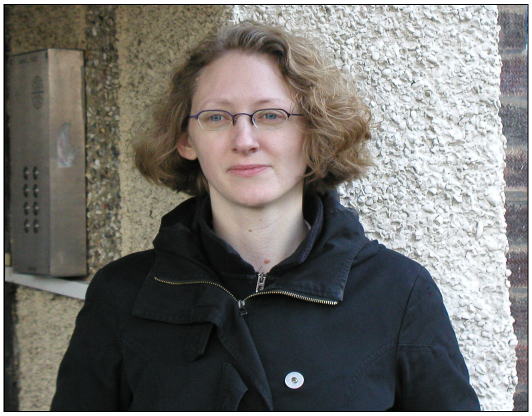

Figure 5: Medium close-up (MCU)

2.6.6 Close-up (CU)

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

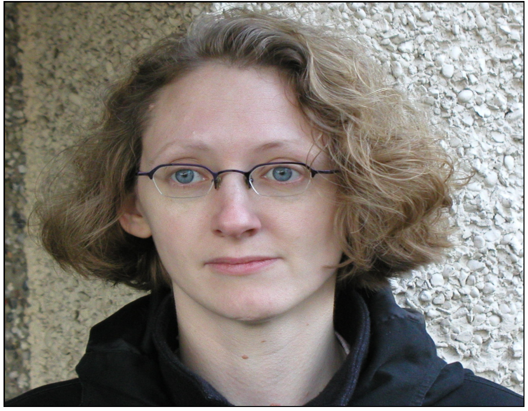

Figure 6: Close-up (CU)

We limit our study to shot types that were most common in the footage used in this study. The presented pictures (Figures 1-6) were not part of the footage used in this study but are representative of the shot types that made up the content.

3 Shot Type Study

TV and cinema content use a mix of shot types with varying lengths. Creating a fully counterbalanced set of stimuli with real content clips is therefore hard to achieve. We decided to drop this requirement for this initial study and classified each shot of video clips of an existing study according to Thompson's shot type classification described in Sec. 2. We drew upon a recent study (Knoche et al., 2005) that employed clips of different content types (news, music, sports and animation) at different resolutions (240x180, 208x156, 168x126 and 120x90) and of considerable length (2:20min). The clips gracefully degraded in quality by a reduction in encoding bitrate from 224kbps down to 32kbps. The content was not manipulated further. The original aim of the study 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.

The different resolutions resulted in four different image sizes on the mobile device (Knoche et al., 2005) to mimic a range typical of current mobile phone display sizes (see Table 1). The study did not directly control for viewing distance. As with normal use, participants were free to adjust the viewing distance to their individual preferences. The viewing ratios (VR) of the different image resolutions indicated in Table 1, however, are based on an average viewing distance of 40cm and expressed in multiples of the picture height.

The iPAQ 2210 used in the study had a physical screen height of 73mm and a vertical resolution of 320 pixels. At a viewing distance of 40cm, the screen vertically subtends visual a visual angles of 10.4°. This translates to a resolution of approximately 15 cycles per degree, which is classified as low to normal resolution in TV terms. Assuming a constant viewing distance this setup results in a constant angular resolution of the different video clip resolutions for the viewer.

In the study we used only one kind of presentation device. This kept the resolution of the display fixed, but we varied the resolution of the video clips and displayed them at their native resolution. In other words the smaller resolution video clips were represented by fewer pixels which resulted in different physical sizes of the video images on the device. However, the participants could freely adjust the viewing distance to the device such that the pixels per degree can be changed according to their preferences.

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	10
(46 x 34.5) 1,587	(208 x 156) 32,448	20	12
(37 x 28) 1,036	(168 x 126) 21,268	20	14
(26.5 x 20) 530	(120 x 90) 10,800	20	20

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 summarised 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, we produced two duplicate sets of clips with different bitrates allocated to the audio channel. The *Low Audio* clips were coded at 16kbps (Windows Media Audio V9) whereas the *High Audio* clips were coded at 32 kbps.

Material

Some mobile TV services employ 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 favour immediate re-use of TV material. These editing rules are not based on empirical research so far but based on expert opinions in the best case. For the purposes of this study, we investigated the acceptability of directly recorded TV or DVD material without any special editing steps to see how the different shot types would be affected by the different encoding settings.

