

Can Small Be Beautiful? Assessing Image Resolution Requirements for Mobile TV

Hendrik Knoche
Dept. of Computer Science
University College London
London. WC1E 6BT UK
+44 (0) 20 7679 3644
h.knoche@cs.ucl.ac.uk

John D. McCarthy
Dept. of Computer Science
University College London
London. WC1E 6BT UK
+44 (0) 20 7679 3644
j.mccarthy@cs.ucl.ac.uk

M. Angela Sasse
Dept. of Computer Science
University College London
London. WC1E 6BT UK
+44 (0) 20 7679 3644
a.sasse@cs.ucl.ac.uk

ABSTRACT

Mobile TV services are now being offered in several countries, but for cost reasons, most of these services offer material directly recorded for mobile consumption (i.e. without additional editing). The experiment reported in this paper, aims to assess the image resolution and bitrate requirements for displaying this type of material on mobile devices. The study, with 128 participants, examined responses to four different image resolutions, seven video encoding bitrates, two audio bitrates and four content types. The results show that acceptability is significantly lower for images smaller than 168x126, regardless of content type. The effect is more pronounced when bandwidth is abundant, and is due to important detail being lost in the smaller screens. In contrast to previous studies, participants are more likely to rate image quality as unacceptable when the audio quality is high.

Categories and Subject Descriptors

H.4.3 [Communications Applications], H.5.1 [Multimedia Information Systems]

General Terms

Design, Experimentation, Measurement, Human Factors, Performance, Economics.

Keywords

Mobile TV, resolution, viewing distance, acceptability

1. INTRODUCTION

There are many different delivery scenarios associated with the term *mobile TV*. These range from Live TV to push-and-store services, and even transfer of content from a Personal Video Recorder (PVR) to a mobile device for watching on the move. Where TV is *received directly* on a mobile device, there are more methods of delivery, including reception via terrestrial (e.g. DVB-H [1], Wi-Fi [2]) or satellite (e.g. SDMB) networks [3].

In addition to different modes of delivery, different types of content are being considered for mobile TV. These range from highly interactive content specifically created for the mobile, to services that relay material produced for standard TV

consumption. With TV material, the content may undergo an additional editing process to prepare it for mobile consumption or it may be directly recorded.

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 Service (QoS) [4] or Quality of Experience (QoE) [5] for this type of service. An important factor is *the resolution of the image* to be delivered to end-users. Image resolution is important for a number of reasons.

1. 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).
2. Mobile devices are operated at 'arm's length'; continued viewing at distances closer than the resting point of vergence – approx. 89cm, with a 30° downward gaze – can contribute to eyestrain [6]. When viewing distances come close to 15cm, people experience discomfort [7]. Paper, keyboard and display objects are typically operated at distances ranging from 30cm to 70cm.
3. If the resolution of TV images can be reduced without affecting the perceived visual quality, less bandwidth is required - always a key concern in the mobile domain.
4. The camera shots used in television range from long shots (LS) to extreme close-ups (XCU) and consider image size and resolution of typical TV setups [8]. It is clear that image size and resolution cannot be reduced indefinitely as important detail will be lost. Logically, there must be some lower limit to the resolution of a watchable TV service.
5. Results from focus groups with people unfamiliar with mobile TV found that concerns about screen size (both in terms of watchability and portability) may inhibit uptake [9]

To address these issues we conducted a study of the image resolution requirements for mobile TV. The aim of the study was to identify the *minimum acceptable image resolution* of mobile TV for a range of difference bitrates and content types. We also wished to assess the impact of reduced image resolution and possible interactions with audio quality on user experience, as an understanding of these problems users experience can inform technical solutions. To ensure the validity of the results, all tests were conducted on a mobile device.

It is important that tests are carried out on a mobile device as it cannot be assumed that the experience of watching a small TV

window on a 17" monitor at a fixed distance is the same as watching the same window on a mobile device. With a hand-held device, users can easily move the screen closer to them. When watching on a large screen, they must move their whole body closer to the display, which requires more effort. TVs are usually watched in a posture where the head is upright. Handheld devices are operated with the head tilting down.

Section 2 reviews previous literature on the effects of image size, resolution and audio-visual quality interaction. Section 3 describes the study on image resolution and presents the results, which are discussed in Section 4. We conclude with a set of recommendations for testing and delivery of mobile TV.

2. BACKGROUND

2.1 The effects of image size

The extant research on mobile TV has focused on a comparison of codecs [10], the effects of frame rate reduction [11], [12], and the tradeoff between audio and video bitrates [13]. There has also been an emphasis on Sports [14], [12], which is one of the more challenging content types to deliver effectively. A factor that has *not* been investigated to date is how the *resolution and size* of the video image and the display affects the perceived video quality of mobile TV. Furthermore, apart from a few exceptions [12], most studies of mobile multimedia quality have not actually been conducted on mobile devices; rather, they displayed small images on a normal (15-17") LCD monitor [13] at fixed distances.

The size of the display in the viewer's visual field depends on both the size of the screen and the distance between viewer and the screen. Viewing ratio (VR) is defined as the viewing distance divided by the picture height (H).

Previous research has examined the impact of increasing the image size in the viewer's visual field by means of large physical displays or projection areas. Typically these studies have compared very large size screens (e.g. 46") to standard sized TV screens (15-20") [15] [16]. The results show that larger image sizes are more arousing than smaller ones, better remembered, and better liked. Other studies also show that users generally prefer bigger image sizes – ideally depicting people and objects up to life-size [17]. Buxton found that the sense of telepresence was so compelling when a video projection screen was used that participants referred to an object on their desk as if it was a shared space, when in fact it was not [18].

