SHOULD WE ACCEPT ANYTHING LESS THAN TV QUALITY : VISUAL COMMUNICATION

F Wilson[1] and P T Descamps[2]

Interaction Design Ltd[1], BARCO NV[2]

ABSTRACT

This paper surveys experimental and empirical evidence which demonstrates that users of full service networks will not be happy with less than TV quality images, especially for personal communication. The paper briefly examines the technical consequences of these findings, and the implications for mature broadcast networks looking towards new service opportunities. The conclusions of this paper contribute to a more detailed exploration of the larger set of technical considerations.

INTRODUCTION

Human communication is a fundamental social activity and has been supported by many of our technological achievements. The advent of multimedia communications technologies, and in particular audio-visual telecommunications, begs some consideration of the visual component of communication and its functional requirements. The inclusion of the moving image may be the most innovative aspect of multimedia communications, and understanding the requirements of humans in utilising these for visual communication may allow us to make best use of the many new opportunities, and to properly address this new area of technical evolution. This paper reconsiders some existing studies in the light of recent extensions of the work and a new analysis of the results. Some conclusions are drawn concerning the consequences for network development which are considered further in a companion paper (Descamps and Wilson (1)).

APPLICATIONS FOR FUTURE NETWORKS

The future Integrated Broadband Communications (IBC) networks for Europe are as yet undefined, but developments include consideration of B-ISDN, ATM, Cable TV, Switched Analogue, and so forth. A key issue in the current debate concerns the applications to be supported by such networks. These applications are the very reason for network usage, and so will be a major determinant of suitable architectures and techniques. It is not surprising, therefore, that research projects such as IBCoBN (Integrated Broadband Communications on Broadcast Networks - ACTS Project AC101) set out to define, implement, and verify applications with real market potential so as to provide a solid basis for evolution towards full service networks (FSN) from mature cable networks.

Consideration of analyses of applications in RACE and ACTS have shown that single services often cannot justify the costs of usage. For example Hopkins et al (2) show that Video on Demand (VoD) may be just such a case. Mature networks preparing for evolution to FSN are therefore considering the whole range of potential services, from Conversational Services (e.g. videotelephony and those value added services supported by it), through Interactive Informatics Services (e.g. Internet Communications and Web Access), to Server Services (e.g. Video on Demand). From these they must identify a 'service menu' which will justify and support the costs of evolution, and deliver applications meeting the real demands of residential and business users alike.

A key element identified in the IBCoBN studies is that of video-communication. This includes desktop or residential videotelephony (VT) and commercial videoconferencing (VC), not only as single services (discreet) but also as elements (combined) in complex services where visual contact may be an intrinsic part of the service menu (for example, teleshopping with options for face-to-face discussions when help or advice are needed).

Whether offered as a discreet or combined service, video-communication is one of the most technically demanding elements of FSN. Video-communication is not only more demanding for the technology, but also for the user, and previous work has shown that the quality requirements of video-communication are quite critical (e.g. Wilson et al (3)). Human communication requires a broad bandwidth of information and a real-time facility if it is to be effective and satisfactory, and our consideration of some empirical evidence will now show that there are many related factors in human communication which add up to a demand for very high quality of service.

EVIDENCE FOR QUALITY REQUIREMENTS FOR VIDEO COMMUNICATION

A number of related studies are considered here, and then some further analysis of their findings is offered for consideration. The interested reader is referred to the source texts for detailed experimental descriptions and statistical treatments of the various data sets.

Studies of effects on simple features of human performance in information processing can easily show that temporal discontinuities in image (such as flicker) affect ability to extract and utilise information (e.g. Steinbach (4)). However, there is great difficulty in generalising from such discreet studies to effects in general usage of information sources for business or leisure purposes. More targeted work (e.g. Frowein et al (5)) has attempted to examine effects of frame rate and other qualities in the context of information reception behaviours directly involving CRT transfer media. However, in that study restricted human behaviours were employed for experimental control, and while a frame rate was proposed for effective human speech reception (15 fps), there may be room for doubt as to the validity of the findings for application in real world settings involving less restricted behaviours, and so also for the claim that this rate and performance are achievable via 128kilobit.sec H261 codecs. (see (3) for full argument).