Previous studies of mobile TV services e.g. (Södergård, 2003) indicated that watching time was likely to be between 2 and 5 minutes and news was a highly demanded content type (Knoche et al., 2004). 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. At a typical viewing distance of 40cm this results in an angular resolution of approximately 15cycles/degree at – classified as low to normal resolution in TV terms (Silbergleid et al., 2000).

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 labelled "ACC." and "UNACC." that allowed for toggling between states that indicated 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.

Methodology and procedure

The methodology used in this study was originally introduced in (McCarthy, Sasse, & Miras, 2004). It has been successfully used in a number of studies. The required rating effort for the participants is minimal as there are few interruptions and the act of rating hardly interferes with the activity of watching TV on the mobile device. It provides results that can be translated into utility curves for service providers.

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.

The participants' ratings, i.e. the taps on the 'Unacc.' and 'Acc.' buttons, were recorded on the device.

There are two possible caveats with the approach at hand that need to be addressed. First, due to the use of the method of limits the experimental design did not present all parts of the video clips at all encoding bitrates. Consequently, the average encoding bitrate at which shot types were encoded were not identical. Second, many video encoders compress e.g. low motion video clips better than clips that include a lot of motion. Some shot types might contain more motion on average than others and therefore look better after encoding in terms of visual quality, e.g. sharpness. Thus even if the shot types had been encoded at identical average encoding bitrates would have not guaranteed equal visual quality of the shot types after encoding.

To control for both the differences in encoding bitrate as well as possible correlations between shot types and encoder performance we used the objective quality measure peak signal-to-noise ratio (PSNR) to obtain a rough estimate of the content's visual quality. We rescaled all degraded clips up to the resolution of the original clips and employed Avisynth's built-in PSNR compare function to compute the degradation of these encoded clips in comparison to their originals (Avisynth, 2005). Since we compared up-scaled versions of the low resolution clips with the reference clip we can expect that the lower resolution clips will in general yield lower PSNR scores. For example a clip with a resolution of 120x90 would be up-scaled by a factor of about four which will result in higher peak signal-to-noise ratio than a clip up-scaled from 240x180 by a factor of two. We only used the PSNR scores as indicators of visual quality between the shot types in clips of the same resolution. We will present the obtained PSNR values of the different shot types for the different content types in Sec. 4.

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, if a participant had been in the unacceptable state during a second it was marked 'unacceptable' for this participant. We decided to exclude all ratings in the three seconds following a scene change to allow for participants' adjustment to the new picture. In doing so we excluded shots that lasted less than three seconds. In addition to the variables analysed in the original study we included *Shot Type* as an independent and *Native Speaker* as a control variable. The latter variable denoted native English speakers.

We analysed the 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 variable *Corrected Vision* indicated whether participants had uncorrected vision or wore contact lenses or glasses.

The regression revealed significant effects of all of the control and independent variables as in (Knoche et al., 2005). Non-native English speakers were less likely to rate the quality of a clip unacceptable than the native English speakers. We excluded the data from the non-native speakers and repeated the regression. All results we present from here on are based on the 72 native speakers that took part in the study.

As expected, higher encoding bitrates and higher resolutions 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. The detailed results about the acceptability of the content types at the different resolutions and encoding bitrates can be found in (Knoche et al., 2005).

In this paper we limit our analysis to shot types. For a given content type we report only the acceptability scores of shot types that each participant had watched for a total of at least 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.

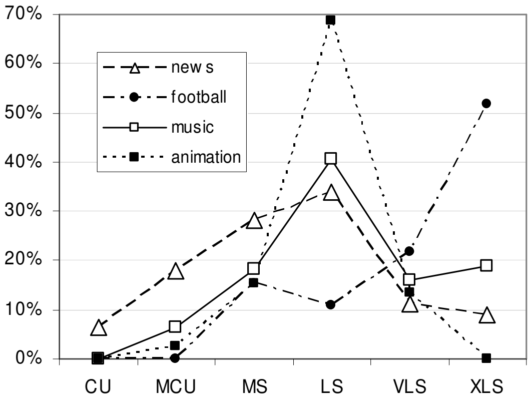

Figure 7: Distribution of shot type usage in experimental clips by content types

Shot type was a significant predictor of acceptability [$\chi^2(1)$=148.4, P<0.001]. Averaged across all content types, resolutions and encoding bitrates the close up and the very long shot were the most acceptable shot types. The extreme long shot (XLS) received the lowest ratings.