When it comes to TV images, the general message from these studies is, '*the bigger the better*'. This clearly presents a challenge to mobile TV where there is a tradeoff 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 [9]. 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 one of the few studies that specifically examined smaller screens, Reeves et al. [15] found no difference in arousal and attention between users watching 2" and 13" screens, although arousal and attention were larger with a very large screen (56") [19].

Other studies have even shown that smaller image resolutions can improve task performance. For example, [20] 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 decrease satisfaction when compared to 320x240 image resolutions [21]. 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 [22].

Another approach to the problem is to identify the possible effects based on known principles. 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.
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. Thus, for a fixed bitrate it is possible that *perceived quality is decreased* when image resolution is reduced.

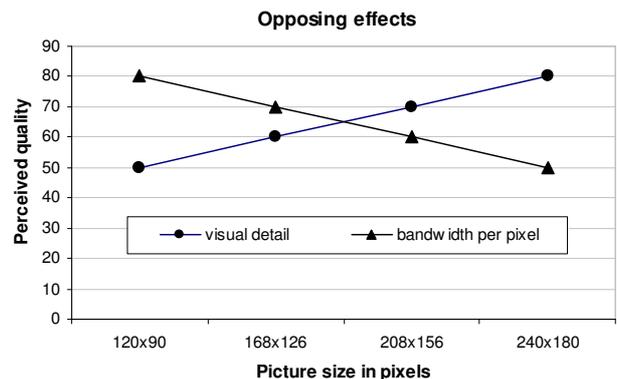


Figure 1: Trading off detail for improved quality

Both of these effects are illustrated in Figure 1 for a range of possible mobile resolutions. A priori, it is unknown which of these effects is dominant for any particular bitrate, or whether an interaction of effects is present.

2.2 Size, Resolution and Viewing Distance

It is important to distinguish the size of the display from the resolution. In print media, resolution is commonly defined in dots per inch (dpi), yet there is no equivalent metric for digital multimedia. Increasing the number of pixels in a given area can increase the perceived quality, but there is an upper limit to visual acuity. An important limiting factor is the viewing distance. As a display is moved further away it becomes more difficult to resolve detail in the display. In [23] the viewing angle is captured by the term *Viewing Ratio (VR)*, which is the ratio of the viewing distance to the picture height (distance/height), and normally expressed in relative units of picture height. Thus, a VR of 10H means that the screen is viewed at a distance of 10 picture heights.

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) [24]. 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 ($1/60^\circ$).

To put this figures in context: the iPAQ 2210 used in our study has a physical screen size of 55x73mm. At a viewing distance of 40cm, this area subtends visual angles of $7.8 \times 10.4^\circ$ in the X and Y dimensions. Taking the figure for 20/20 vision ($1/60^\circ$), users cannot distinguish more than 60 pixels per degree. Thus, on a 55x73mm screen at a viewing distance of 40cm they could not make use of resolutions higher than 468x624 for text discrimination. These calculations indicate that the 240x320 screen on the standard iPAQ can be doubled before it nears the limit of normal human perception.

Other estimates on the resolving power of human vision come from research on TV. Here, visual acuity is often determined using sets of alternating black and white lines of equal width. One black/white line pair (or 2 pixels) represents one cycle. 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. Under some conditions, with high contrast line pairs, human visual acuity extend beyond 40 cycles (or 80 pixels) per degree; but approximately 22 cycles (44 pixels) per degree is perceived as a sharp image [25]. Using this measure, the iPAQ described earlier has a resolution of approximately 15cycles/degree at a distance of 40cm – classified as low to normal resolution in TV terms.

In our study (see Section 3) we kept the resolution of the devices used fixed, while varying the resolution of the video images. In other words, the smaller resolution videos are represented by fewer pixels. As the media player on the device is not capable of scaling the images this also results 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.

2.3 Audio-visual quality interaction

An important consideration for mobile TV is how the bitrates allocated to the audio and video streams interact to affect the perceived quality of service. A recent study by Hands suggested that humans integrate audio and video quality together to evaluate overall multimodal quality [26]. For 'head-and-shoulders' video conferencing material, Hands found that multimodal quality was predicted by a regression equation of the form $AQ+VQ+(AQ \times VQ)$ – where AQ and VQ represented isolated evaluations of audio and video quality. In contrast, with high motion clips a reduced form of the equation with just the interaction term, $(AQ \times VQ)$, gave the best predictions of multimodal quality. In both cases, it was clear that the audio quality had an impact on multimodal quality assessment. The weighting of audio quality was greater for the videoconferencing material.

In a separate study of MPEG4 and AAC standards, Winkler and Faller evaluated perceived quality of very low bitrate mobile TV (40kbps-72kbps) with a range of content typical of this medium [13]. Similar to [26] their results also showed that good modeling of multimodal quality is possible using only a multiplicative or

interaction term ($AQ \times VQ$) although slightly better predictions can be obtained with a simple additive model ($AQ+VQ$).

In another study of audio-visual interactions, Winkler and Faller found that selecting mono audio for a given bitrate gives better quality ratings and that more bitrate should be allocated to the audio for more complex scenes [13].