More recent studies of realistic usage have attempted to clarify quality requirements for normal 'free field' behaviours, but have either concentrated on quality parameters specific to a targeted enquiry (e.g. Wilson et al (6), who establish bandwidth required for spoken English reception using ISDN and H320 codecs), or have focussed behaviours which are not well related to general public usage of video-communication (e.g. Wilson and Selby (7) who look at lip reader performance in ISDN/H320 usage). These latter two studies ((6) and (7)) are reconsidered by Wilson and Selby (8) who have re-analysed the material to answer the frame-rate issue for free field behaviour, and material from this later study is considered here.

Experiment 1 - An experiment was conducted to determine the necessary bandwidth and codec requirements to provide real 'added value' in speech reception for normal hearing people, and to examine the user satisfaction at various bandwidths. A set of six different task types were designed to cover a wide range of normal communication behaviours (message dictation, story interpretation, word lists, etc.). For each task type a set of task instances was enacted and recorded on SVHS PAL to provide a 26 task source (see (6) for a full account). Each task was then processed via an experimental H261 codec* provided by BT and re-recorded in CIF at 64 kilobit.sec, 128 kilobit.sec, 384 kilobit.sec, and 2048 kilobit.sec, (5 qualities of image and 6 experimental conditions including Audio Only, and SVHS PAL). A study was conducted whereby 30 users completed all tasks such that all quality conditions were experienced by each user in a balanced design. Degraded audio was used to ensure that information from the visual stream was required for good performance, and the results showed a steady increase in performance with bandwidth. *(*this codec was benchmarked as comparable/slightly better than current market equipment at its 1993 test date).*

Experiment 2 - A similar experiment to the above was later conducted with deaf people who are lip readers (7). In this case the audio level was set at an estimated 'normal' level for hearing people, and a similar experimental procedure again showed a steady increase in speech reception performance with increase in bandwidth.

Analysis of Results - Both experiments showed that while significant performance gains could be made from ISDN2 systems (128kilobit.sec), a rate of between 384k and 2Mbit.sec was required to approach optimal usage of the visual component of the tasks for hearing people, while deaf lip-readers continued to gain from increased quality.

