Quality of Service for Mobile Multimedia Communication

Video Telephony for sign language

Pascal Dugénie, Amardiya Sesmun, Liang Q.Liu,
Alistair T. Munro, Michael.H.Barton
Centre for Communications Research, University of Bristol

Abstract

Access by anybody and support of multimedia are amongst the leading goals of UMTS. The ACTS project AC027, UMPTIDUMTI has carried out work on Videotelephony for Sign Language (VSL) since it is the most usual and natural means of communication between deaf people. Mobile terminals are of special interest to them because they are unlikely to find specific video equipment when they are on the move [6].

Notable improvement in video compression techniques has created new perspectives for video transport services. Unfortunately, even with the best achievements, video is still a very luxurious medium, especially on public mobile networks for which the effective bandwidth is significantly dearer compared to fixed networks.

Therefore, this study is a pragmatic approach to optimise the cost of a video telephony service for sign language. The main objective is to determine the effective capacity required at minimum thresholds of subjective acceptability.

The methodology applied aims primarily to estimate the capacity requirements of the compressed video signal when a single person performs sign language. The compression techniques are those used in conventional video telephony. At the University of Bristol, the Centre for Communication Research collaborated with the Centre for Deaf Studies to elaborate five sign language sequences recorded in optimum conditions and used throughout the testing.

These video sequences have been played back to a panel of ten deaf users. At this stage, only the frame rate and the image resolution were variable. An ideal channel was assumed.

The analysis of the subjective evaluation reported here indicates the critical areas, in which begin the difficulties of sign language comprehension. A number of signs were misunderstood because of an insufficient resolution for the lip reading or at slow frames rates because of a lack of smoothness in the hands movement.

These results aim to provide a application model for the fixed and mobile network operators, in particular the future UMTS operators, who are intending to offer ad hoc services for the deaf community.

Acknowledgement

In reporting these results the authors are grateful for the support of:

- The Commission of the European Union through the UMPIDUMPTI project in the ACTS Programme; and
- The Mobile Virtual Centre of Excellence, a UK company dedicated to research into future mobile communications and funded by UK government and the mobile communications industry.

We are grateful in particular for the contribution of the members of the UMPTIDUMPTI project to this work.

Background: Aspects of quality of service

General definition

The term Quality of Service is used to designate a set of parameters which are intended to represent measurable aspects of the subjective "perceived quality", but not on the causes of this perception **[1]**.

The criteria taken into account by the user to judge a service change with the nature of the considered service. They involve simple concepts such as service availability or transmission characteristics and subjective estimates.

The ITU defines the QoS as "the collective effect of service performance which determines the degree of satisfaction of a user of the service".

According to the RACE project QOSMIC **[8]**, the term Network Performance (NP), which represents the efficiency of the network in providing services to customers includes many factors influencing QoS.

The essential difference between QoS and NP is that QoS is user oriented, while NP is provider oriented. The terminal will also have a strong influence on the subjective perceived quality.

Factors affecting the QoS

The ETSI ETR003 reviews the factors that affect the QoS as perceived by the user. For a mobile service, the most significant are *call set up delay*, *probability of blocking* and the *effective bandwidth*. These factors can be both network and terminal dependent.

The **call set up delay** can be defined as the time interval from the instant the user initiates a connection request until the complete message indicating call disposition is received by the calling terminal.

The lack of network resources at the user plane as well as the control plane can cause unsuccessful call attempts. The probability of **end-to-end blocking** can occur at the radio link, at the interworking units between the mobile and the fixed networks or at the transit network.

The concept of **effective bandwidth** has been developed over recent years to provide a measure of resource usage **[4]**, which adequately represents the trade-off between sources of different types, taking account of their varying statistical characteristics and the QoS requirements.

Measurements of QoS parameters

QoS parameters are not always directly measurable.

The relative weight of their influence in the user's evaluation depends on the nature of the service. Preponderating criteria will obviously be different for users of interactive applications conversational, messaging or retrieval services and users of distribution services.

On the other hand, NP parameters are measurable. It is the network provider's most important task to "combine" the values of the parameters in such a way that the customers are satisfied with the resulting QoS obtained at an acceptable cost.

The relationship between QoS and NP is usually not a simple one and it is difficult to determine the range of each NP parameter producing any particular desired QoS level. Most of the time, several sets of NP parameters lead to an acceptable QoS. The one involving the lowest network costs is usually chosen.

QoS over mobile links

Existing multimedia transport systems are ineffective when operating in environment where widespread mobility and changing network characteristics are dominant.

These challenges, primarily due to large-scale mobility requirements, limited radio resources and fluctuating network conditions, fundamentally impact on our ability to deliver multimedia flows over mobile and, in general, QoS fluctuating networks. Therefore, one of the major challenges is the delivery of multimedia flows to mobile devices with QoS constraints.