These results provide clear evidence that levels of audio quality contribute to overall multimodal assessment. However, few previous studies have studied how audio quality can affect video quality ratings. As a byproduct in a study on TV viewing experience, Neuman et al. discovered that the perceived video quality was improved by better audio [27]. However, it was only the case for one of the three used content types. Similarly, a study by Beerends and Caluwe, using a 29cm monitor, found that rating of video quality was slightly higher when accompanied by CD quality audio than when accompanied by no audio [28]. The effect, however, was very small and has not been replicated with small screens. To examine whether this holds in a mobile context we asked people to rate video quality at two different levels of audio quality.

3. IMAGE RESOLUTION STUDY

The study adopted a method used in a recent study of quality tradeoffs for mobile sports content [9]. The logic of the method was to gradually change encoding parameters to find the critical point where quality becomes unacceptable.

In the current study, the aim was to evaluate the effects of varying image resolution and encoding bitrate on service acceptability. 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 did not want to control for viewing distance directly. As with normal use, participants were free to adjust the viewing ratio (VR) of the different image resolutions to their individual preferences. Thus, before running the study, the viewing distances participants will adopt is unknown.

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

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 different ways. Within a particular TV clip the bitrate allocated to video was gracefully degraded every 20 seconds by 32 kbps from a maximum of 224kbps down to 32kbps. These intervals are illustrated in Table 2. 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 coded the audio channels at 16kbps (Windows Media Audio V9) whereas the *High Audio* clips were coded at 32 kbps. These values were selected based on results of previous studies on mobile devices in which participants' acceptability of 32bps audio compared to 16kbps audio had declined from 95% to 80% [29].

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

Although the primary task of participants was to rate the video quality, the aim of this manipulation was to examine whether low audio quality would bias people's perception of the video quality as has been indicated by previous studies, e.g. [27]. Finally, we also recorded users while they watched the TV to capture how close they held the device under the different conditions.

3.1 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 [30] 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 will typically be 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 thus 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 recent studies of mobile TV services [9], [31]. These indicated that watching time was likely to be between 2 and 5 minutes, and that news was the most 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, to give 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 Headlines
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, 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 at 12.5fps. These segments were encoded using 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 16kbps 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 in the video quality in any of the segments.

3.2 Design

As shown in Table 4 we ran four different groups, each comprising 32 participants. Each group was presented 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. The dependent variable was *Video Acceptability*. Independent variables were *Image Resolution*, *Content Types*, *Video Bitrate*, *Audio Bitrate*. Control variables were *Resolution Order*, *Sex*, and *Corrected Vision*. The variable *Corrected Vision* coded whether participants considered themselves to have normal vision or whether they wore contact lenses or glasses.

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

3.3 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 transfective 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 [32] to embed the Windows Media Player. It presented the clips along with a volume control and two response buttons to indicate acceptable and unacceptable quality. A screen shot of the application is shown in Figure 2.



Figure 2: Application with volume control on the lower left and to its right feedback buttons ‘Acc.’ and ‘Unacc.’

3.4 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. At the end of the video rating session, we interviewed the participants to find out what aspects of the video quality they found unacceptable for the different types of content.

3.5 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.

3.6 Results

Before analyzing the results, we conservatively coded each 20 second interval of a clip as *unacceptable* if they had given a rating of 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 – *Image Resolution*, *Video Bitrate*, *Content Type* and *Audio Bitrate*. Control variables *Sex*, *Corrected Vision* and *Resolution Order* were also included in this analysis. Post-hoc within-subject tests were performed using non-parametric Friedman and Wilcoxon tests.

The regression revealed significant effects on all of the control variables. *Sex* was a significant predictor of acceptability with women being less likely to rate a clip as unacceptable than men [$\chi^2(1)=12.6, P < 0.001$]. The *Corrected Vision* variable was also a predictor of acceptability [$\chi^2(1)=54.8, P < 0.001$]. Those wearing glasses or contact lenses were less likely to rate a clip as unacceptable than those with normal vision. *Resolution Order* was also a significant predictor of acceptability [$\chi^2(1)=120.7, P < 0.001$]. Those participants who started with large image resolutions that got smaller were generally more likely to rate unacceptable than those who saw clips increasing in image resolution.

3.6.1 Resolution, Video Bitrate and Content Type

As expected the logistic regression also showed a significant effect of *Video Bitrate* on acceptability ratings [$\chi^2(6)=1186, p < 0.001$]. However, there was also an interaction between *Video Bitrate* and *Image Resolution* [$\chi^2(18)=165, p < 0.001$]. This interaction is illustrated in Figure 3 for the two highest and lowest bandwidths. For this and all subsequent figures, the *acceptability*

measure reported can be interpreted as *the proportion of the sample that finds a given quality level acceptable all of the time*.

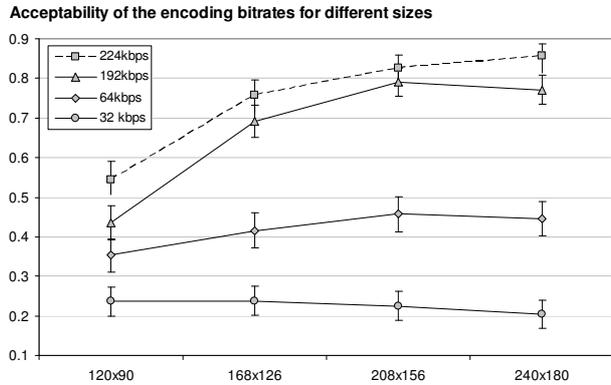


Figure 3: Averaged across content types, resolution effects are more pronounced at high bitrates.

