An Investigation of the Perceived Quality of Digital Media: Research and Research Design Issues

Richard A. Schwier
Earl R. Misanchuk

Abstract: This article reports three experiments which tested the perceived quality of digital images, and discusses issues about conducting research into questions about technical quality in multimedia. Various quality settings are commonly available to multimedia developers, with dramatic effects on the file sizes and consequent loading and execution time in multimedia, yet little is known about whether viewers perceive the differences among different quality settings. Generally speaking, we have found and replicated evidence that for medium speed computers (60-80 Hz), digital video can be recorded and used at 10 fps, at a significant savings in storage space, and with positive effects on the perception of users. For still pictures, it is clear that users prefer larger pictures, and 32 bit color for large images. The results are less clear for 8 and 16 bit images, but there is some evidence that 16 bit images can be used with smaller images with little or no loss to perceived quality. There seems to be an interaction between the detail of images and preferred bit depth of color. For smaller images, greater bit depth was positively correlated with perceived quality. For the image with less detail, 16 bit color was highly preferred, and 32 bit color less preferred. Comments from participants revealed that most were cueing on resolution in still images, and on smooth motion and synchrony in video images for making their selections.
The second part of the paper discusses issues around conducting this type of research, including decontextualization, selection of variables and their values, and presentation protocol for treatments.

Resume: Cet article rend compte de trois experiences qui ont mesure la qualite per9ue d'images numerisees, et discute de problemes concernant la conduite de recherches sur des questions portant sur la qualite technique dans le multimedia. Plusieurs niveaux de qualite sont couramment disponibles pour les concepteurs en multimedia, niveaux qui influencent dramatiquement la grosseur des fichiers, et consequemment le temps de chargement et d'execution de l'application multimedia; par contre nous en savons peu sur la capacite des utilisateurs a percevoir les differences parmi les niveaux de qualite disponibles. De facon generale. nous avons demontre au cours d'experiences repetees, que pour un taux de regeneration de 1'image de 60 - 80 Hz, un video numerise peut etre enregistre et vu a 10 tps (trames par seconde), avec une economie d'espace memoire appreciable, tout en ayant un effet positif sur la perception des utilisateurs. Pour les images fixes, il est clair que les utilisateurs preferent des images plus grandes, et avec une profondeur de couleur de 32 bits pour les grandes images. Les resultats sont moins clairs pour les images 8 et 16 bits, mais il est assez evident que les images 16 bits peuvent etre utilisees pour de plus petites images avec peu ou pas de perte de qualite per9ue. II semble y avoir une relation entre le degre de details des images et la profondeur de couleur preferee. Pour les images plus petites, une plus grande profondeur de couleur est en correlation directe avec la qualite percue. Pour les images comportant moins de details, une profondeur de couleur de 16 bits etait de loin preferee. tandis qu'une profondeur de 32 bits 1'etait moins. Les commentaires des participants ont revele que la plupart determinaient leur choix a partir du degre de resolution pour les images fixes, et a partir du mouvement sans a-coups et la synchronicite pour les images video.

La seconde partie de cet article discute de problemes poses par la conduite de ce type de recherche, incluant la decontextualisation. la selection des variables et leurs valeurs, et un protocole de presentation des images.

Introduction

This article has two purposes. The first is to report the results of experiments on the perceived quality of digital pictures and movies. For these experiments, we chose commonly available computers and common settings for digitizing media, most often those settings which production programs used as defaults, rather than build treatments using high end technology and esoteric production approaches. It is possible to create better digital media by using expensive hardware and through judicious selection of compression alogrithms, but few users have the resources or knowlege necessary to take advantage of these improvements. We wanted to investigate how typical quality selections influenced the perception of viewers. In the second part of this paper, we will discuss several issues researchers must grapple with when conducting experiments to assess the perceived quality of still pictures and digital video in multimedia-issues such as how to define quality, how to select and categorize independent variables, and whether multimedia quality should be assessed in a decontextualized or contextualized treatment. In this way, we hope to throw some light on the tradeoffs which seem to be inherent when treatments are designed to compare multimedia quality variables.

Multimedia Quality Experiment

One problem multimedia designers face when using graphics and movies is the size of picture files. Large files create storage problems, and in some cases they take an inordinate amount of time to load, and therefore cause programs to execute slowly.

There are two ways to reduce the size of picture files: reduce the size of the image; or, reduce the number of colors (bit depth) when creating the files.

Color can be represented at different bit depths, although 8 bit (256 colors). 16 bit (thousands of colors) and 32' bit (millions of colors) are typical choices. Adjustments to either of these variables will affect the quality of the image when it is displayed on screen. Whether or not the quality difference is perceived, and whether the increased file size and loading time is worth the concomitant increase in perceived image quality is often a difficult judgement to make.

A study of the perceived quality of digital still pictures and movies (Schwier and Misanchuk, 1996a) suggested that quality (in the eyes of the learner) may be reflective of technical superiority for larger pictures (640 x 480 pixels), but not for smaller ones (320 x 240 pixels). Viewers generally preferred larger still pictures to smaller ones, and for larger pictures, they preferred those with higher bit depths (see Figure 1). The exception was with small pictures, where those with 16 bit color were preferred to either 8 or 32 bit color in the same size. In fact, the 16 bit,

For 32 bit color, Macintosh systems actually use 24 bit color, but some software packages refer to it as 32 bit. We caution the reader that the actual bit dipt is different than the label suggests.

smaller picture was also preferred to every picture other than the large, 32 bit color rendition. The investigators speculated that this may have been the result of an optimal resolution match between the monitor setting and the bit depth of the picture. The picture used in this experiment was a full color portrait, selected because it had a wide variety of hues and the background had a subtle gradation of value from dark to light[2] .

