Multimedia Conferencing: What Cost to Users?

Gillian M. Wilson
Department of Computer Science, University College London
Gower Street, London, WC1E 6BT, UK
g.wilson@cs.ucl.ac.uk
http://www.cs.ucl.ac.uk/Staff/G.Wilson/

ABSTRACT

Videoconferencing over the Internet is developing into a valuable communicative tool, thus assessment methods allowing user requirements to be determined and accounted for in the design of such applications are vital. Subjective rating scales are the main method employed to assess if multimedia quality is sufficient for a particular task, however relying on this alone has drawbacks. Therefore, we are investigating the use of objective methods to assess the *user cost* of different levels of multimedia quality: physiological indicators of stress are being measured. A three-tier approach to usability evaluation is proposed which incorporates task performance, user satisfaction and user cost, to give a meaningful indication of the multimedia quality required by users.

1.INTRODUCTION

The number of networked multimedia applications, like conferencing over the Internet, is increasing constantly. MMC (Multimedia Conferencing) facilitates communication between two or more users through audio, video and shared workspace tools in real-time (Figure 1). It is becoming more widespread due to falling costs of the hardware required and improvements in networks. MMC is viewed as valuable in a large number of areas, such as distance learning, remote business meetings and distributed project collaboration.

It is possible to send and receive audio and video of a high quality, yet this potentially gives rise to an increase in its financial cost for the user. Most users, individual or corporate, will not want to pay more than is needed for this medium of communication. Therefore, specification of the levels of media quality that allow users to complete their tasks effectively and comfortably is essential information for network providers and application designers. In addition the point at which providing increased quality is of no further benefit to the user is important, as this has positive implications for the conservation of bandwidth.

At present, subjective assessment is the most common method of assessing media quality. However, relying on this alone has drawbacks. Thus, we propose a new method to assess the quality of networked applications: physiological indicators of stress are being measured as an indicator of user cost. User cost is then incorporated with user satisfaction and task performance into a traditional HCI three-tier approach to give a more meaningful indication of the impact multimedia quality has upon the user.

This paper presents a review of traditional methods of assessing multimedia quality (section 2), then a rationale for the new assessment method is given (section 3). Section 4 describes four experimental studies which utilise this new method. Conclusions, contributions and future work are discussed in section 5.

Figure 1: Typical multimedia conference set-up

2.ASSESSING MULTIMEDIA QUALITY

2.1 ITU Scales

The ITU (International Telecommunications Union) subjective rating scales are most commonly used to assess quality in this area. These involve a short section of material being played, after which a 5-point quality/impairment rating scale is administered and a Mean Opinion Score (MOS) calculated. However, recent research has raised concerns about their effectiveness in evaluating multimedia speech and video [20, 21]. The main problems are:

- The scales are one-dimensional, thus they treat quality as being a uni-dimensional phenomenon. This approach is questionable as there are many factors recognised to contribute to users perception of audio

[11] and video [8] quality.

- They were designed to rate toll quality audio and high quality video, whereas MMC audio and video are subject to unique impairments e.g. packet loss.
- The scales are mostly concerned with determining if a viewer/ listener can detect a particular degradation in quality, whereas with MMC it is more important to determine if the quality is *good enough* for the task.
- The short duration of the scales means that there is not the opportunity for the viewer/listener to experience all the degradations that impact upon MMC. Subsequently, a dynamic rating scale for video is now recommended by the ITU (ITU- BT 500-8)[9] in order to account for changes in network conditions.
- The vocabulary on the scales ('Excellent, Good, Fair, Poor, Bad') is unrepresentative of MMC quality. For example it is unlikely that it would ever be classed as 'Excellent'.

In order to address these problems, an unlabelled rating scale was devised [20] and studies showed that users were consistent in their quality ratings using the scale. Yet this is, like the ITU scales, a post-hoc method and it is known that these can result in the cognitive effects of primacy and recency. Therefore, a dynamic software version of this scale was developed, QUASS (QUality ASsessment Slider), which facilitates the continuous rating of the quality of a multimedia conference [3] (Figure 2). The drawback with this method is that continuous rating can result in task interference.