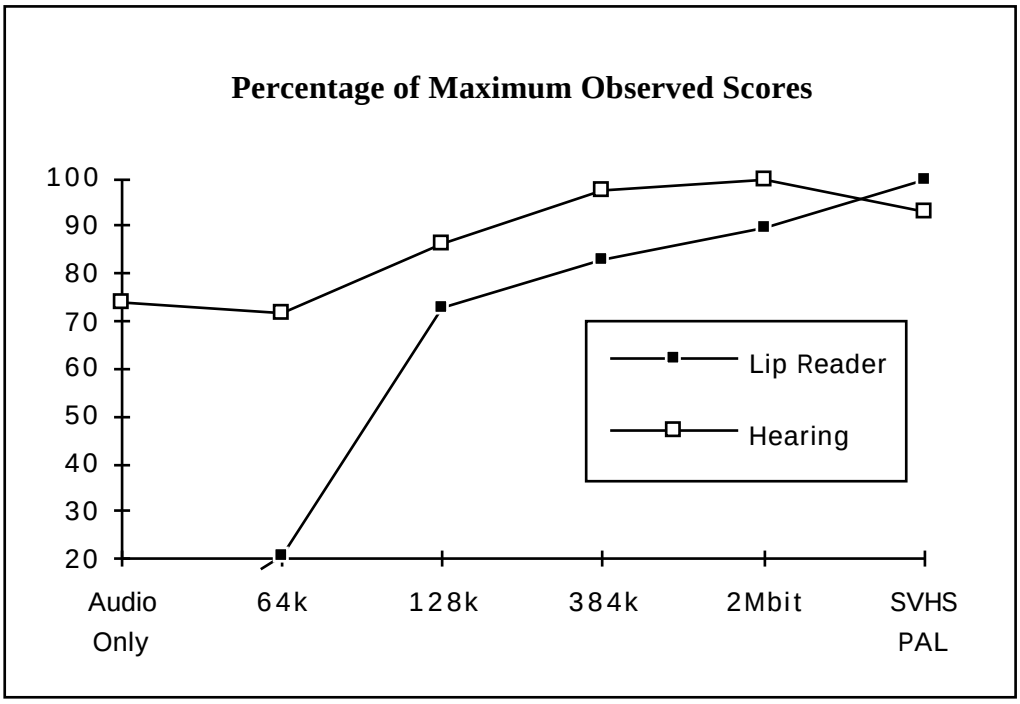

Diagram 1 : Percentage of Maximum Observed Scores over all tasks and subjects in both experiments.

Diagram 1. shows the combined results of these two experiments where the highest recorded actual scores were 63% of the possible maximum for hearing subjects and 32% of the possible maximum for deaf subjects (difficult tasks). The scores are rendered in the diagram as percentage of those "maximum observed scores" to allow easier comparison.

The drop off in performance for hearing people over 2048 kilobit.sec was attributed to 'distraction' by some experimental subjects - apparently the 'smoothed' image at 2 Megabit was better for these very difficult tasks. However, the deaf lip-readers relied on very subtle facial information and so still gained from its retention in the original SVHS PAL version.

The post-test analysis of the source materials has since been conducted and shows that the frame rates generated in the different quality conditions ranged from 2.1 fps to 25 fps.

Table 1 details the results obtained from this analysis for all quality levels (8).The ranges obtained indicate the variability of behaviour in system subjects performing similar tasks for recording, as well as differences based on task structure. The 'best case' was a rather solid young athlete in plain attire who spoke his lines with little movement. He developed nearly 8 fps at only 128k. However, one young lady reduced the system to half that frame rate for a similar task/bandwidth combination - she had quite a lot of 'lively' hair and this, combined with her floral dress and mobile jewellery, provided a much higher image processing load for the system.

Table 1 - Mean frame rates and ranges obtained at different bandwidths for tasks.

BANDWIDTH	MEAN FRAME RATE	RANGE fps
64 kilobit.sec	2.55	2.1 - 3.9
128 kilobit.sec	5.09	4.2 - 7.9
384 kilobit.sec	15.9	12.7 - 19.3
2048 kilobit.sec	25	~25

These data suggest that performance in speech reception is greatly assisted by the visual component of videotelephony. The gains were seen to maximise at between 16 and 25 frames per second in CIF format for hearing people, while deaf people were seen to continue to make gains from 25 fps images at above CIF resolution (SVHS PAL).

Subjective Quality Assessment - As part of the experiments cited, users were asked after each task trial to estimate the quality of image as compared to "TV Quality" on a 10 point scale. Diagram 2 shows that the subjects with normal hearing indicated a subjective assessment at below TV quality even for the SVHS PAL material. However, since there was a correlation between this assessment and observed task quality (*suggesting they tend to blame difficulties on the image*), the extreme task difficulty may be a reason for the depressed score even at 25 fps.

The deaf users reported a steady increase in perceived quality with increase in frame rate / bandwidth, and this may reflect that they are more used to dealing with difficulties in speech reception. It could be argued that the subjective estimates would be higher for the hearing subjects in the absence of the discomfort generated by the very difficult experimental tasks.