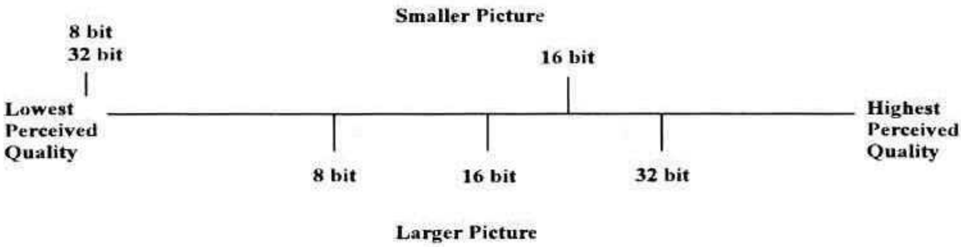

Figure 1. Thurstone Scale of preferences of digital pictures at two size settings and three bit depths of color from Schwier and Misanchuk (1996a).

We questioned how the treatment may have influenced the results of this experiment. Given that it was a portrait used in the treatment, we wondered whether it might have invited subjects to focus on different aspects of the photograph when making their choices. One subject may have preferred a soft focus, another might have preferred clear expressions, and another may have liked a posterized effect on the background. Subjects were not asked how they were making their decisions, so there was little solid guidance available for interpreting the results.

Movie files present similar difficulties to an instructional designer. Large movie files require substantial storage space, and with some applications, large movie files take a great deal of time to load and play within an application.

There are two ways to reduce the size of movie files:

1. reduce the size of the movie window; or
2. reduce the number of frames/second of the recording.

Movies can be recorded in various window sizes, including quarter (160 x 120 pixels), half (320 x 240 pixels) and full screens (640 x 480 pixels). They can also be recorded at any speed up to 30 frames per second (fps), which is the standard rate for NTSC video playback. Each frame of video requires additional file space, so the greater the number of frames per second, the greater the resulting file size. But there is a further complication. Unless fairly sophisticated, high-end production software and hardware is used, computers cannot record 30 frames per second. Indeed, even if recorded, few computers are capable of playing back larger video windows at 30 frames per second.

[2] The photograph can be viewed at: http://www.extension.usask.ca/PapersyMisanchuk/AECT97/Photol.html

Adjustments to either of these variables (window size, frames per second) may also reduce the perceived quality of the image when it is displayed on screen. The viewer may not be able to see a satisfactory amount of detail in smaller windows, and using larger windows may result in a choppier, less fluid display of motion. At the present time and under almost every condition, digital movies are poor in quality. They also demand large file sizes, rendering them slower to load onto screens.

For digital movies (see Figure 2), viewers seemed to favor larger windows (320 x 240) over smaller windows (160 x 120). Frames per second also appeared to be an important variable, but in a counter-intuitive way. Generally speaking, recordings made at lower frame rates were preferred to recordings at higher frame rates. This held for both larger and smaller window sizes (Schwier and Misanchuk, 1996a). The researchers speculated that the lower frame rates more closely matched the processing speed of the computers, resulting in the impression of more fluid, less choppy, movement. The video clip used in this experiment was a fairly monochromatic sequence of a pelican swimming, flying and eating fry in a river, beneath an overflowing weir; there was no accompanying audio. The movement on the original videodisc was fluid and multidirectional, and it contained several visual details[1].

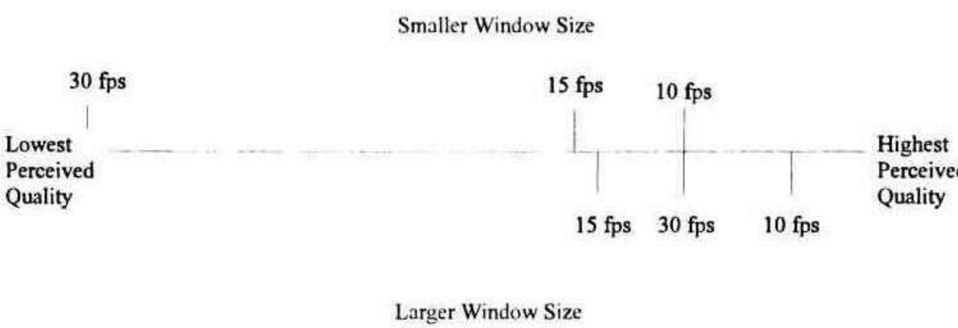

Figure 2. Thurstone Scale of preferences for digital movies at two size settings and three frames per second settings from Schwier and Misanchuk (1996a).

We were unable to find any other research which addressed the influence of these variables on the perceived quality of pictures and movies in multimedia, so we designed an experiment to refine some of the variables, replicate this study and determine the actual costs and perceived quality differences when the window size and bit depths of pictures are changed and when the window size and frames per second of a digital movie are manipulated.

Research questions

How will larger and smaller pictures compare in file size and perceived image quality?

[3] The video can be viewed at: http://www.extension.usask.ca/Papers/Misanchuk/AECT07/Video1 .html

How will altering the bit depth of images influence their file size and perceived quality?

Will a picture which contains a high degree of detail be rated differently than a picture which contains significantly less detail?

How will larger and smaller movie windows compare in file size and perceived image quality?

How will recording digital movies at different frames per second influence their file size and perceived quality?

How will subjects describe the variables they considered when making comparisons?

Subjects

Thirty adult employees and educational technology students at the University of Saskatchewan volunteered to participate in the study.