Averaged across content types, acceptability declines with decreasing image resolution at higher bandwidths. At the lowest bandwidth, there appears to be a slight increase in acceptability. However, a post-hoc comparison revealed no difference between acceptability of the four image resolutions at the lowest bandwidth [$\chi^2(3) = 3.47, P = 0.324$] indicating that there were no quality gains from reducing the image resolution.

This pattern confirms the *visual detail* effect described in Section 2, Figure 1. There is no evidence that acceptability increases by increasing the *bandwidth/pixel*. When bandwidth is abundant, the *primary effect is a loss of detail*. However, even when bandwidth is scarce, and the baseline quality is low, we find no evidence that increasing the *bandwidth/pixel* can increase perceived quality.

The logistic regression also showed *Image Resolution* and *Content* as significant predictors of acceptability, [$\chi^2(3) = 446, P < 0.001$; $\chi^2(3) = 1056, P < 0.001$], and an interaction between *Image resolution* and *Content type* [$\chi^2(9) = 136, P < 0.001$]. A summary of this interaction is shown in Figure 4.

As shown in Figure 4, the different content types have very different levels of acceptability. Not surprisingly, the low motion animation clips received the best ratings – for this type of content there was no significant difference in acceptability as image resolution was reduced from 240x180 to 168x126 [$\chi^2(2) = 0.468, n.s.$], but at the smallest image resolution acceptability dropped off sharply [$Z = -6.49, P < .001$]. For News content the *acceptability significantly increases as the image resolution was reduced from 240x180 to 208x156* [$Z = -2.11, P < 0.05$], after which point there was a steady decline in acceptability with decreasing image resolution. Thus, for News, we do find evidence that bandwidth savings have increased perceived quality. The curve for Music videos was relatively flat, and there was no significant difference in acceptability across the four image resolutions [$\chi^2(3) = 6.1, n.s.$]. Finally, Sports coverage showed the lowest levels of acceptability. There was no significant difference in acceptability between the two largest image resolutions, but at image resolutions smaller than 208x156 acceptability significantly declined [$\chi^2(2) = 25.9, p < 0.001$]. To illustrate these effects in more detail, we subsequently present the results separately for the four content

types at each of the seven video bitrates and report the qualitative comments participants made on the problems they encountered.

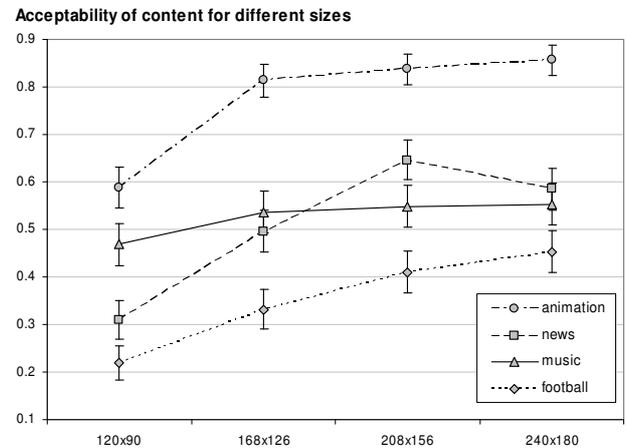


Figure 4: Image resolution effects depend on the content.

3.6.2 News

With News, the largest image resolution did not receive the highest acceptability ratings. When the image resolution was reduced to 208x156, perceived quality of the video improved. The effect was present at all video bitrates apart from 32 and 64kbps. There was also a dramatic reduction in acceptability between 168x126 and 120x90. At 32 kbps no differences in image resolution were observable.

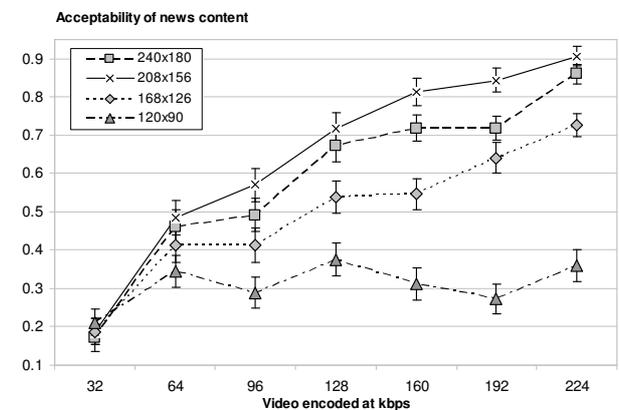


Figure 5: Acceptability of news content

When asked why they rated the News as unacceptable, participants mentioned a number of factors. Across all 128 participants, a total of 290 comments related to the unacceptability of News coverage. Of these comments, 34% related to *text detail*:- the legibility of the news ticker, the headline text, the clock, the logo, or the captions for the people being interviewed by the newscaster. Other problems people reported were *facial details* and expressions, the switch from anchor person to field reports (*shot types*), poor *audio fidelity* and a loss of *general detail*. A summary of these problems and the frequency with which they were mentioned is presented in Figure 6.

