

Vittorio Baroncini (Fondazione Ugo Bordonini),
 Antonino Barone, Angelo Ciavardini, Giancarlo Gaudino,
 Stefano Renzi, Biagio Tancredi (Istituto Superiore CTI)
 E-Mail: lab_QoS@istsupcti.it

**PROVE DI QUALITÀ PER SERVIZI MULTIMEDIALI SU RETE IP
 A LARGA BANDA
 (TEST OF QUALITY FOR MULTIMEDIA SERVICES OVER IP
 BROADBAND NETWORK)**

Sommario: la diffusione dei servizi multimediali su rete IP è spesso eseguita effettuando un compromesso fra banda disponibile e qualità percepita dall'utente finale, senza tener conto della natura dei contenuti (p.es. sport, news, film, video music ecc.). Questo lavoro si inserisce in un progetto che ha come scopo l'individuazione della migliore condizione di compromesso fra banda disponibile e qualità percepita dall'utente, anche in funzione del contenuto del servizio. La ricerca del compromesso ideale passa attraverso una complessa valutazione incrociata della qualità audio e video che comporta anche la valutazione separata dei due segnali al fine anche di determinare ogni possibile "effetto di trascinamento" reciproco. In questo articolo viene descritto il primo di una serie di esperimenti in cui è stato scelto di effettuare la ricerca del compromesso migliore variando la banda del segnale video mantenendo costante (ed elevata) la banda del segnale audio. Questo risultato è stato ottenuto effettuando una campagna di prove soggettive formali di tipo del tutto innovativo in cui audio e video vengono valutati in modo congiunto. I risultati di questa campagna di test hanno permesso di analizzare la risposta di un "pannello di utenti tipo", ovvero un insieme di valutatori opportunamente rappresentativi dell'utenza reale, selezionato grazie al contributo dell'Istituto Tecnico Professionale di Stato per i Servizi Turistici "Giulio Verne" di Roma, che ha consentito la partecipazione di un ampio campione di studenti dell'ultimo anno.

REALIZZAZIONE DI UN SERVIZIO DI IP-TV "SPERIMENTALE"

Un programma di IP-TV sperimentale può essere realizzato sia in tempo reale che in differita.

Al fine di minimizzare i costi, è stata realizzata

Abstract: the diffusion of multimedia services over IP network is usually done applying a compromise between the available bandwidth and the final user perceived quality, regardless the content of the service provided (e.g. sport, news, film etc.). In this regard an effort was done aiming at the definition of the best compromise between available bandwidth and final user perceived quality, taking into account the kind of provided services. The ideal compromise was identified by means of a complex evaluation of the two signals (audio and video) implying a separated evaluation of each signal and trying to find out any possible mutual "leading effect". In this paper a first experiment is described in which the best compromise is seek varying the video signal bandwidth and maintaining constant (and high) the bandwidth assigned to audio. This has been done by means of an innovative quality assessment method, representing one of the first cases of complete audio-visual assessment of multimedia services. The results of this test campaign allowed to analyse the reaction of a typical users' panel, i.e. a set of evaluators properly representing the real users, selected thanks to the kind permission of the "Istituto Tecnico Professionale di Stato per i Servizi Turistici Giulio Verne" of Rome, that gave its assistance in the presence in our laboratories of the students of several classes of the last year of the school courses.

la simulazione di un servizio in differita basato sulla generazione (e successiva messa in onda) di un set di programmi opportunamente preregistrati e codificati.

In questo caso è stato simulato un servizio IP-TV di tipo "consumer" a bassa qualità, ovvero un formato di immagine ridotto e l'utilizzo di canali di

trasmissione a banda sotto 1Mbps.

In ogni caso lo schema qui proposto per l'IP-TV di tipo "consumer", è in linea di massima valido anche per un servizio di tipo professionale; la differenza consiste solo nella capacità computazionale dei dispositivi utilizzati.

Nello schema di Figura 1 vengono evidenziate le fasi che hanno permesso di eseguire la preparazione del materiale video necessario alla realizzazione della simulazione del servizio.

E' stata dapprima realizzata una base dati di programmi "originali", ovvero per i quali non sono state eseguite diminuzioni di qualità del segnale dovute alla limitazione di banda del canale trasmissivo di cui si è voluto eseguire la simulazione. In questa base dati la qualità del segnale dipende dal tipo di codifica effettuato all'atto dell'acquisizione del segnale.

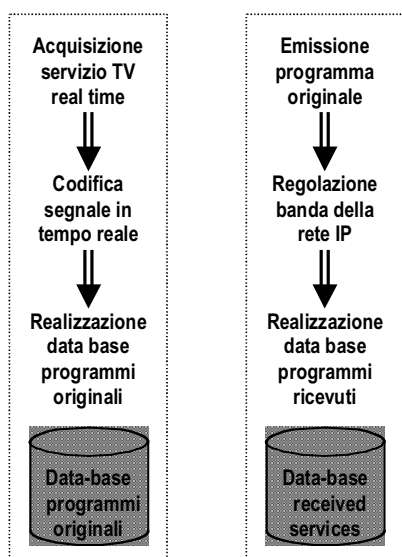


Fig. 1

Una volta eseguita la preparazione dei programmi "originali" si è passati alla simulazione della "messa in rete" dei medesimi, codificando i programmi contenuti nel data base a differenti bit rate, in funzione della casistica dei dispositivi di accesso alla rete IP più diffusi fra l'utenza media in Italia.