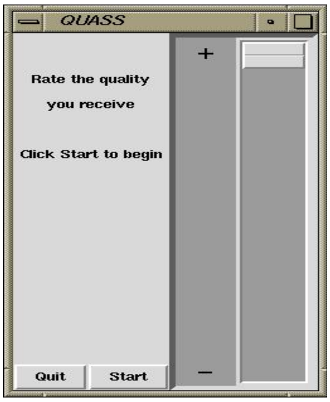

Figure 2: QUality Asessment Slider (QUASS)

2.2 Problems with Subjective Assessment

There are also fundamental problems associated with subjective assessment methods, which center on the fact that they are cognitively mediated. For example, it was discovered that users accepted significantly lower levels of media quality when a notion of financial cost was attached: the accepted quality levels were below the threshold previously established as necessary for the task [2].

Another example of cognitive mediation is given by Wilson & Descamps [24], who showed that the level of task difficulty can affect the rating given to video quality: the same video quality received a lower quality rating when the task being performed was difficult. Therefore, it can be concluded that users may not always be able to accurately determine/judge the quality they need to complete a particular task when contextual variables are operating.

Furthermore, Knoche et al [12] argue that subjective methods are fundamentally flawed as it is not possible for people to register what they do not consciously perceive. As an alternative they recommend that measures of task performance be used to determine how effective the quality is. It is accepted that performance on a task is an essential element of usability, yet to rely on this method in isolation would be to overlook other important effects of the quality on the user.

Subjective assessment methods capture the degree of *user satisfaction* with quality, which is important but not necessarily a reliable indicator of the impact that quality has on the user. Therefore, both task performance and user satisfaction need to be used in conjunction with a measure of *user cost,* as part of a 3-tier assessment approach. User cost is an explicit - if often disregarded - element of the traditional HCI evaluation framework.

3. USER COST

There are subjective methods available to determine user cost via rating scales, yet like all subjective methods they are cognitively mediated. Therefore, we decided to investigate the use of objective methods of assessing the impact of media quality on the user. One way to do this is to monitor physiological responses that are indicative of stress and discomfort. When a user is presented with insufficient audio and video quality in a task context, he/she must expend extra effort on decoding information at the perceptual level. If the user is struggling to decode the information, this should induce a response of discomfort or stress, even if the user is still capable of performing his/her main task. Autonomous physiological responses are not subject to cognitive mediation and collecting such measurements need not interfere with task completion.

3.1 Psychophysiology

The nervous system of humans is separated into the central nervous system (CNS) and the peripheral nervous system (PNS). The PNS comprises the somatic nervous system (SNS) and the autonomic nervous system (ANS). The ANS is divided into the sympathetic and the parasympathetic divisions.

The sympathetic division activates the body's energetic responses. When faced with a stressful situation the ANS immediately mobilises itself without the need for conscious instruction. This is referred to as the 'fight or flight' response [5]. The sympathetic division prepares the body for action by e.g. speeding up the heart rate, dilating the walls of the blood vessels to speed up blood flow to the limbs and releasing glucose into the bloodstream for energy. When the stressful situation has passed, the parasympathetic division takes over to restore the body to its equilibrium.

To measure stress for the purposes of this research, the following signals have been adopted: Galvanic Skin Resistance (GSR), Heart Rate (HR) and Blood Volume Pulse (BVP). These signals were chosen as they are unobtrusive, are good indicators of stress and are easy to measure with specialized equipment - the ProComp, manufactured by Thought Technology Ltd [19] is being used.

3.2 Physiological Stress

Heart rate is viewed as a valuable indicator of overall activity level, with a high heart rate being associated with an anxious state [7]. The function of a rise in HR under stress is to increase blood flow to the working muscles, thus preparing the body for the 'fight or flight' response.