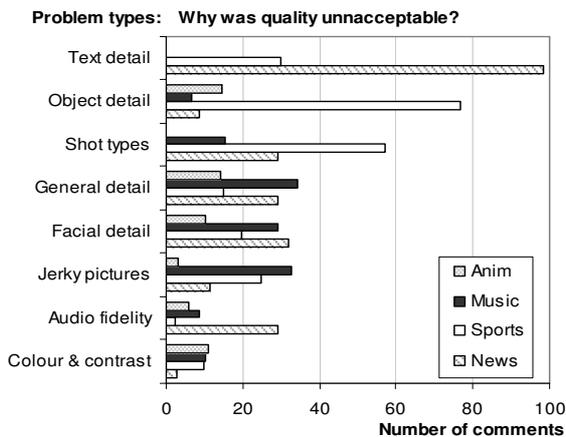


Figure 6: Reasons for unacceptable quality

3.6.3 Sports

With Football clips, acceptability increased with both *Image resolution* and *Video Bitrate*. However, even at the largest image resolution (240x180) and highest bitrate (224kbps) around 30% of participants found the quality to be unacceptable (Figure 7).

Of the qualitative comments collected, 248 related to unacceptability of the Sport material. The main problems participants reported was identifying *object detail*. In particular, participants reported problems seeing the ball and identifying players. The second most common complaint were certain *shot types* - specifically long shots of the entire pitch, - which people found very difficult to watch on the small screen. Other problems included the inability to read *text detail* about the teams and the scores, the *jerky*ness of pictures and the inability to see *facial detail* clearly (See Figure 6).

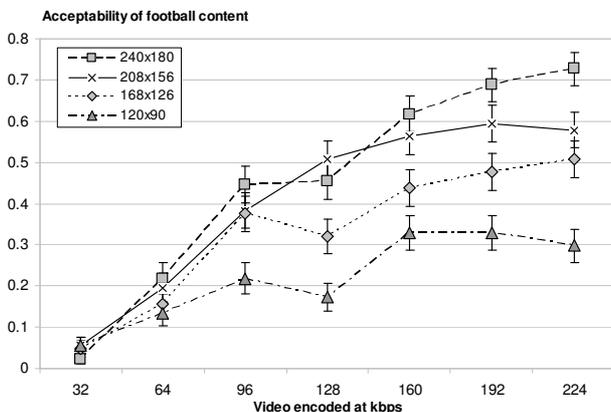


Figure 7: Acceptability of football content

3.6.4 Music

With Music clips the effects of image resolution were less pronounced, but there was a clear interaction between *Image resolution* and *Video Bitrate*. At the lowest bitrate, the smallest images were rated as the most acceptable, but at the highest bitrate they were the least acceptable. Again this is evidence of perceived increases in quality from a reduction in image resolution. For Music clips, there were fewer comments on why quality was unacceptable. Of the 172 comments, made 34% related to *general*

detail - such as blurriness and fuzziness. 33% related to the *smoothness of the frame rate*.

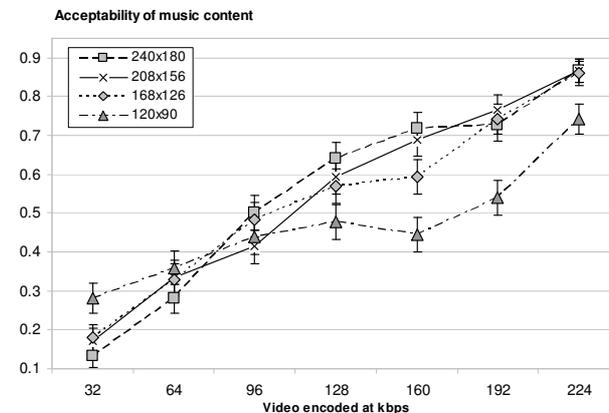


Figure 8: Acceptability of music content.

Interestingly, the proportion of comments relating to frame rate (*jerky pictures*) was much higher with Music than high-motion content such as Sports. For the participants who commented on *frame rate*, the problem seemed to lie with a disruption of the rhythm associated with the music being played. Other major problems include the lack of *facial detail*, special effects and edits (*shot types*) and *colour and contrast* (see Figure 6).

3.6.5 Animation

With the Animation clips, a reduction in image resolution had little effect on acceptability apart from the smallest image resolution where there was a clear reduction in perceived video quality (See Figure 9).

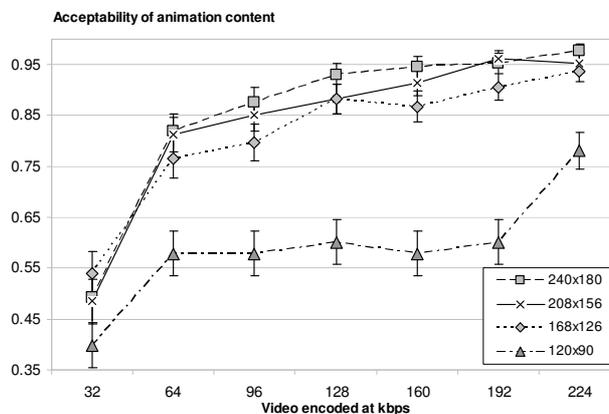


Figure 9: Acceptability of animation content.

Animation produced the fewest comments from participants in the qualitative interviews - only 64 in total, almost five times fewer than comments made about the News content. The most frequent complaint related to problems identifying the animal species in the animation when image resolution was very small. *General detail* was also mentioned and participants had problems when the image was very dark and the contrast was low (*Colour and Contrast*) *Facial detail* - such as the fidelity of the eyes and mouth - was also an issue as was the *audio fidelity* which participants complained was *'echoy'*.