Questo processo ha comportato le fasi qui sotto elencate:

1. individuazione delle categorie di servizio maggiormente significative per un servizio di tipo "consumer" (Cartoon, Sport, Movie, Movie in Bianco/Nero),
2. acquisizione di diversi minuti di programmi per ognuno delle categorie,
3. selezione degli "spot" di maggiore significati-

vità per ogni categoria,

4. individuazione delle condizioni di erogazione di servizi ad un utente consumer (ovvero bit rate e formato dell'immagine),

5. codifica di ciascuno "spot" selezionato al punto 3 secondo i bit rate individuati al punto 4.

In Appendice II vengono elencate tutte le categorie individuate.

In Appendici III vengono elencati tutti gli "spot" selezionati; ad ognuno di essi è associata un fotogramma ed una sommaria descrizione del contenuto

In Appendice IV vengono elencate le condizioni di codifica applicate a gli "spot" selezionati, e che hanno permesso di effettuare una simulazione delle categorie individuate per un servizio di tipo "Consumer".

VALUTAZIONE DELLA SIMULAZIONE DEL SERVIZIO IP-TV

La valutazione della simulazione del servizio IP-TV è stata eseguita limitando la complessità del processo alla sola valutazione di singoli spezzoni di programmi (15-20 secondi), e non a programmi completi, così come intesi nella accezione comune del termine (p.es. un'ora).

Questa scelta è stata dettata sia dalla necessità di rendere gestibile in tempi ragionevoli l'esperimento, che dalla esigenza di condurre i test con le apparecchiature disponibili in laboratorio.

In dettaglio, è possibile con le tecniche più avanzate eseguire la valutazione di programmi reali, ma lo sperimentatore si deve confrontare con una serie di problemi di non poco conto; fra primi il livello di attenzione dei soggetti umani reclutati per eseguire la valutazione; uno sforzo di attenzione prolungato (oltre la mezz'ora) provoca stanchezza, abbattimenti (anche drastici) del livello di attenzione, rifiuto a partecipare ad altri esperimenti.

Inoltre si è considerato che il rapporto costo beneficio della valutazione di uno spezzone di 15 secondi, confrontato con la valutazione di sequenze di lunga durata, non comportava vantaggi di tipo risolutivo nella qualità dei risultati; in altri termini, essendo lo scopo dell'esperimento quello di ottenere un'indicazione del bit rate limite sotto il quale un servizio non è più accettabile per un utente medio, si è ritenuto che dopo 15 secondi un qual-

siasi valutatore, opportunamente addestrato, fosse in grado di fornire le stesse indicazioni ottenibili dopo 30 120 180 secondi, o anche dopo tempi più lunghi paragonabili a quelli di un reale servizio (p.es. 15 minuti - mezz'ora).

Al contrario il livello di innovazione nella valutazione è stato introdotto eseguendo una valutazione congiunta e contemporanea delle qualità video ed audio. È importante rilevare come, al fine di evitare confusione ma soprattutto di ottenere

una "stratificazione" degli stimoli inferti ai valutatori, si è inteso procedere per gradi; ovvero l'esperimento è stato presentato ai valutatori come un contesto in cui la qualità variava sia per il video che per l'audio, mentre in realtà solo il video ha subito variazioni di qualità, mentre l'audio è stato mantenuto a livelli di qualità massima (per lo meno la massima consentita per servizi di tipo "consumer").

CREAZIONE DEL DATABASE DI SEQUENZE DA VALUTARE

Categoria di servizi	Bit rate	Bit rate	Bit rate	Bit rate	Bit rate originale
A = Cartoon	64 kbit/s	128 kbit/s	256 kbit/s	512 kbit/s	1500 kbit/s
B = Sport	64 kbit/s	128 kbit/s	256 kbit/s	512 kbit/s	1500 kbit/s
C = movie	64 kbit/s	128 kbit/s	256 kbit/s	512 kbit/s	1500 kbit/s
D = bianco/nero	64 kbit/s	128 kbit/s	256 kbit/s	512 kbit/s	1500 kbit/s
E = video music	64 kbit/s	128 kbit/s	256 kbit/s	512 kbit/s	1500 kbit/s

Il software utilizzato per la editing video è stato Ulead Media Studio Professional 7.0 che ha permesso di codificare un filmato sorgente in Mpeg2 in Mpeg4 in base ai seguenti template:

File
File name: G:\test_scuola\B\B64.avi
File format: Microsoft AVI files
File size: 2.607 KBytes
Frame rate: 25.000 Frames/sec
Duration: 12.000 Seconds
Data rate: 216.30 kbps

Video
Compression: Microsoft MPEG-4 Video Codec V2
Attributes: 24 Bits, 352 x 288
Total frames: 300 Frame(s)

Audio
Compression: PCM
Attributes: 48,000 kHz; 16 bit; Stereo
Total samples: 576.000 Samples

File
File name: G:\test_scuola\B\B128.avi
File format: Microsoft AVI files
File size: 2.903 KBytes
Frame rate: 25.000 Frames/sec
Duration: 12.000 Seconds
Data rate: 241.01 kbps

Video
Compression: Microsoft MPEG-4 Video Codec V2
Attributes: 24 Bits, 352 x 288
Total frames: 300 Frame(s)

Audio
Compression: PCM
Attributes: 48,000 kHz; 16 bit; Stereo
Total samples: 576.000 Samples