Seyle [18] has linked GSR to stress and ANS arousal. It is also known to be the fastest and most robust measure of stress [4], with an increase in GSR being associated with stress. The precise reason this occurs is not known. One theory is that it toughens the skin, thus protecting it against mechanical injury [23], as it has been observed that skin is difficult to cut under profuse sweating [6].

BVP is an indicator of blood flow: the BVP waveform exhibits the characteristic periodicity of the heart beating: each beat of the heart forces blood through the vessels. The overall envelope of the waveform pinches when a person is startled, fearful or anxious, thus a decrease in BVP amplitude is indicative of a person under stress. This diverts blood to the working muscles in order to prepare them for action, which means that blood flow is reduced to the extremities like a finger or a toe.

3.3 Research Issues

- What aspects of objective quality delivered can be stressful? Factors affecting audio and video quality due to the network, like packet loss, will be investigated along with variables dependent on the end system, such as the color depth of video.
- Is this a reliable way to measure the impact of media quality upon the user? Only further investigation will yield the answer to this question, yet to date the results are promising (section 4).
- How can stress due to the quality be separated from other events which can influence physiology e.g. a cognitive event? The following methods are employed to account for this:

- Baseline measurements are recorded for fifteen minutes prior to any experimentation so that a 'control' set of data is available with which to compare responses under quality conditions, and to allow the sensors and participant to settle down.
- The environment of the experiment is held constant to ensure that events, such as the phone ringing, are not affecting users.
- The tasks used are carefully designed to ensure that they are not overly stressful, yet remain engaging (section 4). All tasks used in this research are taken form the ETNA taxonomy (section 5.1).

4.EXPERIMENTAL STUDIES

4.1 Video Frame Rate

Previous research using subjective assessment methods [1] found that users did not report the difference between 12 and 25 frames per second (fps) when involved in an engaging task - 25fps represents full motion video. If users do not subjectively notice such a difference in frame rate, can it be assumed that it has no effect on them physiologically? Such a finding would have positive implications for the conservation of bandwidth.

To investigate this further, twenty-four volunteers participated in an experiment comparing two frame rates, 5fps and 25fps [25]. The quality difference used in [1] was increased in order to determine if it was noted when made more extreme.

Participants watched two recorded interviews conducted using IP videoconferencing tools on a high-quality computer screen. The interviews were between a university admissions tutor and two school pupils applying to University College London. The tutor and students played themselves in scripted interviews, which had been designed with help of an admissions tutor to reflect common questions and interactions.

The interviews lasted fifteen minutes with each frame rate being held for a period of five minutres. Participants saw two interviews at 5-25-5fps or 25-5-25fps, and were asked to make a judgement on the suitability of the candidates. The frame rate changed twice to counteract any expectancy effect. Audio quality was good and did not change.

For each interview, participants rated the video quality continuously using the QUASS software slider [3]. In addition they had to complete an interviewee assessment form and a questionnaire. The latter addressed how participants felt during the experiment and asked their opinions on the quality and if they noticed any changes. Physiological measurements were taken throughout the experiment. We posited the following hypotheses:

1. There will be different physiological responses to the two frame rates: 5 fps will cause more stress.
2. Participants will not register the frame rate change subjectively.

4.1.1 Results

A Multivariate Analysis of Variance (MANOVA) was performed on the data with the independent variables frame rate and order of presentation. There was no significant effect of order of presentation on any of the signals: GSR ($F_{(1,22)}$=0.383, p=0.542); HR ($F_{(1,22)}$=1.139, p=0.297); BVP ($F_{(1,22)}$=0.680, p=0.418). There was a significant effect of frame rate on each of the signals: GSR ($F_{(1,22)}$=9.925, p=0.005); HR ($F_{(1,22)}$=9.415, p=0.006); BVP ($F_{(1,22)}$=5.074, p=0.035). Examination of the direction of the means (see Figures 3, 4, & 5) showed that GSR and HR significantly increased at 5fps whereas BVP significantly decreased at 5fps. These results are indicative of an increase in stress at 5fps.
However, subjectively only 16% of participants noticed that the frame rate had changed.