3.6.6 Viewing Distance and Fatigue

An analysis of the video recordings of the participants revealed that the vast majority of the participants held the mobile device at a relatively fixed distance throughout the study. For both increasing and decreasing image resolution groups, there was no significant difference in the distance at which the iPAQ was held at the start or end of the study. Overall, the average viewing distance was 27cm with a range of 13 to 45cm. Of those that frequently changed viewing distance throughout the study, this seemed to be more related to adopting a more comfortable posture while holding the device.

In the qualitative interviews, participants made 147 comments that referred to experienced quality across all content types. The most frequent complaints were a general lack of detail, often referred to as a 'blurry' or 'fuzzy' display. There were also a large number of comments specifically citing difficulty when the image size was small. In addition, almost 10% of comments complained about visual fatigue from watching such a small screen – with problems such as 'It's tiring to watch' and 'My eyes hurt'. A further 8% complained about the effort involved when watching the very small screen with people complaining that they 'had to really concentrate to work out what was going on'.

As the viewing distance is relatively constant across different image sizes, this is probably not a problem of vergence, but of effort and fatigue from trying to decode information in such a small display.

Table 5: Problems across all content types

Problem	% of General comments
General detail	20%
Insufficient image size	18%
Fatigue	10%
Effort	8%

3.6.7 Audio-Visual Interaction

Finally, there was a significant effect of *Audio Bitrate* in the logistic regression [$\chi^2(1) = 62.8, p < 0.001$] but not in the direction expected.

As shown in Figure 10, at all video encoding bitrates the acceptability of the video was rated significantly higher at the lower audio bitrate.

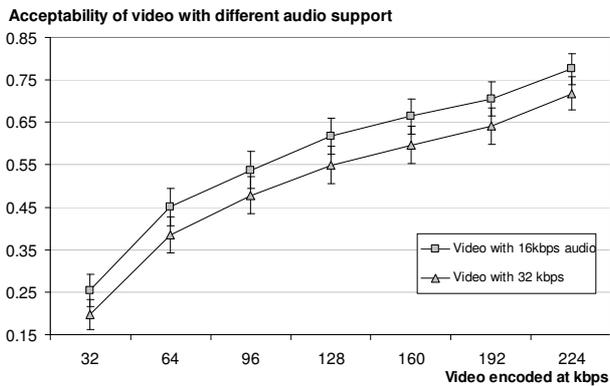


Figure 10: Video supported with different audio levels.

This effect held across different image resolutions and content types and was constant across the full range of bitrates, indicating

that there is no interaction between audio and video quality. One explanation of this surprising effect is in terms of expected quality. If participants have higher expectations of video quality when the audio quality is high, then they will rate quality as unacceptable sooner than when the audio quality is low. This behaviour would produce the pattern of results observed.

4. DISCUSSION

Both quantitative and qualitative results indicate that the primary effect of reducing image resolution is a loss of visual detail. Across content types, the effect of reducing image resolution is more pronounced when bandwidth is abundant. When the encoding bitrate is very low, there is little or no effect of reducing the image resolution, as visual detail is already poor. For all content types at 128kbps and above, there is a sharper reduction in acceptability when image resolution is dropped from 168x126 to 120x90.

The qualitative comments help to identify the source of the problems. Of the eight most frequently cited problems, five relate to identifying or distinguishing detail – such as text, faces, players, animals and the ball. For News, Sports and Music, participants also identified particular shot types that caused difficulty. There were relatively few comments on frame rate, apart from Music clips, in which 'jerky' frame motion seemed to be misaligned with the rhythm of the music and therefore disrupt the overall experience. Overall, audio quality received few comments, with the exception of News.

Apart from News coverage, we find little evidence of any bandwidth savings or increases in perceived quality from reducing the image resolution. For News, the primary detail on which quality was judged was the ability to distinguish textual information – whether the news ticker, the clock, headline text or person names. It seems that the slight increase in perceived quality with a reduction in image resolution to 208x156 was caused by a perceived increase in the quality of the text. If text were coded and transmitted separately from the video we would expect clips encoded at an image resolution of 240x180 to be more acceptable than 208x156.

Somewhat surprisingly, participants were less likely to rate quality as unacceptable when the audio quality was low (16kbps). This was an unexpected result given the findings of previous studies on audio-visual interactions which show that increasing audio quality increases video quality ratings. The explanation may lie in the way the task is framed. Whereas many previous studies required participants to rate video quality on a scale we followed the method recently used by [9] and asked people to indicate when they find it unacceptable. In this context, low audio quality seems to set participants' expectations such that they are less likely to rate the video as unacceptable. By contrast, those given high audio quality have higher expectations and are more easily disappointed with the visual counterpart.

In the previous study by Neuman et al., video quality was comparatively high (standard NTSC TV display) [27]. In the study by Beerends et al., participants judged the two lower video quality levels (where the video bandwidth was limited to 0.15 MHz and 0.025 MHz) worse when they were presented with audio than without audio. The two higher video qualities had received better ratings with audio than without audio [28]. Unfortunately, there are no detailed data available in the paper on the influence of the different audio qualities on the perceived video quality.

In terms of viewing distance we found no evidence that participants modified this for the smaller image resolutions. Consequently viewing ratios were higher for the smaller screen size – much higher than those typically observed with normal TV. An illustration of this is shown in Figure 11 which plots viewing ratio vs. vertical screen resolution for standard TV and the mobile TV resolutions we tested. What is evident in the figure is that standard TV is much closer to the limits of human perception than mobile TV.