File
File name: G:\test_scuola\B\B256.avi
File format: Microsoft AVI files
File size: 3.593 KBytes
Frame rate: 25.000 Frames/sec
Duration: 12.000 Seconds
Data rate: 298.50 kbps

Video
Compression: Microsoft MPEG-4 Video Codec V2
Attributes: 24 Bits, 352 x 288
Total frames: 300 Frame(s)

Audio
Compression: PCM
Attributes: 48,000 kHz; 16 bit; Stereo
Total samples: 576.000 Samples

File
File name: G:\test_scuola\B\B512.avi
File format: Microsoft AVI files
File size: 4.230 KBytes
Frame rate: 25.000 Frames/sec
Duration: 12.000 Seconds
Data rate: 351.59 kbps

Video
Compression: Microsoft MPEG-4 Video Codec V2
Attributes: 24 Bits, 352 x 288
Total frames: 300 Frame(s)

Audio
Compression: PCM
Attributes: 48,000 kHz; 16 bit; Stereo
Total samples: 576.000 Samples

File
File name: G:\test_scuola\B\OrigB.avi
File format: Microsoft AVI files
File size: 4.843 KBytes
Frame rate: 25.000 Frames/sec
Duration: 12.000 Seconds
Data rate: 402.70 kbps

Video
Compression: Microsoft MPEG-4 Video Codec V2
Attributes: 24 Bits, 352 x 288
Total frames: 300 Frame(s)

Audio
Compression: PCM
Attributes: 48,000 kHz; 16 bit; Stereo
Total samples: 576.000 Samples

SERVIZIO IP-TV A LIVELLO "CONSUMER"

Il livello di qualità "consumer" per l'IP-TV è quello fruibile (con l'attuale tecnologia) da un utente collegato in rete mediante una connessione telefonica commutata su rame (doppino telefonico).

Rientrano in questa fascia (e quindi nella valutazione) le velocità di connessione comprese fra 24Kbps (modem V90 su linea di media qualità) e 1,6 Mbps (modem XDSL su linea di elevata qualità).

Non rientra in questa fascia la fruizione del servizio IP-TV mediante collegamento su fibra ottica, non perché esso non abbia come "target" l'utente finale, ma in quanto i livelli di qualità raggiungibili, con questo tipo di collegamento, richiedono l'impiego di procedure e di terminali "professionali", del tipo utilizzato per la valutazione della fornitura di un servizio TV tradizionale (broadcast on air); in altri termini la valutazione di un servizio IP-TV fornito su fibra richiede l'impiego di terminali e server video a livello "broadcast professionale"; al contrario la valutazione del servizio IP-TV fornito a livelli di qualità consumer può essere eseguita utilizzando terminali e server di tipo "consumer" (PC con schede video evolute, monitor per PC a tubo o LCD)

VALUTAZIONE FORMALE DELLA QUALITÀ

Al fine di eseguire una valutazione formale della qualità ottenuta in differenti condizioni del canale, è necessario, utilizzare uno dei programmi registrati nel data base di programmi originali e gli elementi del data base di programmi ricevuti ricavati ricevendo quel determinato programma.

I segnali relativi a varie condizioni di rete sono stati presentati (insieme a quello originale) in ordine casuale; i valutatori, seduti di fronte a monitor situati in una cabina silente, hanno espresso il loro giudizio di qualità, apponendo un segno su un questionario opportunamente progettato.

Il risultato ha portato alla generazione di grafici di qualità, in cui viene visualizzata la variazione di qualità in funzione della larghezza di banda del canale trasmissivo.

PROTOCOLLO DI VALUTAZIONE FORMALE DELLA QUALITÀ - MMDSIS

LE PREMESSE

La valutazione formale della qualità ottenuta in differenti condizioni di fruizione di un programma deve rispondere a precisi protocolli codificati da tempo dalla Commissione 6 del Radio Sector dell'International Telecommunication Sector (ITU-R).

In particolare la base operativa nella definizione del protocollo di valutazione è stata ricavata in funzione dei dettami contenuti nella Raccomandazione 500-11.

Su questa base sono stati apportati gli adattamenti ritenuti più idonei alla sperimentazione in corso.

In particolare, trattandosi di valutare un servizio (nell'accezione globale), si è cercato di associare alla valutazione video la valutazione audio.

Il protocollo di valutazione risultante da questa associazione, è del tutto simile a quello riconosciuto valido nella raccomandazione 500-11 per il solo video, con l'aggiunta di una ulteriore richiesta di valutazione (per l'audio) alla fine della presentazione di ogni singolo spezzone di programma multimediale (da qui in poi denominato "Spot").

Questo protocollo, denominato Double Stimulus Impairment Scale (DSIS) prevede la presentazione di un clip video non codificato a cui fa seguito la presentazione dello stesso clip video codificato (con intensità di elaborazione più o meno profonda). Nel nostro esperimento il protocollo in questione è stato modificato al fine di eseguire la valutazione di clip video completi del segnale audio associato, in modo tale da riprodurre (seppure per un tempo limitato) un reale servizio.

Tale modifica si è riflessa anche nel tipo di questionario utilizzato per la raccolta dei giudizi, che, a differenza di quanto canonizzato nella raccomandazione 500-11, presenta non più una ma due scale di valutazione della qualità: una per il video ed una per l'audio.

PERCHÉ UN NUOVO PROTOCOLLO

Ma vi è da mettere in evidenza anche il fatto, di non poca rilevanza, che la Raccomandazione 500-11 contiene metodi di valutazione soggettiva studiati e definiti esclusivamente per la valutazione di materiale video di tipo televisivo, e quindi generato e rappresentato (secondo la Raccomandazione

ITU-R 601) in formato a scansione di schermo interallacciata.