Summary of evidence - The evidence presented here is derived from sources which together suggest a number of important issues for service provision in FSNs with applications involving the use of desk-top conferencing, videotelephony, and videoconferencing. Firstly, the data clearly indicate that human behaviour is quite variable, and requires a service level which accommodates the most demanding in terms of image processing. It would also seem that human behaviour varies in terms of the level of subtlety in the actions of the performance. This is pointed up by the lip readers who are extracting more and more useful information when the hearing participants are no longer making gains. This serves to emphasise the need for retention of detail for any user who may require it (e.g. medical, engineer, etc.). The general case seems to be one of demand for higher quality than is available from current ISDN networks, and so future usage of moving image as part of information and communication systems must take careful account of whether the quality provided actually adds something useful to communication.

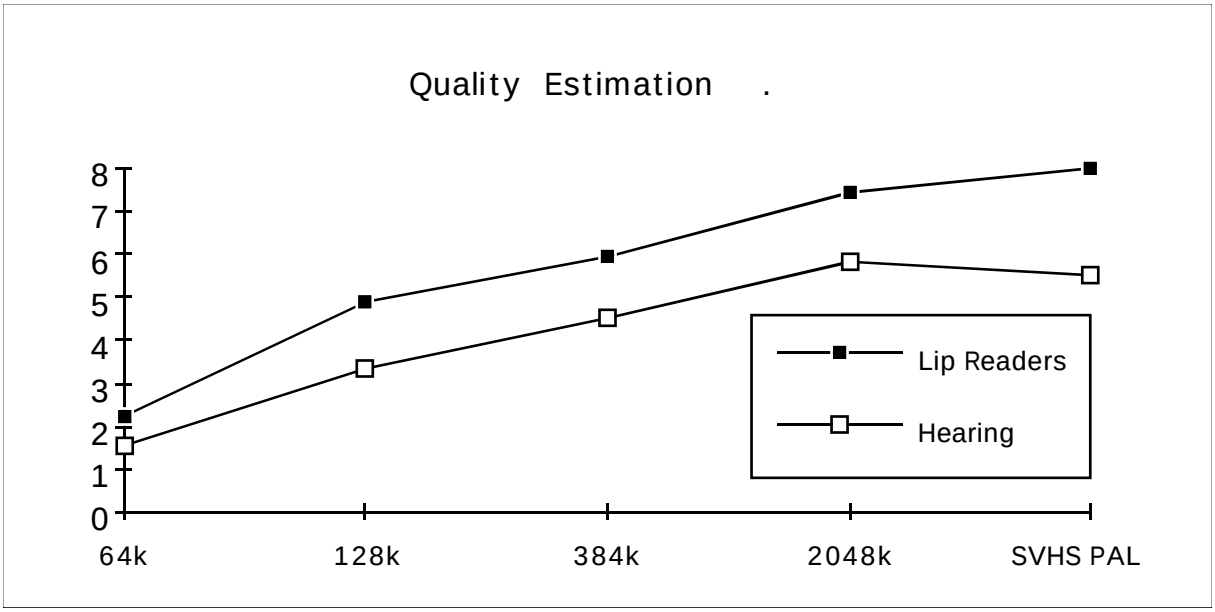

Diagram 2 : Subjective quality estimates as a function of frame rate and bandwidth.

TECHNICAL ISSUES FOR FULL SERVICE CABLE NETWORKS

As considered earlier, a single application such as videotelephony cannot carry the full costs of digitising an analogue broadcast CATV network or its conversion to carry fully symmetrical services (return path). However, sufficient evidence exists to indicate that videotelephony is flexible in providing a component of more complex services, and other applications may be introduced more easily when videotelephony has been enabled since videotelephony is technically quite demanding (fully symmetrical real-time high bandwidth service).

Given that the market sector for the CATV networks is the residential user, and the user would prefer TV quality moving images when these form part of any service, be it broadcast TV or videotelephony, then it may easily be that video telephony will form a significant part of the financial driving force for the evolution of CATV networks. In this case, it will also be likely that CATV networks should progressively implement fibre to the curb as a key part of their evolutionary strategy.