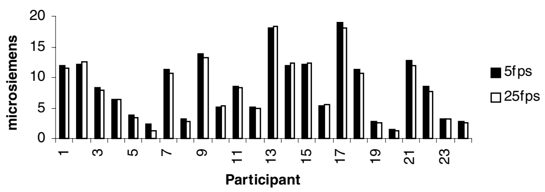

Figure 3: Mean GSR for each participant at 5fps and 25fps.

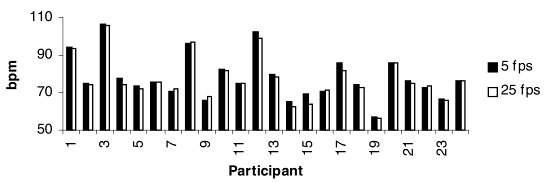

Figure 4: Mean HR for each participant at 5fps and 25fps.

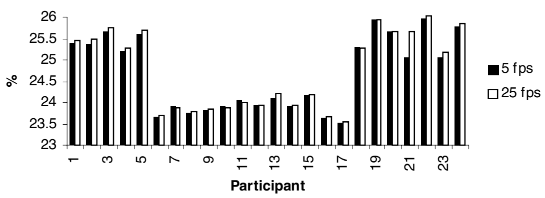

Figure 5: Mean BVP for each participant at 5fps and 25fps

4.1.2 Discussion of Results

The results from this experiment show that there was a statistically significant effect of frame rate on participants' physiological responses in the direction predicted: 5fps caused responses indicative of stress. Thus, hypothesis 1 is supported.
The questionnaire results showed that 84% of participants did not notice the frame rate change subjectively, thus hypothesis 2 and the Anderson et al [1] results are supported.
In addition, there was no significant correlation between subjective (QUASS) and physiological results, which indicates that physiological measurements are tapping into a mechanism that subjective, cognitively mediated, responses do not register.
From this result it can be recomended that designers deliver high frame rates in order to make users more comfortable. Additionally, the three–tier approach to media quality assessment (section 2.2) should be employed to assess the impact of media quality on the user. In this experiment if solely subjective assessment had been used, the important physiological effects on the user would have been overlooked.

4.2 Audio Degradations in a Passive Listening Task It is well known in this area that good audio quality is necessary for effective MMC [10, 17]. To date, it is typically assumed that any problems encountered are due to the network - i.e. packet loss - and that increasing the amount of bandwidth will cure this. Yet, in a large-scale trial where sufficient bandwidth was available, audio problems were still reported in a large number of trials [16]. Therefore, do problems due to hardware affect users as much as those due to the network?
To investigate this further, a joint study was performed [22, 26] which examined six audio degradations resulting from network or hardware effects. The material used was six two-minute segments of a dialogue between two male speakers - this was created from original recordings of multicast meetings with names and locations changed.

Twenty-four participants listened to each of these conditions and rated the quality. The conditions were:

1. **5% packet loss** on both speakers
2. **20% packet loss** on both speakers
3. Audio recorded by one speaker with a **bad microphone**
4. Audio recorded by one speaker that was **quiet**
5. Audio recorded by one speaker that was **loud**
6. One speaker used an open microphone and speakers, as opposed to a headset, which meant that the other speaker generated **echo**