Quindi, nel caso di servizi Multimediali era necessario eseguire degli ulteriori adattamenti, legati al tipo di rappresentazione del segnale video sul display.

Questo fattore, se a prima vista risulta essere non importante dal punto di vista del protocollo di valutazione, ha un impatto rilevante dal punto della logistica di realizzazione del test, ovvero sulla scelta della apparecchiature e di programmi software da utilizzare per rappresentare ai soggetti i clip multimediali.

Infatti la raccomandazione 500-11 descrive come deve essere preparato il laboratorio, la distanza dallo schermo di visione, l'illuminazione dell'ambiente ed le caratteristiche dei display da utilizzare, che, come detto sopra, sono (anzi "devono" essere) di tipo televisivo. Inoltre per permettere una fedele e corretta visione delle immagini di test sui display è necessario utilizzare un video registratore a nastro di tipo "D1" (che registra e riproduce immagini televisive senza introdurre distorsioni) o un server video di tipo televisivo.

Ma essendo in presenza di un protocollo nuovo per consentire, per la prima volta, di valutare congiuntamente audio e video in ambiente multimediale, detto protocollo è stato ri-battezzato MM-DSIS, ovvero protocollo a doppio stimolo per materiale Multimediale.

IL PROTOCOLLO MM-DSIS

La struttura del protocollo di valutazione MM-DSIS prevede l'osservanza delle seguenti regole:

- ogni singolo caso da valutare viene compiutamente rappresentato mediante la presentazione di un originale e di un elaborato realizzati utilizzando lo stesso clip multimediale,

- il singolo caso da valutare è inserito in un sotto-protocollo (definito protocollo della cella elementare di test - Basic Test Cell - BTC) strutturato come indicato in fig.2 .

- alla fine di ogni BTC il valutatore (vedi Figura 2) riceve un messaggio che lo invita a votare indicando la qualità video; dopo 5 secondi riceve un secondo messaggio che lo invita a votare indicando la qualità audio; i due messaggi che invitano a votare sono

- o "Vota Video N" (per il video)

- o "Vota Audio N" (per l'audio)

dove "N" indica il numero della BTC in fase di valutazione.

- Il voto associato al video viene emesso dal valutatore apponendo un segno su una di cinque caselle poste una sopra l'altra in verticale, a cui vengono associati degli indicatori semantici che aiutano il valutatore a comprendere il significato del voto espresso (vedi "Appendice I - Foglio di valutazione").

- Terminata la BTC "N" al valutatore viene presentata la BTC "N+1", fino al completamento del test.

- Ogni test non può impegnare i valutatori per più di 30 minuti; nel caso in cui il numero di casi da valutare comportasse un numero di BTC la cui lunghezza totale supera i 30 minuti, si riorganizzano le BTC in sessioni separate, la cui lunghezza non deve comunque superare i 30 minuti.

- Nel caso di utilizzo di più sessioni, un valutatore deve osservare un periodo di riposo pari almeno alla durata della sessione appena svolta.

- Al fine di rendere più agevole a ciascun soggetto la creazione di una corretta scala di valutazione, ogni sessione di test inizia con una fase detta di "stabilizzazione"; durante questa fase (composta da un numero di BTC comprese fra 3 e 5) vengono presentati ai soggetti degli spot caratterizzati da un livello di qualità ottimo, medio ed estremamente basso; in questo modo ogni soggetto si abitua subito al range di qualità che dovrà votare durante la sessione, creandosi il più presto possibile una scala di valori corretta. Per una buona riuscita della "stabilizzazione" dei soggetti, è necessario che gli BTC contengano materiale che fa parte della sessione di test in fase di svolgimento.

- La fase di test effettiva deve essere preceduta da due importanti fasi che vengono eseguite una volta sola all'inizio di ogni esperimento:

- o Lo "screening" del sistema visivo dei valutatori,

- o L'addestramento dei valutatori.

- I dati raccolti (eseguendo una o più sessioni fino al completamento del test) vengono sottoposti ad analisi statistica, tale da ottenere, per ogni caso sottoposto a valutazione) i seguenti dati:

- o Valore medio delle opinioni dei valutatori (Mean Opinion Score - MOS),

- o Varianza associata ad ogni MOS,

- o Indice di confidenza (Confidence Interval - CI) del campione utilizzato per calcolare ogni valo-

re di MOS, calcolato al 95% della distribuzione.

Le suddette regola sono state osservate nella effettuazione del test di servizi "Consumer" di tipo IP, di cui questo documento fornisce i risultati.



Figura 2 - Esempio di BTC, utilizzato nel presente esperimento

IL SOFTWARE QUALIVIEW, PER LA GESTIONE DEL PROTOCOLLO MM-DSIS

A questo punto si poneva il problema di reperire un nuovo "tool" (hardware o software) in grado di rimpiazzare le suddette apparecchiature televisive, e al tempo stesso in grado di eseguire la presentazione delle immagini secondo l'ordine imposto dal protocollo di valutazione soggettiva formale.

Dopo alcune ricerche, è stato deciso di acquisire, da una ditta specializzata in prodotti per la valutazione della qualità (la Genimedia) un software che permettesse di rappresentare degli spot multimediali (audio + video) secondo il protocollo di valutazione soggettiva formale scelto, ovvero l'MM-DSIS. Inoltre detto software era in grado di eseguire questo compito utilizzando un PC di buon livello a cui sono stati collegati dei display LCD.