All shot types became more acceptable with increased resolutions see (Figure 8). The extreme long shot was by far the least acceptable shot at all resolutions when averaged across the content types.

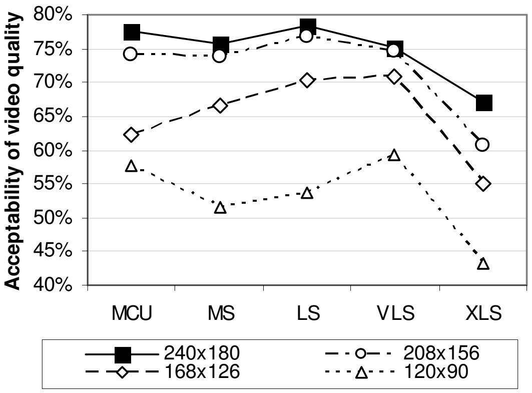

Figure 8: Acceptability of shot types at different resolutions

Furthermore, the regression revealed an interaction of *Shot Type* and *Content Type* [$\chi^2(1)$=1337.1, P<0.001]. In Figure 9 we present the acceptability scores of the different shot types by content type averaged across the four different resolutions and all encoding bitrates.

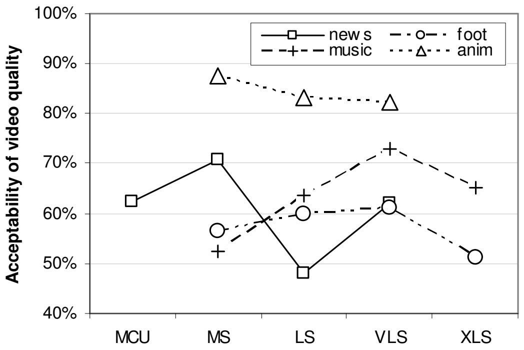

Figure 9: Acceptability of shot types by content type

We will subsequently address each content type in turn. For each resolution the acceptability scores of the shot type are averaged across all encoding bitrates. The figures below present these values with standard error bars based on the participants' acceptability averages in these conditions.

4.1 News

News content is made up of a mixture of different material and therefore had the biggest range of shot types in our experiment as can be seen in Figure 7. Typically the anchorman announced a topic that was then covered in more detail by means of field reports, graphs, illustrations or interviews. The field reports used a wide variety of shot types to depict the topic and to situate the audience. The video quality of the field reports was usually worse than the footage shot in the studio.

The shot type that yielded the highest acceptability of video quality across all resolutions was the MS. One must keep in mind that this shot is typically used when presenting the anchor man in a static posture. The LS was the least acceptable shot type across all sizes. The acceptability of the video quality of the shot types at the two highest resolutions did not differ significantly; Mann-Whitney [Z=-1.7, n.s.]. The acceptability of the different shot types is summarised in Figure 10.

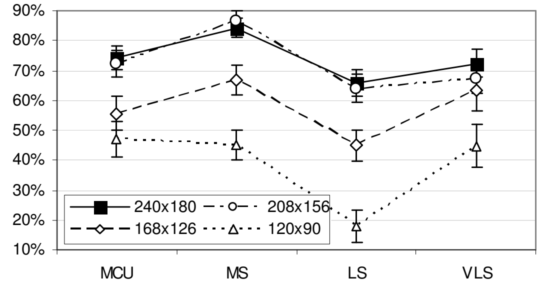

Figure 10: Acceptability of shot types of news content

The PSNR values of the different shot types presented in Figure 11 looked very similar to the acceptability scores. The values for the MCU and VLS were about the same and the MS was slightly above and the LS slightly below in value. This provided evidence that for the news content the differences in acceptability between the shot types were merely due to differences in visual quality.