Treatments

To measure perceived picture quality, two full-color photographs were scanned on an HP ScanJet II cx/T at its highest quality settings, one photograph for each of two matched treatments. Speculating about the results from the pilot study (Schwier and Misanchuk, 1996a), researchers questioned whether the amount of detail in the original treatment provided the strongest cue for making quality discriminations. For the two treatments in this experiment, one treatment employed a reproduction of an impressionist painting, Bathers at La Grenouillere by Claude-Oscar Monet, which had little precision and detail, but a full range of color (Roland, 1996). The second treatment employed a photograph of the interior of the National Gallery of Art in London, a photograph which contained a high degree of detail in addition to a full range of color (Roland, 1996).

For each treatment the scanned photograph was imported into Adobe Photoshop(™) to create six versions of the picture, including two image sizes (640 x 480 pixels and 320 x 240 pixels) and three bit depths of color (32 bit, 16 bit, and 8 bit). Each was saved as a PICT file without any type of compression algorithm. Each image was imported into a program created with Authorware Professional™) v. 3.5, and a series of "pages" created to provide a paired comparison of every possible combination of image variables for each picture. The order of comparison was constructed according to recommendations by Ross (1934) for conducting paired comparisons to eliminate effects of picture location and presentation order. A Thurstone Scale was constructed for comparing resultant data (Torgerson, 1958).

To measure perceived movie quality, six versions of a 30 second clip of video were digitized as QuickTime™ movies. The original video was recorded on a Sony Betacam™ and transferred to videodisc. A segment was chosen which had few colors (to reduce possible contamination from this variable) but a great deal of motion which was, in this experiment, accompanied by a synchronized sound track (with key fram synchronization at half-second intervals). The segment was a

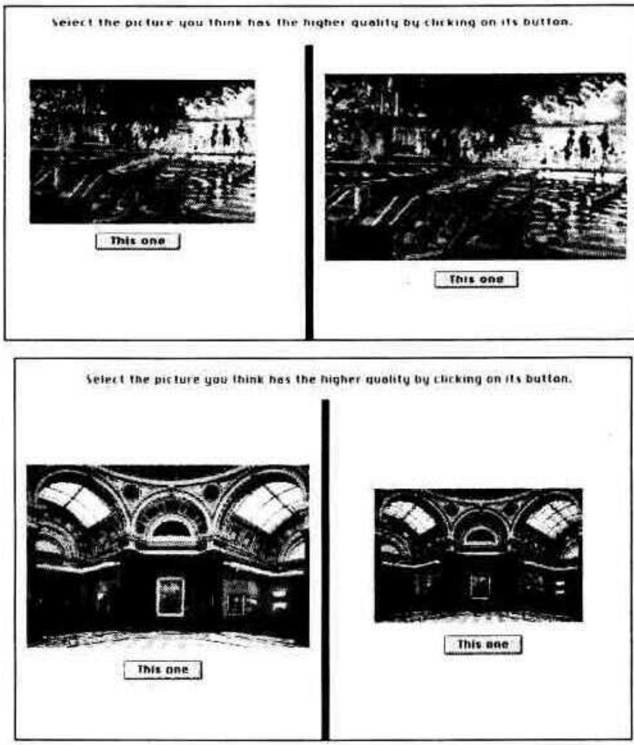

Figure 3. Screen samples from the picture quality experimental treatments. The photographs can be viewed at http://www.extension.usask.ca/Papers/Misanchuk/AECT97/Photos2.html

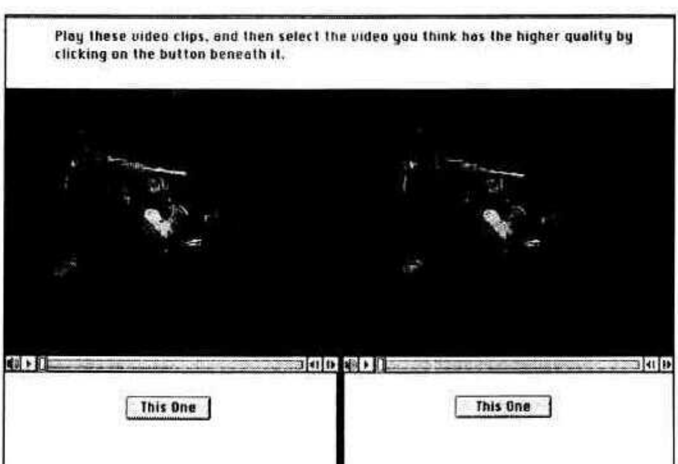

Figure 4. Screen sample from the movie quality experimental treatment. The video can be viewed at http://www.extension.usask.ca/Papers/Misanchuk/AECT97/Videol.html

a farcical dance scene from the puccini opera La Boheme (Hofsteter, 1986), originally recorded to videodisc at the University of Delaware. One of the reasons we selected this segment was because the sound track and dance provided a strong indication of the stability of the synchronization between sound and picture, a feature that was missing in the original pilot study. The digital versions were recorded in quarter screen (160 x 120 pixels) and half screen (320 x 240 pixels) sized windows, and three settings of frames per second (30, 15 and 10 fps) using Apple's Fusion Recorder™ 1.0.2 on a Power Macintosh™ 8100/80AV with 32 Mb of RAM and 2Mb of VRAM. Each movie file was imported into a program created with Authorware Professional v. 3.5 and a series of "pages" were created to provide a paired comparison of every possible combination of movie variables under study. The order of comparison was constructed according to recommendations by Ross (1934) for conducting paired comparisons to eliminate effects of picture location and presentation order. A Thurstone Scale was constructed for comparing resultant data (Torgerson, 1958).