Il software Qualiview è estremamente pratico e dispone di un approccio professionale. Infatti, nella preparazione delle sessioni di test, non impone un'interfaccia di lavoro a finestre (che va bene per lavori a bassa complessità) ma permette di sviluppare delle sessioni di test molto complesse ed articolate attraverso la semplice compilazione di file di testo.

In Figura 3 viene fornito un esempio di un file di gestione di una sessione di test necessario per permettere al programma Qualiview di eseguire una corretta presentazione del materiale da valutare.

All'avvio del programma (vedi) è sufficiente richiamare il file di testo opportunamente compilato, per lanciare la sessione di prova.

```
%  
% Sample DSIS script  
%  
% Initialise session  
%  
[Start Setup]  
    [Background]    100 100 100 % Background colour  
    [Display]       360 288     % Set video frame size  
[End Setup]  
%  
% Execute test  
%  
[Start Cell]  
    [Annotation]    'Test starting in 3 seconds' 1000  
    [Annotation]    'Test starting in 2 seconds' 1000  
    [Annotation]    'Test starting in 1 seconds' 1000  
[End Cell]  
  
[Start Cell]                                % Initiate new cell  
    [Annotation]    ''      500           % Show message for 500ms  
    [Annotation]    'A'    2000          % Show blank screen for 2s  
    [Video]         'c:\M&C_or.avi'      % Play reference video  
    [Annotation]    ''      500           % Show blank screen for 500ms  
    [Annotation]    'B'    2000          % Show message for 2s  
    [Video]         'c:\M&C_64K.avi'     % Play processed video  
    [Annotation]    ''      500           % Show message for 500ms  
    [Annotation]    'A*'  2000          % Show blank screen for 2s  
    [Video]         'c:\M&C_or.avi'      % Play reference video  
    [Annotation]    ''      500           % Show blank screen for 500ms  
    [Annotation]    'B*'  2000          % Show message for 2s  
    [Video]         'c:\M&C_64K.avi'     % Play processed video  
    [DSIS]          % shows the voting interface  
[End Cell]                                % Terminate cell  
  
[Start Cell]  
    [Goodbye]      'END'  
[End Cell]  
%  
% End of script  
%
```

Figura 3 - Esempio di file di gestione di una sessione di test

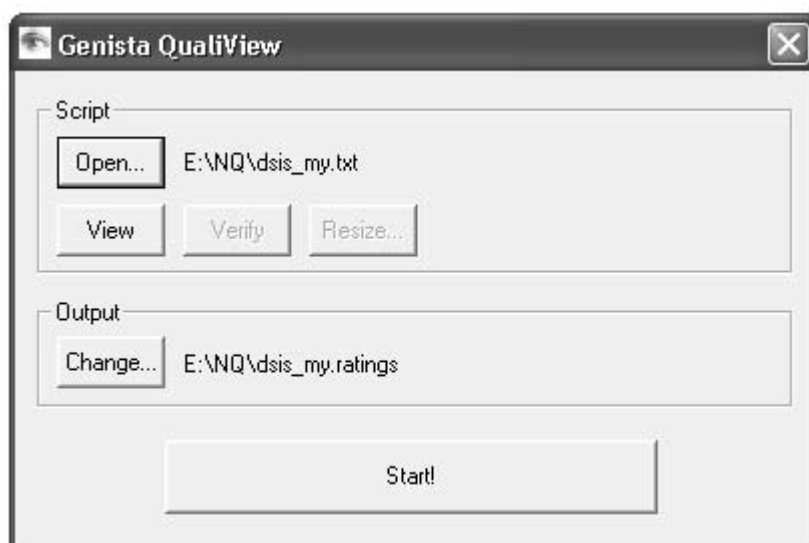


Figura 4 - Programma Qualiview; finestra di selezione del file di gestione e di avvio del test

SVOLGIMENTO DEI TEST

I test di valutazione di servizi "Consumer" si sono svolti presso il laboratorio di valutazione soggettiva dell'Ufficio IV dell'ISCTI.

L'ambiente selezionato è stato la camera silente ivi presente.

Screening del sistema visivo dei soggetti

I soggetti utilizzati facevano parte di una classe dell'ultimo anno di un istituto tecnico; la distribuzione fra i sessi era omogenea (vicina al 50%) e tutti i soggetti hanno superato lo "screening" di valutazione del sistema visivo.

Detto "screening" è consistito in due fasi:

- verifica dell'acutanza visiva, mediante tavola di Snellen (vedi Figura 5),

- verifica della corretta visione dei colori, mediante test di Ishihara. (vedi Figura 6).

Qui di seguito vengono riportati un esempio di tavola di Snellen ed due delle immagini utilizzate per il test di Ishihara.