4.2.1 Results
Graphs of the mean responses can be seen in figures 6, 7, & 8.

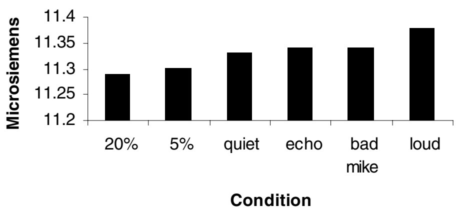

Figure 6. Mean GSR of all participants

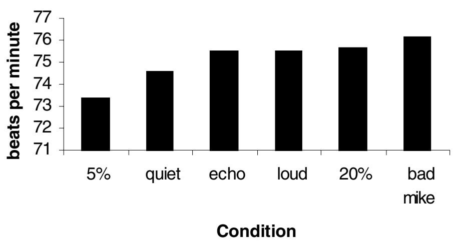

Figure 7. Mean HR of all participants

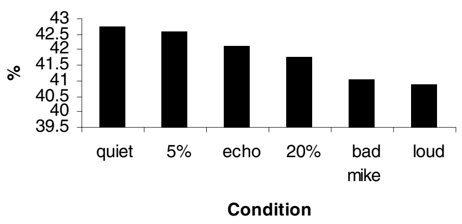

Figure 8. Mean BVP of all participants

A Multivariate Analysis of Variance (MANOVA) was performed on the data with the independent variable audio degradation. There was a significant effect of condition on HR and BVP signals, but not on GSR: HR ($F(5,115)=4.106$, $p=.002$), BVP ($F(5,115)=3.316$, $p=.008$). Pairwise comparisons revealed where the differences were:

•**Bad mike** was significantly more stressful than **quiet** and **5%** loss in both HR and BVP at the .05 level.

•**Loud** was significantly more stressful than **quiet** and 5% loss in both HR and BVP at the .05 level.

•**20%** loss was significantly more stressful than **5%** loss and **quiet** in both HR and BVP at the .05 level

•**Echo** was significantly more stressful than **quiet** in the HR signal only at the .05 level

The finding that GSR did not produce any significant results needs to be explained. The direction of the means corresponds to that of HR and BVP, with the exception being that **20% packet loss** is the least stressful, however the difference between the highest and lowest conditions is tiny: 0.09 microsiemens. It is known in the psychophysiology community, that autonomic signals do not correlate with each other all the time [15]. However, we could suggest that audio degradations do not affect GSR, whereas video frame rate affected GSR more than HR and BVP [25], thus there could be different types of discomfort to media quality degradations - only further research will determine if this hypothesis can be substantiated.

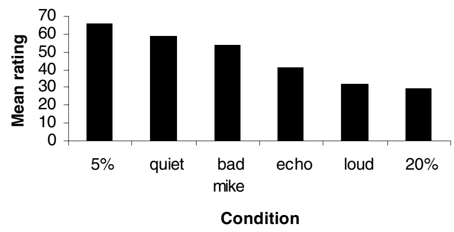

Figure 9. Mean subjective rating of all participants

Analysis of the mean subjective results (Figure 9) showed that there was no significant difference between the **5%** loss and **quiet** conditions (Qobt = 2.39), but that the **5%** condition was rated significantly higher than **echo** (Qobt = 9), **loud** (Qobt = 12.41) and **20%** loss (Qobt = 13.43) at the 1% probability level and at the 5% level for **bad mike** (Qobt = 4.17, Qobt = 4.33). In addition, there was no significant difference between the **20%** loss condition and the **echo** and **loud** conditions at the 1% level (Qobt = 4.43 and 1.02 respectively), despite **20%** loss being subjectively rated the lowest.

To summarise these results, firslty the **bad mike** condition is the first (HR) and second (BVP) most stressful physiologically, yet it is not subjectively rated as being poor. Secondly, subjectively the **20% packet loss** condition was rated as being of the poorest quality yet physiologically this is not the case. Thus, again the discrepancy between subejctive and physiological results can be seen. However, it must be noted that there was a convergence at the other end of the scale where **quiet audio** and **5% packet loss** were subjectively rated as being of good quality and also had no significant impact physiologically.

The implication of this study is that hardware problems can affect people as much as, if not more so, than problems due to the network. Therefore, more attention needs to be focussed on improving hardware set-up and giving an increased awareness of the problem and its causes and effects. In addition, this study provides further support for the three-tier approach (section 2.2).