The treatments were administered on two matched Power Macintosh 6100/60AV computers running under System 7.5.5 with 15" Apple Audioscan™ monitors. Headphones were worn for the audio portion of the treatment.

Each subject completed the treatments individually and without consultation. Subjects were asked to compare pairs of images and movies as they appeared on the screen, and judge which image or movie had the higher quality. No definition of the term "quality" was offered: subjects were instructed to employ their own definitions of the term. Selections were made by clicking on buttons beneath each image or movie (see Figure 3). The movies remained on the screen until a selection was made, and both clips remained under the complete control of the subject. They were able to adjust the volume on either clip independently, and they could select and replay portions of the clips and still frames as many times as they wanted.

At the end of each treatment, subjects were asked to describe the criteria they used to make their judgements about quality. They typed their responses, and their comments were saved along with the paired comparison data.

Results (still pictures)

The paired comparisons data were used to construct a Thurstone scale. Figures 4, 5 and 6 are graphic displays of the Thurstone scale points for the picture and movie treatments. One of the advantages of Thurstone scaling is that it provides a method for representing distances meaningfully. Graphically, it is easy to describe the relative positions of the quality ratings. The values and rank orders of Thurstone scale points are presented in Tables 1, 2 and 3 (1 =highest perceived quality rating, 6=lowest perceived quality rating).

File sizes of the picture and movie files were obtained from the "Get Info" system function on Macintosh System 7.5.5.

Table 1: File sizes and quality ratings for treatment with lower amount of detail at two sizes and three bit depths of color.

Window Size and FPS	Movie Size (Mb)	Thurstone Scale Point	Thurstone Scale Ranking
640 x 480 (32 bit)	1528	5.07	1
320 x 229 (32 bit)	876	-3.52	6
640x480 (16 bit)	1039	-1.51	4
320x240 (16 bit)	600	1.46	2
640 x 480 (8 bit)	430	0.92	3
320 x 240 (8 bit)	266	-2.42	5

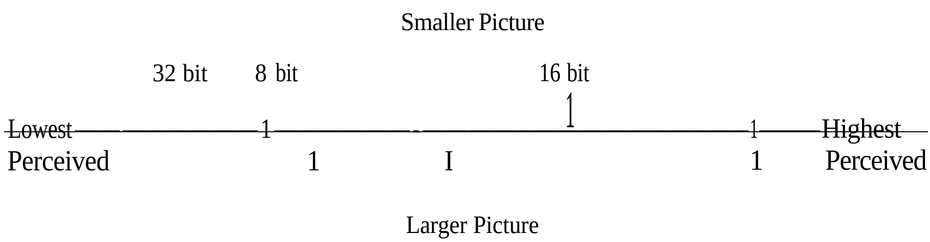

Figure 4. Graphic representation of Thurstone scale points for picture treatment with lower level of detail in the picture at two sizes and three bit depths of color.

For larger pictures in the lower detail treatment, there was a clear preference for the image with the greatest bit depth of colour. Curiously, the 8 bit version of the image was preferred to the 16 bit version of the same picture.

For smaller pictures in the lower detail treatment, bit depth of colour seemed to have little to do with the quality ratings given the pictures. The similar low ratings of smaller 8 and 32 bit pictures could indicate that the two images are inseparable visually. This is not likely, however, given that the smaller 16 bit picture had a higher quality rating than the other two smaller pictures. This unexpected finding was similar to the original finding in the pilot experiment (Schwier and Misanchuk, 1996a), and the difference between these results and the results of the second treatment seems to support our suggestion that this treatment, with its lower level of detail, was similar to the portrait used in the pilot study. At that time, we speculated that there may be an optimal colour depth for different size images, one which takes maximum advantage of the colours available. It could also be that picture size is so influential that the companion variable (bit depth of colour) is virtually ignored. The replication of this finding lends support to these speculations, particularly for images judged to have lower levels of detail or precision.

Table 2: File sizes and quality ratings for different size pictures and different bit depths for treatment with higher amount of detail.

Window Size and FPS	Movie Size (Mb)	Thurstone Scale Point	Thurstone Scale Ranking
640x480 (32 bit)	1313	1.95	1
320 x 229 (32 bit)	716	-0.80	4
640 x 480 (16 bit)	890	0.47	3
320 x 240 (16 bit)	487	-0.89	5
640 x 480 (8 bit)	383	0.56	2
320 x 240 (8 bit)	210	-1.29	6

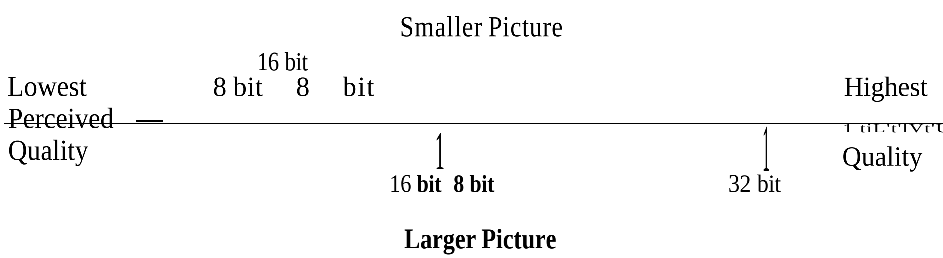

Figure 5. Graphic representation of Thurstone scale points for picture treatment with higher level of detail at two sizes and three bit depths of color.