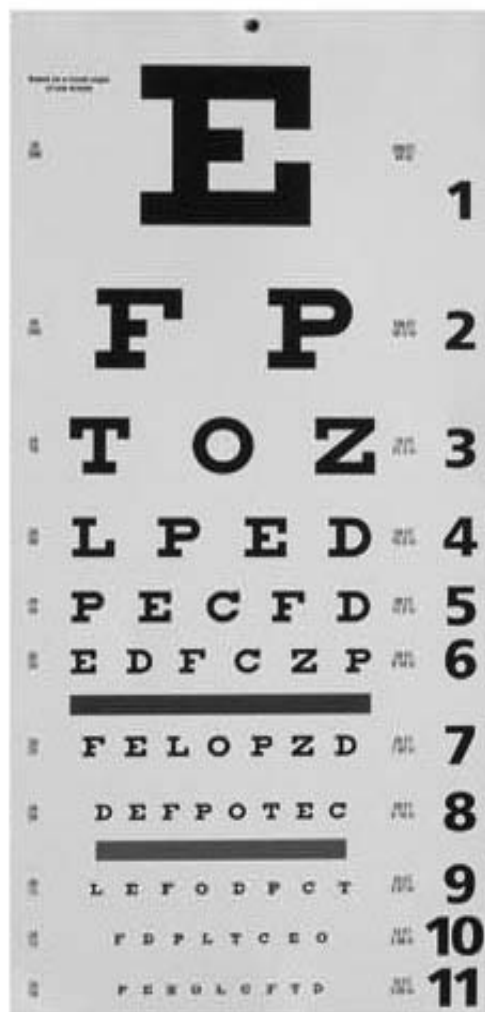


Figura 5 - Esempio di tavola di Snellen

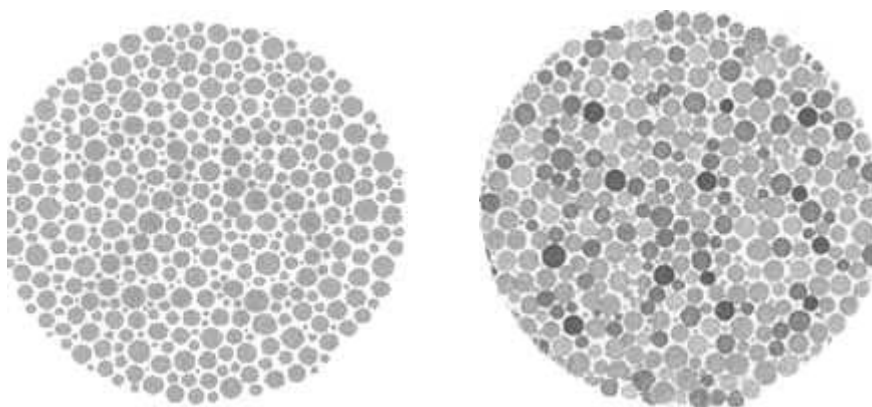


Figura 6 - Esempio di tavole di Ishihara

Addestramento dei soggetti alla valutazione

Terminata la fase di "screening" visivo si è proceduto alla fase di addestramento; durante questa fase ogni soggetto ha eseguito un minitest (comprendente 5 BTC) durante il quale ha familiarizzato con il foglio di valutazione e la tempistica del test. La fase di addestramento è stata preceduta da una spiegazione dei compiti da svolgere, mentre durante l'addestramento gli sperimentatori erano posizionati all'interno dell'area di test e pronti a fornire eventuali consigli suggerimenti ai soggetti impegnati nell'addestramento. Concluso l'addestramento, i soggetti sono stati invitati a porre

domande tese a chiarire ogni eventuale dubbio o quesito, tale da renderli confidenti in una corretta esecuzione del compito che li attendeva durante il test reale.

È bene tenere in conto l'estrema importanza della fase di addestramento dei soggetti.

Infatti lo scopo di questa fase preliminare ha, oltre allo scopo di rendere agevole il compito di persone che per la prima volta si accostano ad esperimenti di questo genere, anche l'obiettivo di ottenere dei risultati tali da potere eseguire una corretta ed efficace discriminazione (dal punto di vista dell'analisi dei dati statistici) dei casi in esame.

Come detto altrove, la base dell'analisi dei risultati di un test di tipo soggettivo è di tipo statistico e si basa sull'analisi della varianza.

Quindi per rendere questo tipo di analisi un mezzo efficace di discriminazione dei casi sottoposti a test (ovvero per potere dire con certezza statistica che una condizione è migliore o peggiore di un'altra) è necessario rendere il più possibile omogeneo il comportamento dei soggetti che partecipano all'esperimento.

Un tale risultato si ottiene in due modi:

- selezionando i soggetti per cate-

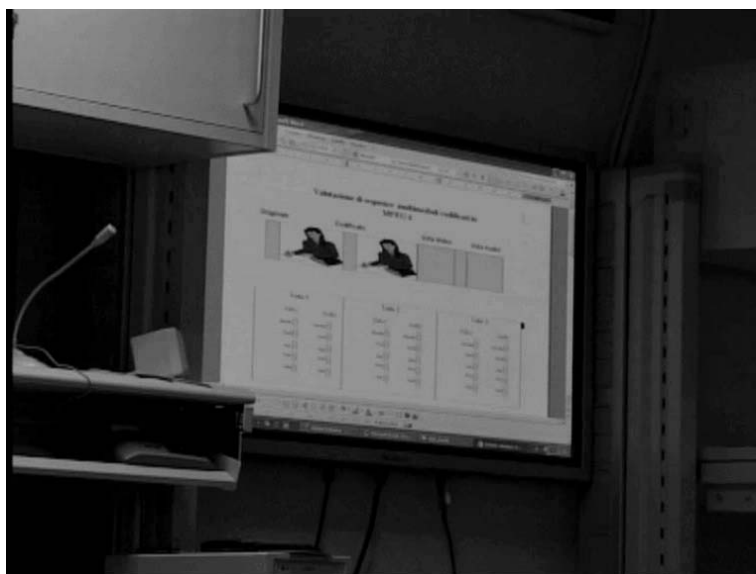


Figura 7 - Una momento della fase di addestramento dei soggetti

gorie sociali omogenee (studenti, operai, casalinghe ecc.),

- addestrando i soggetti su come reagire agli stimoli a cui sono sottoposti.