The effect of the network on the transmission of traffic is dependent on a number of factors including the load on the network and its resulting effects on loss of packets and retransmissions **[3]**.

In a mobile and wireless network, the nature of the medium itself introduces problems owing to limited bandwidth and high bit error rate. The signal strength determines how well traffic is received; multipath effects have to be dealt with; mobility of the users requires procedures such as handovers to be executed, which has a direct impact on quality of service **[5]**.

Handover is a process whereby a connection may be lost temporarily, during which time traffic may get lost. The resulting delays and loss of traffic have a direct bearing on the deterioration of quality of service.

User perception

How a user characterises QoS is an important issue. On the one hand, there is a measure of quality associated with a particular instance and on the other hand, the assessment of the quality of a particular video clip depends on its actual length in time.

Assume that a clip is transmitted across a network, which is affected by various types of disturbance in terms of congestion. This is reflected in the number of corrupted packets and retransmissions required. The question of time therefore is important, especially that network effects can cause disruptions for different periods of time. The manner in which a user assesses a clip where there are frequent but short disruptions to the quality of the clip is different from the case where there is a longer disruption of quality but only once during the whole clip.

For example, if a particular error occurs in a number of successive frames, the user might observe this as being of lower quality than if the error occurred the same number of times but at different instances in time.

Methodology

Approach

This methodology, inspired from the ISO N0999 recommendation **[2]**, combines collections of objective and subjective data. These data are analysed in order to determine the critical thresholds, in which begins the difficulties of SL comprehension.

The **objective analysis** is intended to determine the relation between some video format parameters (frame rate and image quantification) and the required *effective bandwidth*. Furthermore, traffic profiles of video sequences containing sign language can be compared with idle sequences.

The **subjective analysis** is intended to determine the rate of comprehension of the signing within the same parameter range.

Tools

Five sign language video sequences of approximately twenty-five seconds have been used for both objective and subjective evaluations. The total duration of these video sequences is 160 seconds including 4-5 seconds idle time between each sequence. Series of four questions are associated with each sequence for the subjective evaluation.

These video sequences have been recorded on a PAL VHS format in optimum lighting and framing conditions. The actual experiment is carried out as shown in the figure 1.

The pre-recorded sequences are played back through a Unix station equipped with an S video interface. The VIC* video telephony tool is used to enable the communication between two Unix workstations. H.261 compression and CIF video format has been used throughout experiments.

The deaf subjects watched the video sequences at the other Unix workstation. Their understanding of the sequence was recorded in the form of answers to four questions for each sequence.

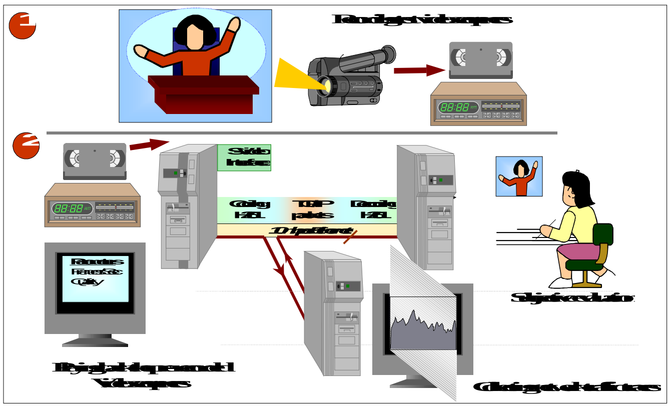

Figure 1: Synopsis describing the assessment of the sign language video sequences

Results

Results reported here are series of traffic measurements of H.261 sign language video telephony over IP (10 Mbps Ethernet) and the associated subjective quality assessment.

The two variable parameters are the frame rate (in the range 5 to 10 frames per second) and the image quantisation parameter (in the range 9 to 25 points), known as QUANT (macro bloc quantisation factor) in H.261 terminology.

Network traces collection

TCPDUMP tool was used to collect UDP packets, generated by the H.261 codec. This tool enables collection of data in different formats. The parameters of interest for our objective evaluation were the timestamps and the lengths of the packets. Given the timestamps and the size of the packets, we can determine the total number of bits of traffic transmitted over a specified time interval, giving an indication of the required instantaneous bandwidth.

Twenty-six traces have been collected for various values of the parameters of the video format. The chronograms (figure 2) represent the instantaneous bandwidth required over an integration period of one second. Each dot represents the cumulated amount of traffic in kilobits divided by the integration period.

The solid line corresponds to the envelope of this traffic when the factor QUANT is 19 whereas the dashed line shows the variation of traffic when QUANT becomes 9. For both cases, the measurement was done with six frames per second.