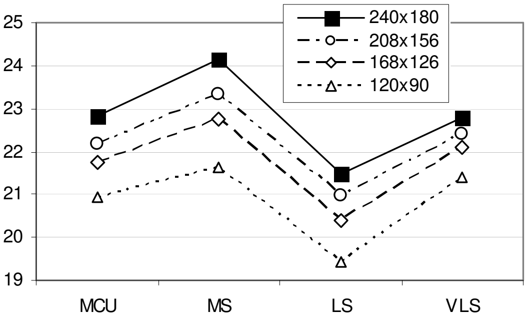

Figure 11: PSNR scores of news content at different resolutions by shot types

4.2 Football

Almost all of the scenes in the football footage depicted players in motion or camera pans of the pitch. Thus motion

Shot types closer than a medium shot are not common in football coverage. It is hard to zoom in on and follow players because they often move in unpredictable ways. The extreme long shot provides the viewer with an overview of what is going on in the playing-field. It is very popular and even in the highlights material used in the study this shot was used approximately 50% of the time.

Non-parametric tests showed that there was no significant difference in acceptability of the extreme long shot at the highest resolution when compared to the other shot types [$\chi^2(3)=2.34$, *n.s.*]. However, at all resolutions lower than 240x180 the results confirm the qualitative feedback about the extreme long shots in (Knoche et al., 2005). Here the extreme long shot was the least acceptable shot type.

Surprisingly, the acceptability of the medium shot depicting the greatest amount of detail in the football material declined much more than the long and the very long shot at lower resolutions (see Figure 12).

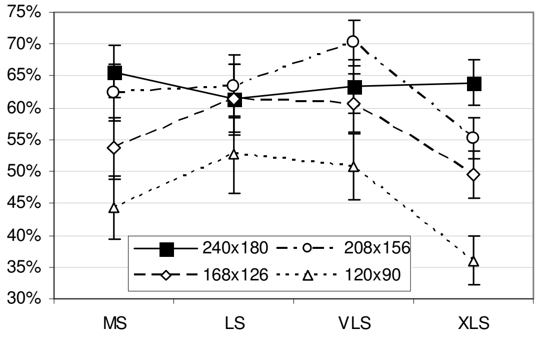

Figure 12: Acceptability of shot types of football content

In the computed PSNR values depicted in Figure 13 we find no evidence that the lower acceptability of MS and XLS might be induced by lower visual quality as was argued for the news content earlier. Both the MS and the XLS yielded considerably higher PSNR values in comparison to the LS and VLS.

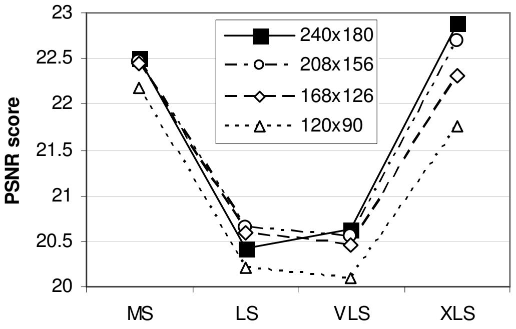

Figure 13: PSNR scores of football content at different resolutions by shot types

4.3 Music

The visuals of the music clips were dynamic with many camera pans. Across all resolutions the medium shot was the least acceptable and the very long shot the most acceptable in the music clips. The acceptability of the less detailed shots (LS and VLS) increased with a corresponding decrease in the level of detail. The acceptability of the extreme long shot changed dramatically with different image resolutions. At the smallest resolution its acceptability was only slightly above but not significantly different from the least acceptable medium shot. At the highest resolution, however, it was only slightly below and not significantly different from the most acceptable shot type – the very long shot (see Figure 14).

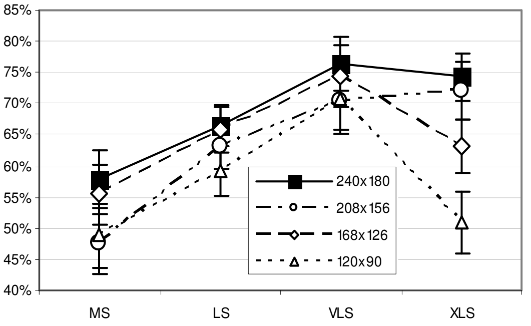

Figure 14: Acceptability of shot types of music content

Apart from the XLS image resolution seemed to have little effect on the acceptability of the more detailed shots. We could neither explain the VLS's high acceptability across all resolutions nor the XLS's reduction in acceptability at lower resolutions with just differences in visual. We can see in Figure 15 that the VLS had the lowest PSNR scores of all shot types and they were close to the PSNR scores of the MS. Despite the low PSNR scores the acceptability of the VLS was the highest of all shot types for the music clips. The PSNR values provide no indication of the degradation of the XLS at lower resolution that was evident from the acceptability scores.