La prima condizione prescinde dalla fase di addestramento, ma è stata tenuta in conto in quanto i soggetti che hanno partecipato all'esperimento erano tutti studenti dell'ultimo anno delle scuole medie superiori.

La seconda condizione si ottiene spiegando in modo accurato e preciso il significato semantico del processo di espressione di un voto di qualità.

Nel nostro caso, ogni soggetto aveva a disposizione una scala a cinque livelli, ad ognuno dei quali era associato un significato semantico di qualità. Questa associazione, se può facilitare l'omogeneizzazione dei comportamenti, non basta a garantire risultati ottimali, in quanto esiste sempre un diverso grado di interpretazione degli aggettivi associati ai livelli di qualità da parte di ogni soggetto, interpretazione legata ovviamente al significato che ognuno di noi attribuisce ad un aggettivo.

Per rendere maggiormente omogenea l'interpretazione della scala utilizzata, si è proceduto a spiegare cosa lo sperimentatore intendesse associare ad ogni aggettivo di qualità associato ad un livello di qualità.

Quindi è stato suggerito di apporre il voto nella casella identificata dall'aggettivo "Eccellente" solo nel caso in cui il soggetto non rilevasse nessuna differenza fra l'originale e l'elaborato; altrettanto è stato fatto per l'aggettivo "Buono" da usarsi quando le differenze erano visibili ma solo prestando una certa attenzione; è stato quindi suggerito di scegliere l'aggettivo "Accettabile" quando le differenze erano visibili anche senza che fosse necessario prestare una particolare attenzione; l'aggettivo "Mediocre" è stato suggerito per i casi in cui le differenze con l'originale fossero evidenti; infine l'aggettivo "Cattivo" è stato suggerito in caso di forte

degrado dell'immagine elaborata.

In questo modo sono stati ottenuti dei risultati che, a seguito dell'elaborazione statistica, hanno mostrato una forte omogeneità fra i soggetti (bassa varianza) e quindi una elevata capacità di discriminazione fra i vari casi sottoposti a confronto.

STRUTTURA E CONDUZIONE DEL TEST

Il test si è articolato in due sessioni distinte composte ciascuna da un totale di 26 BTC.

Ogni BTC era composta da:

- 2 secondo di grigio (preparazione allo stimolo),
- 15 secondi di clip multimediale non compressa ("Reference"),
- 2 secondo di grigio (preparazione allo stimolo),
- 15 secondi di clip multimediale compressa ("Codified"),
- 6 secondi di messaggio "Vota Video N"
- 2 secondo di grigio (preparazione al voto successivo),
- 6 secondi di messaggio "Vota Audio N"

Pertanto la durata totale di un BTC era pari a 48 secondi, che ha reso ciascuna sessione pari ad una lunghezza di 20 minuti circa.

È stata la prima volta che sono stati utilizzati valutatori non universitari, ma delle quinte classi dell'Istituto Professionale di Stato per i servizi turistici "Giulio Verne" di Roma.

I soggetti hanno eseguito i test senza mostrare difficoltà o segni di stanchezza, fornendo risultati la cui distribuzione è risultata di tipo gaussiana e pertanto sottoponibile ad analisi della varianza.

I test sono stati eseguiti in due sessioni in base al seguente schema, dove le lettere rappresentano il tipo di filmato (categoria di servizi) e i numeri la velocità in bit rate:

Sessione 1	d64	e512	c128	a64	e128	c128	d256	d64	c256	b128	e512	a256	b64
------------	-----	------	------	-----	------	------	------	-----	------	------	------	------	-----

Sessione 2	c64	b512	a128	b256	e64	c512	d128	d512	c64	b512	e256	a512	a128
------------	-----	------	------	------	-----	------	------	------	-----	------	------	------	------

AMBIENTE DI TEST E POSTAZIONI DI LAVORO

Le foto qui sotto riportate illustrano l'ambiente in cui si è svolto il test (le foto si riferiscono al laboratorio come era prima dell'allestimento per il test in questione).

L'ambiente in questione è rappresentato dalla "camera silente" realizzata presso il laboratorio dell'Ufficio IV, caratterizzata dalle seguenti specifiche costruttive:

- dimensioni: lunghezza 7,00mt; larghezza 4,20 mt; altezza 2,40 mt.
- abbattimento sonoro 40dB a partire da 500Hz.
- tempo di riverberazione compreso tra 0,3 s. e 0,4 s.

L'illuminazione è ottenuta mediante coppie di lampade fluorescenti sistemate sul pavimento e sul soffitto ed orientate in modo da evitare proiezione diretta di luce sulla postazione di lavoro.

La luce emessa dalle lampade è regolabile attraverso regolatori di tipo elettronico dotati di memoria dell'impostazione.

Le lampade fluorescenti sono di tipo "daylight" (a largo spettro di emissione) e pilotate con trasformatori ad alta velocità di commutazione (circa 100 KHz); questa soluzione permette di evitare qualsiasi interferenza con le frequenze di scansione di quadro dei display impiegati per i test (tipicamente variabili fra i 50Hz ed i 100 Hz).