From the results of this joint experiment, Watson & Sasse [22] recommend that firstly, audio tools incorporate a fault diagnosis option, where users search though a list of terms that describes their problem in terms most commonly generated by users (e.g. fuzzy), and a list of potential actions to remedy this be offered. Secondly, they put forward the idea that designers could offer an expert system style diagnosis on a speech stream to identify likely problems

4.3 Audio Degradations in a Multimedia Conference

The results from the audio degradations in the passive listening experiment (section 4.2) were incorporated in to an experiment using the recorded interview task. The purpose of this was to determine if similar effects emerged when the task was made more engaging and the video channel involved. Twenty-four participants experienced five minutes of normal quality, then the audio quality changed (or vice-versa) to five minutes of:

1. 20% packet loss on both speakers
2. 5% packet loss on both speakers
3. Audio recorded using a bad microphone
4. Audio that was loud

The video frame rate was held at a level established that does not adversely affect people: 25fps [25]. The order was randomised and counterbalanced. The results from this experiment are being analysed at present.

It is recognised that that the bad mike and the volume conditions in the two audio degradation experiments are subjective, as they have not been quantified. In addition a microphone that produces ‘bad’ audio with one soundcard will not neccesarily produce ‘bad’ audio with another. However, they were vital to investigate as a previous field trial [16] indicated that they do affect users – e.g. users reported ‘tinny’ or ‘hummy’microphones - and this needed to be considered in a lab-based setting. Now it has been established that these parameters do impact upon users (section 4.2), further resarch needs to be conducted in order to quantify the conditions and to investigate them in a fine-grained manner.

4.4 Interactive Task

As part of the ETNA project (section 5.1) an interactive task was performed with two levels of both audio and video quality. The participants were eleven undergraduate admissions tutors at University College London whose task was to interview four candidates at Glasgow University, UK. The candidates were actors and the interviews were conducted over the network in real-time.

In each interview the audio quality was either low (15% packet loss) or high (0% packet loss) and the video frame rate was either low (5fps) or high (20fps). Physiological, subjective and objective data regarding network behaviour were gathered and are being analysed at present. The conditions were:

1. High video, high audio
2. High video, low audio
3. Low video, high audio
4. Low video, low audio

The order of presentation was counterbalanced and randomised. This study is an important step-forward in this research for two reasons. Firstly, the task was interactive as opposed to passive: this will allow the effects of the quality levels upon users when they are performing a task with cognitive demands to be established. The nature of the task means that more sophisticated analyses need to be performed on the physiological data - to do this we are extracting a number of features from the signals then entering those features into statistical analyses.

Secondly, we varied both audio and video together, as opposed to looking at their effects in isolation. Therefore, we will be able to determine the interactive influence of one upon the other.

5.CONCLUSIONS

Three main conclusions can be made from this research. Firstly, different levels of media quality cause different physiological responses in users and can be detected through common physiological measurement techniques.

Secondly, subjective assessment and measures of task performance do not pick up all the effects of poor quality in the short-term, e.g. in an hour-long experimental study. It is possible that the negative effects of poor quality would emerge from these assessment methods in longer-term studies, yet for laboratory-based experiments physiological responses give a more instant account of how the quality affects the user. We therefore argue that the 3-tier approach to multimedia quality assessment, (section 2.2) should be utilised to determine if a certain level of media quality is usable.

Finally, we suggest that the largely neglected element of user cost should be given due attention in usability evaluation of any technology, and that objective measures - such as physiological responses - may be more reliable indicators of user cost than subjective methods, which are cognitively mediated.

Critics of this approach may argue that it is not proven that stress responses are a reliable indicator that a factor - e.g. a level of media quality - is actually bad for the user. In our view, it is reasonable to assume that a significant deviation between conditions in the direction of stress indicates that the user has to work harder, and that this might manifest itself in a usability problem with prolonged use. At a time where the negative effects of stress in the workplace are debated, indications that a particular aspect of technology - such as the level of video quality - may be inducing stress deserves further investigation.

5.1 Contributions

Our continuing research in this area is working towards two substantive contributions.