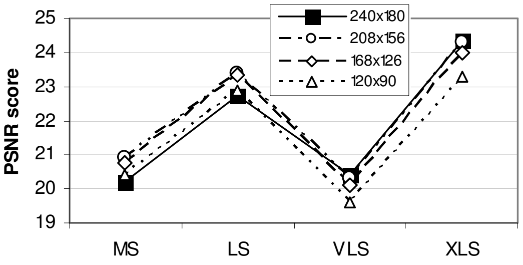

Figure 15: PSNR scores of music content at different resolutions by shot types

4.4 Animation

The claymation creature comforts material relied mainly on three shot types: VLS, LS and MS. Shots with more detail than the medium shot are possibly not desirable as the imperfections of the claymation process, e.g. fingerprints, might become more visible. The animation content depicted fairly static scenes with few camera pans. Of all content types this was the easiest for the encoder to encode as can be derived from the PSNR scores, which are the highest of all the four content types (see Figure 17). In the fairly static animation content the medium shot presenting the most visual detail (MS) was the most acceptable. There were no significant differences between the long and very long shot in terms of acceptability.

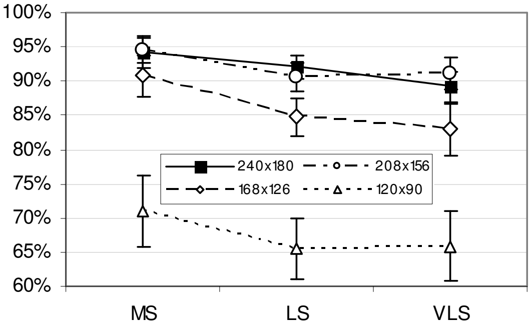

Figure 16: Acceptability of shot types of animation content

The PSNR scores for the shot types of animation content depicted in Figure 17 showed that the visual quality of the MS was the best and of the LS was the worst. The scores of the VLS lay between these two. The PSNR quality differences between the LS and the VLS were not reflected in the subjective acceptability values presented in Figure 16 where LS and VLS were almost the same.

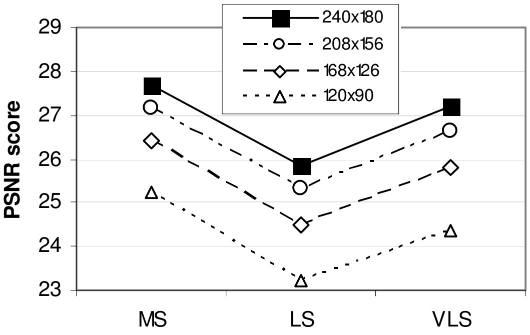

Figure 17: PSNR scores of animation content at different resolutions by shot types

5 DISCUSSION

The acceptability of the extreme long shot declined most at resolutions of 208x156 and lower in both music and football content. The acceptability of the very long shot, which shows a little more detail than the extreme long shot, was not degraded as much by these lower resolutions. This is encouraging news for intelligent cropping approaches (Dal Lago, 2006), (Holmstrom, 2003) that zoom in on part of the footage. Cropping brings the depicted content of an extreme long shot closer to what is seen in a very long shot, which had a much higher acceptability at all resolutions lower than 240x180. More research is required to evaluate the potential benefits of cropping for resolutions of 240x180 and higher, e.g. 320x240 that will be supported by DVB-H (ETSI, 2005).

The medium shot received the worst ratings of all shot types in the music clips. In the football clips only the extreme long shots received worse ratings.

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 and low encoding bitrates 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 noted in (McCarthy et al., 2004) which found that visual detail was more important in football coverage than a smooth frame rate. Our results could be interpreted a way that when producing for resolutions below 240x180 that content producers should not favour the medium shot over other shot types when the subjects are in motion.

If we consider the visual content of the news clips to be the most similar to soap operas we found no reason to use medium close-up shots instead of medium shots, for example. The medium shot allows for more body language to be presented in a frame and was not significantly worse but at most resolutions more acceptable than the medium close-up shot.

6 CONCLUSIONS

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, we cannot generally support these adaptations for mobile consumption from our results.