Generally speaking the Thurstone scale suggests that for larger and smaller pictures, where detail and a high degree of precision is important, the higher bit depth setting is desirable. For 8 and 16 bit images, there appears to be an interaction between image size and bit depth of color. For larger pictures, there appears to be little difference between 8 and 16 bit images. For smaller pictures, there appears to be a preference for 16 bit images over 8 bit images. One might speculate that with larger images, the role of bit depth was less influential for discerning details in the pictures than the size of the picture. For smaller images, greater bit depth may appear to compensate for some of the detail lost by size. This position is bolstered somewhat by the observation that there was a clear preference for larger pictures in the higher detail treatment, with the smaller 16 bit image providing an interesting anomaly.

Comments from subjects cast some light on the data. First of all, it was clear from their comments that subjects attempted to compare the pictures carefully and systematically.

> I tried to select the picture that had the sharper edges and outlines initially. When I could not seem to see much difference in this quality I pay particular attention to the light coming in through the ceiling windows and also the spotlights on the portraits. It seemed that when it was particularly difficult

> to choose between the two, the only way I could decide was through this quality of "light", (subject)

A number of comments indicated that the participants were able to identify different bit depths of color, even though it was articulated in different ways. It was apparent that for the most part, shallower bit depth was interpreted as louder, more garish color. Greater bit depth gave the impression of more natural, pleasant color.

> I also prefer smooth lines and less bright colours as opposed to bright colours and a more digitized look, (subject)

> If I see nothing but bold colours and no in-between colours then I suspect that the picture has no colour depth. The best way I have of telling is to look at a colour that changes shade with distance. The reasons for this is that jagged edges and bold garrish [sic] colours are HARSH! Smooth lines and mellow colours are pleasing, (subject)

> On the orange wall, you could see sometimes the shading change abruptly, (subject)

> Once posterization sets in, except for special effects, these were automatically rejected. (subject)

> I also looked at whether or not the picture looks realistic or like a "computer generated" image. The final aspect I took into account was the shading of the colors. (subject)

There was a great deal of commonality among subjects about what they were cueing on to make selections. Most were looking for clarity, sharpness, and detail- all different terms which we would probably combine under the label "resolution." This seemed to be the preferred variable, regardless of the type of picture.

> When I was making a selection, I was looking at the detail in the boats, the color clarity and the sharpness of the people and objects in the picture. (subject)

> I also paid particular attention to the rowboats in the foreground of the painting and also the reflections in the water. I found that these particular parts of the painting were affected quite significantly in the specific examples. when the color and line quality were blurred or less sharp I was able to make the easiest decisions. (subject)

> I was looking for picture clarity. Crisp detail in the paintings caught my eye. (subject).

The last subject quoted above pointed at the second most often mentioned element - size. All other things being equal, most subjects seemed to prefer larger pictures.

> Size didn't seem to matter but where 1 could not decide I believe I chose the larger picture. (subject)

> I picked those that appeared clearer, crisper, sharper... once again. Where they appear equal, I picked the larger image. (subject)

> I like bigger pictures... no logic at all... went with the gut reaction.... (subject)

> Also, if the "quality" looked the same to me. I chose the bigger picture because I think this one would be harder to achieve. (subject)

> If they appeared to have the same compression ratio 1 chose the larger one because it offered more detail. (subject)

But there was a curious counter point to this perspective. A few subjects occasionally selected smaller images for very specific reasons. Sometimes the reasons revealed a misunderstanding of the medium employed.

> My primary focus was on picture clarity, but where both pictures seemed equal I always picked the larger. In the case of a smaller, clear picture vs a larger pixelated images I always chose the smaller. Interestingly enough, when both small and large pictures were pixelated I prefered the smaller as it seemed to hide the imperfections soemwhat better. (subject)

> In the very first example of different sized photos I saw little or no difference and choose the smaller (just to save on printing costs). (subject)

It was interesting to note that subjects did. on occasion, approach the two different pictures (higher detail vs. lower detail) quite differently. When we designed the experiment, we felt that this might be a significant issue, and a few comments supported that position.

> It was much harder to choose a "higher quality" sample in this section because of the style of the picture. A photograph is easier than an Impressionist painting! I looked for clarity and fuzziness, focussing on the woman (on the left of the trio) in the right side of the painting, the boats, and mainly the leaves on the trees. But I wouldn't dare tell an Impressionist that clarity = higher quality... (subject)

> For some reason, these pictures seemed more clear when they were smaller. I think this is because the picture itself is one in which the objects do not have very definite lines. In these types of pictures, the smaller the picture the clearer they seem to be. In the first set of pictures (the ones of the inside of the art museum) the objects were created with very definite lines and the larger they were, the clearer they seemed. (subject)

Results (video)

For digital video, it appears that the frame speed of the video is more important than the size ofthe window (see Table 3 and Figure 6). Subjects, on the whole, preferred the smaller images, with the exception of the large frame recorded at 10 frames per second. The computers used in this instance were evidently not able to process the higher frame rates (15 or 30 fps) with the larger window; something less than that was actually displayed. They were, however able to process 10 frames per second with the larger window, and came closer to keeping up with all of the frame rates of the smaller window. We speculate that the resultant similarity between the recording rate and playback rate may result in a more fluid appearance to the displayed video. Where the larger and smaller versions were both "fluid," the larger window was preferred.

Table 3: File size and quality ratings for digita video at two window sizes and three fps (frames per second) settings.