Firstly, the minimum levels of multimedia quality for certain tasks at which users can successfully perform, without significant user cost, will be determined. The impact of problems caused by the network will be investigated. However, quality is not uni-dimensional and encompasses more than variables effected by the network. Thus, the effects of other contributing factors must be examined. This will allow network providers to allocate resources with end users' requirements clearly specified, thus improving applications for them.

These findings will be incorporated into the ETNA (Evaluation Taxonomy for Networked Multimedia Applications) Project, which aims to produce a taxonomy of real-time multimedia tasks and applications and their corresponding quality requirements. This will greatly assist network providers and application designers, as they will have guidelines on the quality they need to deliver for specific tasks.

Secondly, we are working on providing physiological feedback to the user whilst they are engaged in an application. For example, this could be a happy face in the corner of the screen that changes to sad when the user begins to show signs of stress. Such feedback would give an increased awareness to users of their physiological state and would allow them to act upon this information if neccesary.

Finally, a methodological contribution will be made: guidelines stating the most appropriate physiological measurements to indicate a specific impairment in quality will be produced. This will pave the way for much needed further research in this area.

5.2 Future Applications

An idea for a future application of this work is that a utility curve could be built. Utility curves provide a mechanism by which the network state can be related to the end user. They are usually formulated by the results of subjective assessment, however by using physiological measurements an adaptive application could be built. This would enable the application to receive continuous feedback on the state of the user. In the future a user 'wearing' a discrete computer, like those being developed at Massachusetts Institute of Technology Media Lab [15], could have their physiological responses fed into a videoconferencing application. If the computer detected that the user was under stress, it would automatically adjust the variable of the videoconference causing stress to reduce user cost and increase user satisfaction. If network congestion was occurring, the computer would then refer to the utility curve to deliver the next best quality possible.

Discussions with British Telecommunications (BT) are in progress about the possibility of using physiological measurements as a method of stress detection to evaluate a new interface. The MUI (Motivational User Interface) was developed by Bournemouth University and BT's Bournemouth '150' call centre [14]. It aims to motivate and provide feedback to call centre operators, thus reflecting their positive attitude back to the customer. BT are interested in determining if operators are put under more or less stress when using the MUI, as opposed to the traditional interface.

This example of industrial interest illustrates that the ability to detect discomfort and stress unconsciously has wide-ranging implications in product assessment and also in areas like teaching, stress control and providing 'emotionally sympathetic' user interfaces.

ACKNOWLEDGEMENTS

My grateful appreciation is given to my supervisor, Dr Angela Sasse. This research is funded by an ESPRC CASE studentship with BT Labs.

REFERENCES

[1] Anderson, A., Smallwood, L., MacDonald, R., Mullin, J., Fleming, A. & O'Malley, O. (2000), *"Video Data and Video Links in Mediated Communication: What Do Users Value?",* International Journal of Human Computer Studies 52(1), 165-187.

[2] Bouch, A. & Sasse, M. A. (1999), *"Network Quality of Service: What do Users Need?",* Proceedings of the 4th International Distributed Conference, 22nd – 23rd September 1999, Madrid.

[3] Bouch, A., Watson, A. & Sasse, M. A. (1998), *"QUASS – A Tool for Measuring the Subjective Quality of Real-time Multimedia Audio and Video",* in J. May, J. Siddiqi & J. Wilkinson (eds.), HCI '98 Conference Companion, pp.94-95, 1st – 4th September 1998, Sheffield, UK.

[4] Cacioppo, J. T & Louis, G. T. (1990), *"Inferring Psychological Significance from Physiological Signals"*, American Psychologist 45(1), 16-28.