The medium shots that are used frequently in football highlights appear to be more sensitive to degradations due to low resolutions than some of the shot types with less detail. Extreme long shots in football coverage were not significantly less acceptable than more detailed shot types at a resolution of 240x180. At lower resolutions this shot type might benefit from cropping off the safe area or intelligent cropping, which would show a part of the screen in more detail. Clearly, the results at hand warrant more research that could control for movement and other possible covariates of shot types. More insight will 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 not equally exposed to all encoding bitrates. Furthermore, the overall Quality of Experience of a mobile TV service might differ from the mere acceptability of the video quality, i.e. despite low acceptability ratings shot types might be important to understanding of the content. Interactive TV and games like content might have very different requirements from passively consumed content. On average our study focused on a fairly young population. Older viewers with less focussing power (accommodation) might have different requirements as they compensate this deficiency by holding e.g. newspapers at a greater distance. Data loss – a relevant problem for broadcast services and on the perceived visual quality (Jumisko-Pyykkö, Kumar, Liinasuo, & Hannuksela, 2006) – 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 perceived video quality alone.

Intelligent cropping mechanisms that present an enlarged part of the original image are another promising improvement for mobile TV that we would like to explore.

REFERENCES

Mediacollege (2006). http://www.mediacollege.com/video/shots/

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

Gwinn, E. & Hughlett, M. (2005). Mobile TV for your cell phone. Chicago Tribune Available: http://home.hamptonroads.com/stories/story.cfm?story=93423&ran=38197

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

Song, S., Won, Y., & Song, I. (2004). Empirical Study of User Perception Behavior for Mobile Streaming. In *Proceedings of the tenth ACM international conference on Multimedia* (pp. 327-330). New York, NY, USA: ACM Press.

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

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

Westheimer, G. (1992). Visual acuity. In W.M.Hart (Ed.), *Adler's Physiology of the Eye: Clinical Application* (9th ed., St. Louis, Mo: CV Mosby.

Campbell, F. W. & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology, 181,* 576-593.

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

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

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

Jesty, L. C. (1958). The Relation between Picture Size, Viewing Distance and Picture Quality. In *Proceedings of IEE* (pp. 425-439).

Westerink, J. H. & Roufs, J. A. (1989). Subjective Image Quality as a Function of Viewing Distance, Resolution, and Picture Size. *SMPTE Journal*.

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.

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 Conf.of the Assoc.for Education in Journalism and Mass Communication*.

Okada, K.-I., Maeda, F., Ichikawaa, Y., & Matsushita, Y. (1994). Multiparty videoconferencing at virtual social distance: MAJIC design. In *Proc.ACM conf.on Computer supported cooperative work* (pp. 385-393).

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.

Horn, D. B. (2002). The effects of spatial and temporal video distortion on lie detection performance. In *Proceedings of CHI '02*.

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.

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

Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many
pixels does it take to carry identity? *European Journal of Cognitive Psychology, 3,* 87-103.

Bathia, S., Lakshminarayanan, V., Samal, A., & Welland, G. V. (1995). Human face perception in degraded images. *Journal of Visual Communication and Image Representation, 6,* 280-295.

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

Dal Lago, G. (2006). Microdisplay Emotions. http://www.srlabs.it/articoli_uk/ics.htm

Holmstrom, D. (2003). *Content based pre-encoding video filter for mobile TV.* Unpublished thesis: Umea University, http://exjob.interaktion.nu/files/id_examensarbete_5.pdf.

Nemethova, O., Zahumensky, M., & Rupp, M. (2004). Preprocessing of Ball Game Video-Sequences for Robust Transmission over Mobile Networks. In *Proceedings of the CIC 2004 The 9th CDMA International Conference*.

Guardian (2005). Romantic drama in China soap opera only for mobile phones. Guardian Newspapers Limited Available: http://www.buzzle.com/editorials/6-28-2005-72274.asp

Avisynth (2005). http://www.avisynth.org/

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

Odyssey software inc. CFCOM (2003). http://www.odysseysoftware.com/

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

ETSI (2005). Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines. http://webapp.etsi.org/action/PU/20050301/tr_102377v010101p.pdf

Jumisko-Pyykkö, S., Kumar, V. M. V., Liinasuo, M., & Hannuksela, M. (2006). Acceptance of Audiovisual Quality in Erroneous Television Sequences over a DVB-H Channel. In *Proceedings of the Second International Workshop in Video Processing and Quality Metrics for Consumer Electronics*.