Window Size and FPS	Movie Size (Mb)	Thurstone Scale Point	Thurstone Scale Ranking
Quarter (30fps)	11.488	1.74	3
Half (30fps)	41.411	-3.90	5
Quarter (15fps)	6.126	4.42	1
Half (15fps)	20.683	-4.00	6
Quarter (l0fps)	4.092	-0.11	4
Half (l0fps)	14.329	1.85	2

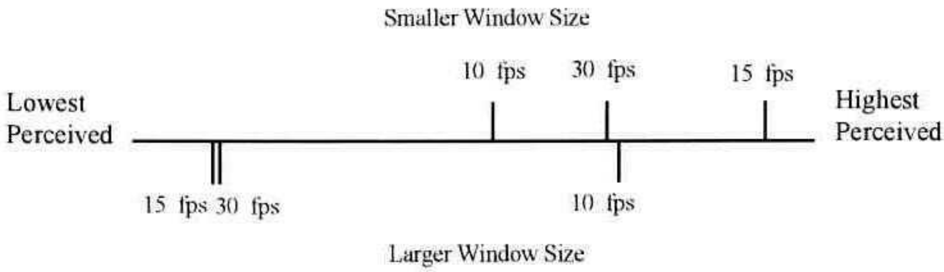

Figure 6. Graphic display of Thurstone scale points for digital video at two window sizes and three-fps (frames per second) settings.

Comments from the experimental subjects supported this interpretation of the data. Subjects grappled with the way the treatment was administered. There were two clips on each screen, and a number of decisions were required for participants to make comparisons. We noted that subjects took considerably longer to arrive at their judgements on the video segments than they did on the still pictures. For each clip, volume had to be controlled, and the video might be started, stopped and positioned for purposes of comparison. This meant that different subjects developed unique strategies for analyzing the data. This may have contributed some contamination to the way(s) video clips were analyzed.

> I first watched the pushing off the knee bit (in the clip on the left) to see if there was blurring of the sleeve as the fellow fell. Then I watched that same two-second bit in the second video, then let the second video run had the kerchief putting-on. Then I stopped the second video and went back to the first to watch the pushing to the kerchief bit. Usually by that time 1 could tell which I thought was clearer, less choppy, more fluid, and had less blur. If in doubt, I would continue watching, or rewind and compare the first scenes. (subject)

At the same time, and all other things being equal, most subjects cued very deliberately on the fluidity and how closely the video and sound were synchronized.

> The major factor I used in evaluating the video segments was the flow or frame rate of the video. If the video was choppy I did not like it. Again the smaller the window the better the frame rate appeared. The actual digital quality of the picture was not as important as the was the action moved. (subject)

> I tended to pick the bigger pictures when I couldn't decipher a change in the audio. (subject)

> The jerkiness of the movement was the key factor. When the one actor pointed his toes in time to the music, you could catch how smooth it looked. Also, when the actor's knee struck the floor, often the sound would be out of sync. The smaller image was able to display more frames per second than the larger one. The smaller of the two always was better (in my opinion). (subject)

Any difference in apparent audio synchrony would probably be an artifact of the computer speed - its inability to keep up with the frame rate. For QuickTime video, audio is always given priority for smooth playback, with video display adjusted to keep pace with the audio. Key frames are set in digital recordings to establish specific points to resynchronize the audio and video. For the experiment, key frames were recorded at half-second intervals, so if the video displays on the treatments were able to keep pace with the audio in all cases, the synchronization would be identical for all treatments. The perceived lack of synchrony in some clips indicate that the computers were unable to keep pace with some of the video segments.

At the same time, the data suggest that synchrony of audio and video may indeed be a critical factor in determining the perceived quality of video. When video has low fidelity (is choppy and indistinct), as is the case with most digital video, it is often very difficult for the viewer to determine which treatment is worse. At these times, the viewer looks to synchrony to make a judgment of video. it is often very difficult for the viewer to determine which treatment is worse. At these times, the viewer looks to synchrony to make a judgment of quality, and this is consistent with other research which suggests that synchrony of audio and video is an important variable in how people evaluate media, whereas video fidelity appears to be insignificant (Reeves and Nass, 1996).

Conclusions

Clearly more investigation is needed before robust guidelines can be formulated to help designers choose optimal (as opposed to maximal) size, colour depth, and frame rates. For the moment, file size and the initial findings reported here can be used as guides, but we caution designers to generalize these findings carefully. Generally speaking, we have found and replicated evidence that for medium speed computers (60 - 80 Hz), digital video can be recorded and used at 10 frps, at a significant savings in storage space, and with positive effects on the perception of users. For still pictures, it is clear that users prefer larger pictures, and 32 bit colour for large images. The results are less clear for 8 and 16 bit images, but there is some evidence that 16 bit colour can be used with smaller images with little or no loss to preceived quality. For smaller images, there seems to be an interaction between the detail of images and preferred bit depth of colour. For those images with more detail, greater bit depth was positively correlated with perceived quality. For those images with less detail. 16 bit colour was highly preferred, and 32 bit colours less preferred - a finding that was consistent with the earlier pilot study (Schwier and Misanchuk, 1996a).

Research Issues

A host of questions and issues face researchers studying the quality of multimedia. As we went through the process of designing and conducting this and a previous series of studies, we grappled with and learned from a number of issues we believe will help other researchers interested in conducting similiar research about the quality of multimedia images, particularly questions about the perceived quality of digital images by users (Schwier and Misanchuk, 1996a. 1996b).

Definitions of interactive multimedia and quality

A definition we have used previously describes interactive multimedia instruction (IMI) as "an instructional program which includes a variety of integrated sources into the instruction with a computer at the heart of the system" (Schwier and Misanchuk, 1993, p.6). It is a fairly common definition and one which is similiar to other definitions of **MI** (c.f., Gayeski, 1993; Schroeder and Kenny, 1995).

But the most elusive element to define is quality. Do we examine the quality of the contribution made by multimedia to instruction? Do we study the quality of aesthetic decisions made in screen design? Deo we examine technical quality of the multimedia elements? Certainly all of these issues, and more, are important in a comprehensive study of interactive multimedia instruction, and any path chosen will impose a series of very different demands on the researcher.