[5] Cannon, W.B. (1932) *The Wisdom of the Body*. (Reprinted 1963.) New York: WW Norton.
[6] Edelberg, R. & Wright, D. J. (1962), "*Two GSR Effector Organs and their Stimulus Specificity*", Paper Read at the Society for Psychophysiological Research, Denver, 1962.
[7] Frijda, N. H. (1986), *"The Emotions, chapter Physiology of Emotion",* Studies in Emotion and Social Interaction, Cambridge University Press, Cambridge, pp124-175.
[8] Gilli Manzanaro, J., Janez Escalada, L., Hernandez Lioreda, M. & Szymanski, M. (1991), *"Subjective Image Quality Assessment and Prediction in Digital Videocommunications*", COST 212 HUFIS Report.
[9] ITU-R BT.500-8 Methodology for the subjective assessment of the quality of television pictures: http://www.itu.int/publications/itu-t/iturec.htm
[10] Kawalek, J. (1995), "*A User Perspective for QoS Management*", Proceedings of 3rd International Conference on Intelligence in Broadband Services and Network, IS & N 1995, Crete, Greece.
[11] Kitawaki, N. & Nagabuchi, H. (1998), "*Quality Assessment of Speech Coding and Speech Synthesis Systems"*, IEEE Communications Magazine, October 1998, pp.36-44.
[12] Knoche, H., De Meer, H. G. & Kirsh, D. (1999), "*Utility Curves: Mean Opinion Scores Considered Biased*", Proceedings of 7th International Workshop on Quality of Service, 1st - 4th June 1999, University College London, London, UK.
[13] Lacey, J.I. & Lacey, B.C. (1958), *"Verification and Extension of the Principle of Autonomic Responses Stereotypy",* American Journal of Psychology, 71, 50-73.
[14] Millard, N., Coe, T., Gardner, M., Gower, A., Hole, L. & Crowle, S. (1999), *"The Future of Customer Contact",* British Telecommunications Technology Journal, http://www.bt.co.uk/bttj/vol18no1/today.htm
[15] Picard, R.W. & Healey, J. (1997), *"Affective Wearables*", Personal Technologies, 1(4), 231-240.
[16] PIPVIC-2 http://wwwmice.cs.ucl.ac.uk/multimedia/projects/pipvic2/
[17] Sasse, M.A., Biltung, U., Schulz, C-D & Turletti, T. (1994), "*Remote Seminars through Multimedia Conferencings: Experiences from the MICE Project"*, Proceedings of International Networking Conference, pp251/1-8, 13th-17th June 1994, Prague, Czech Republic.
[18] Seyle, H. (1956), *"The Stress of Life*", McGraw-Hill.
[19] Thought Technology Ltd. http://www.thoughttechnology.com/
[20] Watson, A. & Sasse, M. A. (1997), "*Multimedia Conferencing via Multicast: Determining the Quality of Service required by the End User*". Proceedings of AVSPN '97 – International Workshop on Audio-visual Services over Packet Networks, pp.189-194, 15th - 16th September 1997, Aberdeen, Scotland, UK.
[21] Watson, A. & Sasse, M. A. (1998), *"Measuring Perceived Quality of Speech and Video in Multimedia Conferencing Applications",* Proceedings of ACM Multimedia '98, ACM New York, pp.55-60, Bristol, UK.
[22] Watson, A. & Sasse, M.A. "*The Good, the Bad, and the Muffled: the Impact of Different Degradations on Internet Speech",* To appear in Proceedings of ACM Multimedia 2000, 30th October – 3rd November, Los Angeles, California.
[23] Wilcott, R. C. (1967), *"Arousal Sweating and Electrodermal Phenomena*", Psychological Bulletin 67, 58-72.
[24] Wilson, F. & Descamps, P. T. (1996), "*Should We Accept Anything Less than TV Quality: Visual Communication"*, Paper presented at International Broadcasting Convention, 12th – 16th September 1996, Amsterdam.
[25] Wilson, G. M. & Sasse, M. A. (2000) *"Do Users Always Know What's Good For Them? Utilising Physiological Responses to Assess Media Quality"*, To appear in Proceedings of HCI 2000, 5th – 8th September, Sunderland, UK.
[26] Wilson, G. M. & Sasse, M. A. (2000) "*Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Methods"*, Paper submitted to OZCHI 2000, 4th – 8th December, Sydney, Australia.