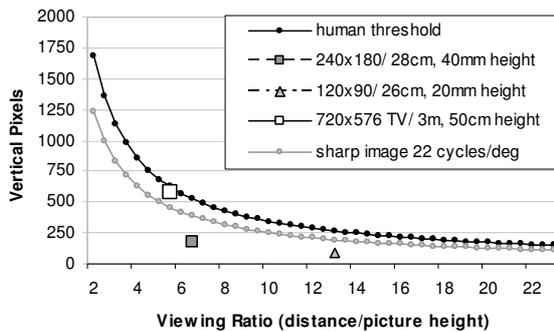


Figure 11: Mobile TV resolutions compared to standard TV.

5. CONCLUSIONS

5.1 Substantive

In reviewing the background literature there were a number of different effects associated with the image resolution used for mobile TV. Three effects that can be delineated are:

1. *Visual Detail*
2. *Bandwidth/Pixel*
3. *General Arousal*

The results of the study indicate that the dominant effect of reducing the image resolution is a loss of *visual detail*. This conclusion is reinforced by the qualitative comments on the problems participants experienced. Of the 921 qualitative comments collected, 63% relate directly to a problem identifying specific detail in the image.

This effect, however, is not universal. For News, we found that the increased *bandwidth/pixel* with a slightly smaller image gives improved acceptability ratings. However, as the image resolution is reduced further the loss of detail again dominates. Another exception was very low bandwidth music clips. This was the only content type where the smallest image resolution was actually rated the most acceptable.

No comments made by participants related to *general arousal*. Previous studies indicate that arousal is related to the visual angle subtended by the image, thus if arousal were of primary interest to participants we would expect them to adjust for the smaller image size by moving the device closer. In our laboratory setup we found no evidence of such an adjustment as the image size and resolution were reduced.

An additional effect that was not predicted at the outset is *fatigue*. Comments from a significant number of participants give evidence for eyestrain with prolonged viewing at the smallest image size. This is similar to observations of Wilson et al. on perceptual strain with low quality video conferencing [33]. The

potential health impact of small screen mobile TV needs to be fully investigated.

Another unexpected result was the effect of *audio bitrate* on acceptability. In line with previous results, we expected that better audio would lead to better video ratings. This was not the case. Instead, better audio made participants more likely to rate the video as unacceptable. This mirrors the findings of Bouch et al. showing that lower expectancies can produce more positive QoS evaluations [34].

Overall, the results indicate that there is a lower limit to the image resolution of mobile TV services for mass market consumption. When the content consists of unedited re-purposed TV material, and bandwidth is abundant, TV displays less than 168x126 in image resolution give a sharp drop in acceptability. The general recommendation to service providers would be to encode at the largest image resolution possible for any particular content type. The two exceptions to this rule are for News and very low bandwidth Music videos. At the lowest bandwidth (32kbps), Music videos were more acceptable at the smallest image resolution. For News coverage, legibility of text is an important issue and may be improved by reducing the image resolution of the content prior to encoding. More generally, however, the recommendation is to stream the text information separately to the device.

Irrespective of the size however, many participant complained about the clarity of the image. By examining how the resolutions and viewing ratio compare to normal TV it is clear that even the maximum resolution we tested is well below that required for a sharp TV image.

5.2 Methodological

In adopting the method used by [9], we found that rating quality through a binary acceptable/unacceptable response comes natural to users, and does not interfere with the viewing experience. Other methods use unnaturally short clips, and constantly prompt users to assign a label (e.g. excellent, fair, bad) to the quality. The results from binary responses can be easily translated into percentages of satisfied customers, which is of high relevance to service providers. The measure is also independent of any particular dimension or video quality, and when used alongside qualitative interviews, provides a clearer insight into the actual problems that users experience.

5.3 Future Work

To improve the perceived quality of TV material repurposed for mobiles, there are a number of different avenues to explore. Firstly, as the primary problem is a loss of visual detail, one approach is to focus on the problem that dominates – namely insufficient text detail. Here it would be much more efficient and effective to stream the text alongside the TV coverage. Thus protocols as e.g. SMIL [35] should be integrated into the mobile TV production process to synchronize text and video streams to mobile devices.

Secondly, as VGA resolution mobile devices are now available, it may be possible that encoding at resolutions that preserve visual detail but displaying at 640x480 might realize gains in acceptability. This requires further study.

Finally, on the evaluation side, further work is needed to understand how audio and video qualities interact to bias users' perception of video quality acceptability. As audio quality has a clear impact on perceived quality it is important to evaluate which

audio quality is the best match for any particular level of video quality to maximize service acceptance.

Acknowledgments

This study was designed and executed by Hendrik Knoche as part of his PhD program. The work was funded by the EU IST-project MAESTRO. The CFCOM software was provided by Odyssey software, inc.