For these experiments, we chose to examine technical quality in isolation, and to the exclusion of most instructional or aesthetic considerations. The defintion of technical quality may be approached from at least two different perspectives: perceived technical quality and actual technical quality. In these experiments, we emphasize perceived technical quality; that is, how adequate is the still picture and the digital video in the opinion of the viewer? This was contrasted with assumed technical quality. For example, video recorded at 30 frames per second was assumed to be of higher quality than video recorded at 15 frames per second.

One measure of perceived technical quality in digital video is how smooth the motion appears to be to the viewer. Digital video is notoriously choppy, and can be very distracting to a viewer who is used to high-quality, commercial video. How should video clips be selected which emphasize the technical characteristics under study? It it simportant to use material which samples the range of technical qualities under scrutiny, but doesn't introduce extraneous variables. For example, in the quality experiments reported here, we selected a video clip which contained almost continuous motion, and had motion running in several directions simultaneously. We knew that motion was a critical variable, and therefore wanted to make sure that we selected a clip which included more than one type and direction of motion, but we also selected a clip that was relatively monochromatic, because we didn't want to emphasize the influence of color in the experimental treatment of video, we were concerned that subjects might be attracted to (or distracted by) particular colors or combinations, or that certain colors might act as contaminants to other motion variables. For instance, a dancer wearing a red sweater who is moving rapidly from one side of the picture to another might cause some flaring of the color which would be interpreted as a problem with the technical quality of the digital video.

There are also actual measures of technical quality, typically represented by increasing file size. File size is an easily obtained technical measurement, yet one which is not less important to instructional designers than perceived quality, due to the intrusive influence of large files on the storage and execution of computer-based instruction. Large files require huge areas for storage, and with many multimedia authoring tools, large image and movie files execute very slowly.

Variable and value selections

Two variables were included in the digital video experimental treatment reported earlier in this paper; window size and frame rate. These were chosen because they can have a profound effect on file size and execution of the clip. Other variables, such as key frames, the compression algorithm (CODEC), and audio quality may be important when audio accompanies the video segment. Key frames are used to synchronize audio with video (the greater the number of key frames, the higher the degree of synchronization).

Should audio be included in video treatments? In most cases, of course, audio is an integral part of a video segment. However, for testing the perception of video quality, we questioned whether audio might be a contaminating variable because of its interaction with a visual. One needs only to view a popular film with and without the audio track to experience how dramatically audio can influence the interpretation of the visual.

result, we chose to study only the visual in video for the earlier experiment (Schwier and Misanchuk, 1996a). But for these studies, we included audio because we wanted to test typical treatments that use the types of settings most developers might employ, and we speculated that developers will most commonly digitize video that has an audio track. Our studies suggested that the synchrony may in fact be an important feature used by viewers to judge the perceived quality of video. Several experimental subjects commented on their attempts to match the timing of the dancers to the timing of the music on the audio track to make their decisions about which video was best. This supports other recent research that suggests that synchrony is not only important, it may be more important to the viewer than the actual quality of the visual (Reeves and Nass, 1996)

Full motion video is played at 30 frames per second. Logically then, when measuring the perceived quality of digital video, the highest quality setting should be 30 frames per second. However, most computers cannot play video at very fast frame rates, especially at larger window sizes, so is it reasonable to do comparisons at faster fps settings? We decided that it was important to record and test video at 30 frames per second, because developers are sometimes unaware of the limitations of computers, or intentionally record video at higher frame rates in anticipation of faster hardware in the near future - hardware which will be able to handle 30 fps video. We included fast frame rates in the experimental treatments in order to determine whether actual quality differences were perceived by viewers on typical multimedia computers of this generation. Other frame rates may be chosen randomly; we have no reason to believe one set of selections would be better than another. We chose to record segments at 15 frames per second and 10 frames per second. Fifteen frames per second was selected because it was close to the average rate of playback on the computer systems we used for testing. Ten frames per second is a very common setting used by multimedia producers, and it results in significant savings in file sizes.

We also speculated that larger movies might be more attractive to viewers than smaller movies, all other things being equal. In our experiment, we limited the different frame sizes (window sizes) to quarters and half frames. These are common sizes used in multimedia presentations, and any larger window sizes result in much slower compression/decompression of the images, which in effect dramatically reduces the frame rate at which video is recorded and played back. Original video which is digitized should be of the highest quality possible, so that any criticisms of quality in experimental treatments are not related to poor source material.

Still pictures share some characteristics with video, but present a strikingly different challenge to the design of an experimental treatment. We speculated, based on a pilot study, that viewers would assign higher quality ratings to pictures which were sharper and had a natural looking range of colors. So for the experiment, we selected pictures which included a wide range of colors and gradient lighting. Still pictures have at least two characteristics which may influence actual and perceived quality-window size and bit depth of color. These variables are easy to control when producing instructional treatments. High quality scanners allow the conversion of photographs into equally high-quality digital source material, and programs such as Adobe Photoshop(™) permit a researcher to accurately control the size, resolution and bit depth of a still image. The production of treatments is dependent on a system's ability to reproduce color, and upon the quality of the monitor, however, so care must be taken to produce and display materials on a system which is capable of creating the images you require. In our experiments, we started with very high quality photographic reproductions and scanned them at the highest quality settings available on an HP ScanJet II cx/T. These images were used as the source material from which we derived two commonly used picture sizes and three different bit depths of color to produce our stimulus materials.