The main technical issues are:
• HFC architectures, Headends and TMN: how can mature cable networks evolve in order to bring a good mix of distribution, information access, and conversational broadband tele-services?
• service layers: how can a flexible, efficient and robust modulation method for the return channel be standardised, and an optimum multiple access protocol be introduced?
• how will a broadband upstream capability be introduced in the terminals (set tops)?

In order to answer such questions, two groups of broadband teleservices must be addressed:
• symmetrical real time services (quality video telephony) and cable LAN,
• asymmetrical fast data services (e.g. WWW).

With these issues in mind, the ACTS project AC101-IBCoBN has been formed to provide a forum and strategy for exploration of such themes, and demonstration of potential technical solutions. It is the main cable project in ACTS which investigates CATV as a future-proof communications platform by providing quality videotelephony through a representative group of cable operators in seven EU member states and in Russia. IBCoBN uses two measurable objectives for planning future large trials:
1)to implement Integrated Broadband Communications on broadcast networks within 5 years.
2)to identify and verify a scenario for transforming mature and new cable TV networks to Integrated Broadband Communication Networks.

In a companion paper in these proceedings (1) a fuller description of the strategy for transforming CATV networks to Integrated Broadband Communications is considered. This uses the introduction of video communications and teleservices utilising videotelephony, and is described in terms of the technical issues and the existing experiments which inform the technical considerations.

CONCLUSIONS

From the preceding considerations, we conclude that whilst no single service will justify the costs of network evolution from broadcast CATV to FSN, the added value of videotelephony and its utility as a component in more complex service offerings make it a candidate for inclusion in any 'core service' menu. The experimental findings considered in this study suggest that to implement videotelephony successfully, and in a manner which responds to real requirements, the video quality must approach TV quality. Existing low quality solutions such as ISDN2 may interest and stimulate the market, but delivery of real quality over FSNs will be required to really exploit the larger market for videotelephony and the many videotelephony-assisted applications.

REFERENCES

(1) Descamps, P.T. and Wilson, F. 1996 "Adapting Broadcast Networks for Broadband Interactive Services", in Proceedings of IBC'96, IEE, Amsterdam.

(2) Hopkins, M., Connell, S., Cools, J.P.,Esteban, D., Seetzen, J., and Kellner, M. 1994 "Guidelines for Introduction of Advanced Communications", RACE Project 2091 - URSA, deliverable D16.b1, Analysys Ltd, UK.

(3) Wilson, F., McEwan, J. and Selby, C. 1995 "Evaluation Report on the Success of the ACEs and the Concept", RACE Project 2033 - TELECOMMUNITY, deliverable D12.b1, IDL, UK.

(4) Steinbach, M.J. 1969, "Eye tracking of self moving targets : The role of efference" Journal of Experimental Psychology, 82, 366-376.

(5) Frowein, H.W., Smoorenburg, G.F., Pyfers, L. and Schinkel, D. 1991. Improved speech reception through videotelephony: an experiment with the hard of hearing. IEEE Journal on Selected Areas in Communications, 9, 611-616.

(6) Wilson, F., Wakeman, I. and Smith, W. 1993 "Quality of Service Parameters for Commercial Application of Videotelephony", Human Factors in Telecommunication - HFT'93, R V Deckers verlag, Darmstadt, Germany.

(7) Wilson, F. and Selby, C. 1995 "Final Report of the RACE 2125 Project -ADEPT", ADEPT project deliverable D11.b1, IDL, UK.

(8) Wilson, F. and Selby, C. "Human Communication in Desktop Videotelephony" - submitted to IWC journal of the British Computer Society for publication in 1996.

(9) Wilson, F. 1994 "Description of a Generic IBC Application : Conversation" RACE Common Functional Specifications and Common Practice Recommendations, CEC DGXIII, Brussels.