As can be seen, the traffic generated by the codec during the five signing periods is significantly higher than the one generated during the 5 seconds idle period (without signing) in between the sequences.

---

* VIC: Video Conferencing Tool, by Lawrence

Subjective quality assessments

Individual reactions from the subjects lead to scattered subjective quality assessment results. Twice as many subjects would significantly increase the confidence in the results.

Nevertheless even with 10 subjects, a correlation between the quality of the video and the degree of comprehension has been identified and classified (see table 1) in three fuzzy groups:

- G1: Most of the subjects answered correctly at least 75% of the questions with some of them reaching high scores. (lightest on the table)
- G2: The achievements were random, sometimes lower than 50% and no subject reached scores above 85%. Difficulties often arose from lip reading.
- G3: Both lips reading and signs were sometimes difficult to interpret. Scores ranged between 15% and 60%. (darkest on the table)

Quantiser value		Min. (25 points)	Average (19 points)	Max. (9 points)
Frames/second				
10	*Scores*	?	81%	81-100%
	Number of tests	0	1	3
8	*Scores*	40-68%	65%	75-86%
	Number of tests	2	1	2
6	*Scores*	25-60%	38-75%	56-85%
	Number of tests	3	4	4
5	*Scores*	15-20%	25-45%	40-75%
	Number of tests	2	2	3

Table 2: Classification of the degree of comprehension

Synthesis

The following two charts (figures 3 and 4) represent the average traffic versus the frame rate and the image quantification for signed and non-signed periods.

An equation is associated with each trend line to indicate the variation rate of the average bandwidth required in function of the frame rate. Here the variation rate range from 21.23 kb/s/frame (lowest value of QUANT) to 29.17 kb/s/frame (highest value of QUANT). R is the minus square factor indicating the deviation of the measurements from this trend line.

Table 2 summarises both objective and subjective results.

The values are the average effective bandwidth (in kbps) required for the transmission of sign language using H.261.

The cell shading depends on the subjective quality, according to the groups defined previously: G1=lightest, G3=darkest.

The ratio of effective bandwidth required between signed and non-signed sequences is almost constant in the range 4.9 and 5.0

Quantiser value		Min. (25 points)	Average (19 points)	Max. (9 points)
Frames/seconds				
10	*Signing*	198	242	339
	Not signing	40	42	72
8	*Signing*	157	180	280
	Not signing	35	37	59
6	*Signing*	118	135	218
	Not signing	30	33	47
5	*Signing*	96	110	185
	Not signing	28	30	38
	Average/frame, signing	**19.2 to 19.7**	**22.0 to 24.2**	**33.9 to 36.3**
	..., not signing	**4.0 to 5.6**	**4.2 to 6.0**	**7.2 to 7.8**

Table 2 : Average effective bandwidth required (in kb/s)

Discussion

Coding algorithms of both H.261 and H.263 standards combine inter-picture prediction and transform coding to remove temporal and spatial redundancy.

This study showed that, for a given quality, the relative bandwidth for the video sign language sequences compressed with H.261 is about **five times** higher than the sequences without sign language. This figure has to be verified for H.263 coding techniques.

According to our results, 242 kbps is the minimum effective bandwidth required for an acceptable videotelephony service for sign language. Below 135 kbps, there is no possibility to run such a service, even with low quality assumptions. In other words, this means that 4 ISDN or 25 GSM channels would be required to bear such a service, assuming an optimum channel quality and an insignificant overhead for the mobility management.

Thus, can "pay per bit" be envisaged for this particular service, as it remains to be the ATM approach. Will this approach be acceptable in any case for UMTS?

The GAM chain, in their guideline for residential videotelephony, recommends to introduce the concept of "service value" which is also a normal commercial practice. The problem for the future is to package services in an unstructured and connectionless environment.

There can only be a market for wide-spread broadband when the providers recognise their social duty to provide services and not quantities.

References

[1] Personnal Communication and Traffic engineering in ITU-T, Davide Grillo, Fondazione Ugo Bordoni, dec. 1996

[2] MPEG-4 Testing and Evaluation Procedures, ISO/IEC N0999, 1995

[3] Measured Performance of data transmission over GSM, Timo Alanko, University of Helsinki, Nov. 1994

[4] Stochastic Networks, Theory and applications, F.P Kelly, Oxford Science Publications, 1996, http://www.statslab.cam.ac.uk/~frank/eb.html

[5] L. LIU Performance method for mobility management in cellular networks. ICUPC 96.

[6] UMPTIDUMPTI D07: Mobile communication and Special Needs

[7] Generic Access for Multi Service Delivery to Residential Users (GAM chain from ACTS) - GAM-G3 Videotelephony for Residential Users.

[8] Project R1082 QOSMIC - QoS verification and tools for Integrated Broadband Communication: deliverable "QoS and Network Prformance".