6. REFERENCES

- [1] Reimers, U. *Digital Video Broadcasting* IEEE Communications Magazine, 36 (6) 1998.
- [2] IEEE *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications* (Rep. No. IEEE 802.11) The Institute of Electrical and Electronics Engineers, 1999.
- [3] Narenthiran, K., Karaliopoulos, M., Tafazolli, R., Evans, B. G., Vincent, P., Selier, C. et al. *S-DMB System Architecture and the MODIS DEMO* in IST Mobile and Wireless Communications Summit 2003, 2003
- [4] ITU-T *End-user multimedia QoS categories* (Rep. No. G.1010), 2001.
- [5] Jain, R. *Quality of Experience* IEEE Multimedia, 11, p. 95-6, 2004.
- [6] Owens, D. A. & Wolfe-Kelly, K. *Near Work, Visual Fatigue, and Variations of Oculomotor Tonus* Investigative Ophthalmology and Visual Science, 28, p. 743-9, 1987.
- [7] Ankrum, D. R. *Viewing Distance at Computer Workstations* Work Place Ergonomics, p. 10-2, Sep/Oct 1996.
- [8] Weiner, R. *Webster's New World Dictionary of Media and Communications* (rev. and updated ed.) New York, NY: Macmillan, 1996
- [9] Knoche, H. & McCarthy, J. *Mobile Users' Needs and Expectations of Future Multimedia Services* in Proceedings of the WWR12, 2004
- [10] Winkler, S. & Dufaux, F. *Video Quality for Mobile Applications* in Proc.of SPIE, p. 593-603, 2005
- [11] Winkler, S. & Faller, C. *Audiovisual quality evaluation of low-bitrate video* in Proc.SPIE/IS&T Human Vision and Electronic Imaging, p. 139-48, 2005
- [12] 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
- [13] Winkler, S. & Faller, C. *Maximizing audiovisual quality at low bitrates* in Proc.of Workshop on Video Processing and Quality Metrics, 2005
- [14] Wikstrand, G. & Sun, J. *Determining utility functions for streaming low bitrate football video* in Proc.of IMSA 2004, 2004
- [15] Reeves, B. & Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places* University of Chicago Press, 1998
- [16] Lombard, M., Grabe, M. E., Reich, R. D., Campanella, C., Ditton, T. B. *Screen Size and viewer responses to television: A review of research* in Annual Conf.of the Assoc.for Education in Journalism and Mass Communication, 1996
- [17] Okada, K.-I., Maeda, F., Ichikawaa, Y., Matsushita, Y. *Multiparty videoconferencing at virtual social distance: MAJIC design* in Proc.ACM conf.on Computer supported cooperative work, p. 385-93, 1994
- [18] Buxton, W. A. S. *Telepresence: Integrating shared task and person spaces* in Proceedings of Graphics Interface '92, p. 123-9, 1992
- [19] Reeves, B., Lang, A., Kim, E., Tartar, D. *The effects of screen size and message content on attention and arousal* Media Psychology, 1, p. 49-68, 1999.
- [20] Horn, D. B. *The effects of spatial and temporal video distortion on lie detection performance* in Proceedings of CHI '02, 2002
- [21] Kies, J. K., Williges, R. C., Rosson, M. B. *Controlled Laboratory Experimentation and Field Study Evaluation of Video Conference for Distance Learning Applications* (Rep. No. HCIL 96-02) Virginia Tech, 1996.
- [22] Barber, P. J. & Laws, J. V. *Image Quality and Video Communication* in R. Dampier, W. Hall, & J. Richards (Eds.) Proceedings of IEEE International Symposium on Multimedia Technologies & their Future Applications, p. 163-78, London, UK: Pentech Press, 1994
- [23] ITU-R *Methodology for the subjective assessment of the quality of television pictures*. (Rep. No. BT.500-7), 2004.
- [24] Westheimer, G. *Visual acuity*, In W.M.Hart (Ed.), Adler's Physiology of the Eye: Clinical Application (9th ed., St. Louis, Mo: CV Mosby, 1992
- [25] Silbergleid, M. & Pescatore, M. *The Guide To Digital Television* (3rd ed.) Miller Freeman Psn Inc, 2000
- [26] Hands, D. S. *A Basic Multimedia Quality Model* IEEE Transactions on Multimedia, 6 (6), p. 806-16, Dec. 2004.
- [27] Neumann, W. R., Crigler, A. N., Bove, V. M. *Television Sound and Viewer Perceptions* in Proc.Joint IEEE/Audio Eng.Soc.Meetings, p. 101-4, 1991
- [28] Beerends, J. G. & de Caluwe, F. E. *The influence of video quality on perceived audio quality and vice versa* Journal Audio Eng.Soc., 47, p. 355-62, 1999.
- [29] McCarthy, J., Miras, D., Knoche, H. *TN01-1.1.03_UCL_MAESTRO_bandwidth_study_V02*, 2004.
- [30] VQEG *Final Report from the video quality experts group on the validation of objective models of video quality assessment* (Rep. No. <http://www.vqeg.org>), 2000.
- [31] Södergård, C. *Mobile television - technology and user experiences Report on the Mobile-TV project* (Rep. No. P506) VTT Information Technology, 2003.
- [32] *Odyssey software inc. CFCOM*, <http://www.odysseysoftware.com/>. 2003
- [33] Wilson, G. & Sasse, M. A. *Do Users Always Know What's Good For Them? Utilising Physiological Responses to Assess Media Quality*. in S. McDonald, Y. Waern, & G. Cockton (Eds.) Proc.of HCI: People and Computers XIV, p. 327-39, 2000
- [34] Bouch, A. & Sasse, M. A. *Why Value is Everything: A User-Centered Approach to Internet Quality of Service and Pricing* in Proceedings of 9th International conference on Quality of Service (IWQoS'01), 2001
- [35] W3C-Recommendation, *Synchronized Multimedia Integration Language SMIL 1.0 Specification*., <http://www.w3.org/TR/1998/REC-smil-19980615>. 1998