We also confronted the issue of picture content. In the pilot study, a portrait was used, and we wondered whether the type of content could have influenced the results. Portraits don't typically emphasize detail, and can have a softer look than some other types of photographic images. So for these experiments, we first selected an impressionist painting with little detail and precision. The photographic reproduction of the painting captured the vibrant colours, but the style of the image was soft, sweeping and subtle. For a second experimental treatment, we selected a very realistic architectural photograph which contained a high degree of detail and a very sharp focus and depth of field. These two treatments stood in direct contrast to each other, and invited subjects to analyze the images differently.

Contextualized versus decontextualized presentation

It is difficult, yet important, to acknowledge and accommodate the context of multimedia instruction, while attempting to decontextualize technical assessments of quality. Multimedia assumes a context, primarily because anything that exists as part of a multimedia environment by definition coexists with other media. Should the treatments be presented to the subjects in an instructional module? Should the context for comparison be educational? Will it matter whether participants are entertainment media clups or educational media?

In interactive multimedia instruction, we are interested in how several component media converge and contribute to a complex symphony of instruction. But although we are interested in rich, contextual influences of multimedia, technical assessments of quality demand that we isolate characteristics of individual digital media. For example, the quality of a Quick Time movie can be measured reliably, but once the Quick Time movie is embedded in an instructional context, the learner's assessment of its technical quality may vary based on the learner's assessment of the instructional quality. So our position varies, depending on the dependent variable under study. If learning influence or potential is under scrutiny, then an instructional (or at least educational) context for the treatment is essential. If, however, strictly technical assessments are required, we believe it is important to decontextualize the treatment as much as possible, due to the potential contaminating influence of the instructional context.

Presentation and comparison

Should images which are being compared be shown side-by-side for direct comparison, or should they be shown individually, with the subject able to toggle bit depth of color). Many of us have experienced the "television showroom phenomenon" where a wall of televisions reveal a startling difference in picture quality among the various screens, whereas differences are much less apparent when one looks first at one screen and then at another. It is therefore important to provide an experimental context which invites such direct comparison of multimedia.

At the same time, it is impractical and unwise to compare all of the treatment images at one time. For one thing, it would be difficult to find a monitor large enough to display all of the images simultaneously for comparison while preserving the quality a viewer would experience while viewing a desktop monitor. Second, the order effect would be difficult or impossible to manage, as the perception of any single image would certainly be influenced by its position among the cluster of images being compared. Therefore, we recommend a paired comparison approach. In paired comparisons, each image is compared one at a time with every other image, and the viewer is required to select the image with the highest quality. The order in which pairs are presented to the viewer can be randomized, as can the relative positions of images on the screen.

When multimedia treatments are reported, it is important to describe in exhaustive detail the systems used for presentation of the treatment. Microprocessor speeds vary considerably, and can have dramatic effects on multimedia displays. Similarly, perceptual differences may be associated with monitor resolution, colour settings and brightness/contrast settings. When designing treatments, researchers need to consider the colour and clutter on the portion of the computer desktop or background which is showing. Further, where more than one system is used in an xperiment, it is very important for studies to employ matched systems with identical monitors, processors and operating systems.

Summary

Conducting research on the perceived quality of multimedia is difficult, not only because of the elusive nature of human perception, but because of the array of technical and contextual variables that can influence outcomes. Still, with careful attention to technical and experimental details, much can be learned about how viewers respond to various visual treatments, and within precisely defined contexts, inform the work of instructional designers working with multimedia.

References

Gayeski. D. M. (1993). Making sense of multimedia. In D. M. Gayeski (Ed.), *Multimedia for learning: Development application, evaluation.* Englewood Cliffs, NJ: Educational Technology.

Hofsteter, F. T. (1986). *The University of Delaware videodisc music series. Newark,* Delaware: The University of Delaware.

Misanchuk, E. R., and Schwier, R. A. (1996). *Benefits andpitfalls ofusing HTML as a CD-ROM development tool.* Paper presented at the Annual Convention of the Association for Educational Communications and Technology, Indianapolis, IN. February 12-16. 1996.

Reeves, B., and Nass, C. (1996). *The media equation: Howpeople treat computers, television, and new media like realpeople andplaces.* New York: Cambridge University Press.

Roland, A. (1996). *Videos on art.* Peasmarsh, East Sussex, England: The Roland Collection.

Ross, R. T. (1934). Optimum orders for the presentation of pairs in the method of paired comparisons. *The Journal ofEducational Psychology,* 25, 375-382.

Schroeder, E. E., and Kenny, R. F. (1995). Learning strategies for interactive multimedia instruction: Applying linear and spatial notetaking. Canadian *Journal ofEducational Communication, 24(1),* 27-47.

Schwier, R. A., and Misanchuk, E. R. (1993). *Interactive multimedia instruction.* Englewood Cliffs, NJ: Educational Technology.

Schwier, R. A., and Misanchuk, E. R. (1996a). *Assessments oftechnical quality in multimedia: A report ofresearch and discussion ofresearch design issues.* Proceedings of the annual conference of the Association for Media and Technology in Education in Canada, June 2-6, Vancouver, BC.

Schwier, R. A., and Misanchuk, E. R. (I996b). Designing multimedia for the hypertext markup language. *Journal of Interactive Instruction Development, 8(4).* 15-25.

Torgerson, W. S. (1958). *Theory and methods ofscaling.* New York: John Wiley and Sons.

AUTHORS

Richard A. Schwier. Curriculum Studies College of Education, University of Saskatchewan, Saskatoon. SK S7N 5C8

Earl R. Misanchuk. Extension Division University of Saskatchewan Saskatoon, SK S7N 5C8