Il livello di emissione luminosità delle coppie di lampade viene regolato in funzione del tipo di display impiegato. La regola seguita per la regolazione



Figura 7 - Camera silente del laboratorio QoS dell'Ufficio IV dell'ISCTI

Come si vede dalle figure ogni postazione è realizzata separando un unico ambiente mediante tendine scorrevoli che permettono ai soggetti di operare senza distrazioni esterne.

Ogni postazione è realizzata di fronte ad uno sfondo realizzato in pannello di colore grigio medio e in grado di eliminare alcun tipo di riflessione diretta di luce.

della luce è quella di fare sì che lo sfondo emetta un flusso luminoso di pari intensità di quello emesso dal display utilizzato quando esso viene pilotato con un'immagine grigia uniforme pari alla metà del livello di bianco (grigio medio).

All'interno della sala di valutazione i soggetti sono stati disposti in quattro postazioni di lavoro (vedi) ognuna delle quali era equipaggiata con una



Figura 8 - Alcuni momenti della valutazione

scrivania ed una sedia. Su ogni scrivania è stato sistemato un display LCD da 17" dello stesso modello e marca di fabbricazione.

La distanza del display dal bordo della scrivania era uguale per ogni postazione e pari a 40cm. In questo modo è stato riprodotto un ambiente tipico di fruizione di un servizio multimediale. Ogni soggetto aveva di fronte a se un foglio di valutazione, che è stato sistemato nella posizione di cui solitamente è posta la tastiera di un computer.

Le foto qui sopra riportate illustrano alcuni momenti dello svolgimento dei test.

RISULTATI

I risultati ottenuti mediante i test soggettivi formali, eseguiti presso il Laboratorio Valutazione QoS Multimediali Ufficio IV dell'ISCTI, sono illustrati nei grafici riportati in questo capitolo.

Ogni grafico mostra come, per ogni condizione di accesso "Consumer" (ovvero bit rate di accesso alla rete IP), il livello di qualità ottenuto dipenda dal contenuto delle immagini, ovvero dal tipo di categoria di servizio (ovvero Sport, Commedia, News, Film in bianco e nero, ecc.).

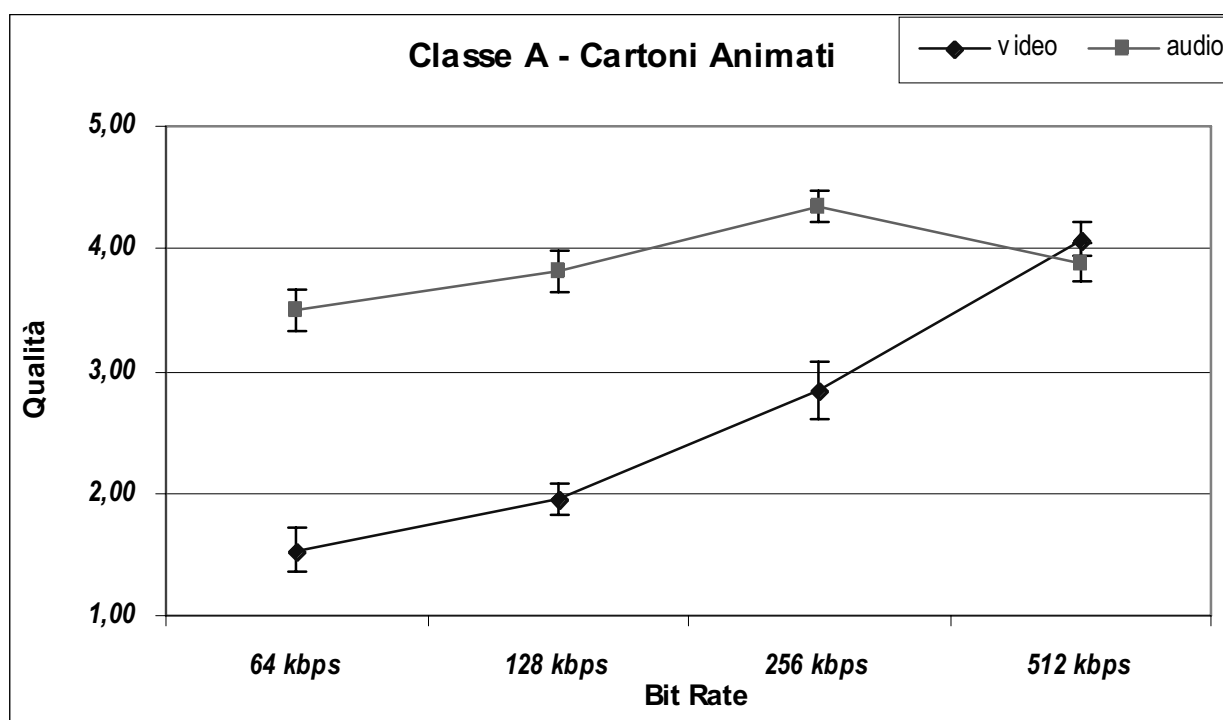


Figura 9 - Valutazione della qualità Video ed Audio per contenuto "Cartoni Animati"

Inoltre osservando i grafici è possibile avere un'indicazione del tipo di degrado che ogni categoria di servizio (in ragione del contenuto semantico del segnale) subisce in relazione alla diminuzione di banda con cui un potenziale utente ha accesso alla rete IP.

E' interessante notare che in tutti i clip sottoposti a test la qualità audio è stata mantenuta costante.

Le oscillazioni dei valori assegnati alla qualità audio indicano la naturale predisposizione dei soggetti ad esprimere un giudizio di qualità anche in

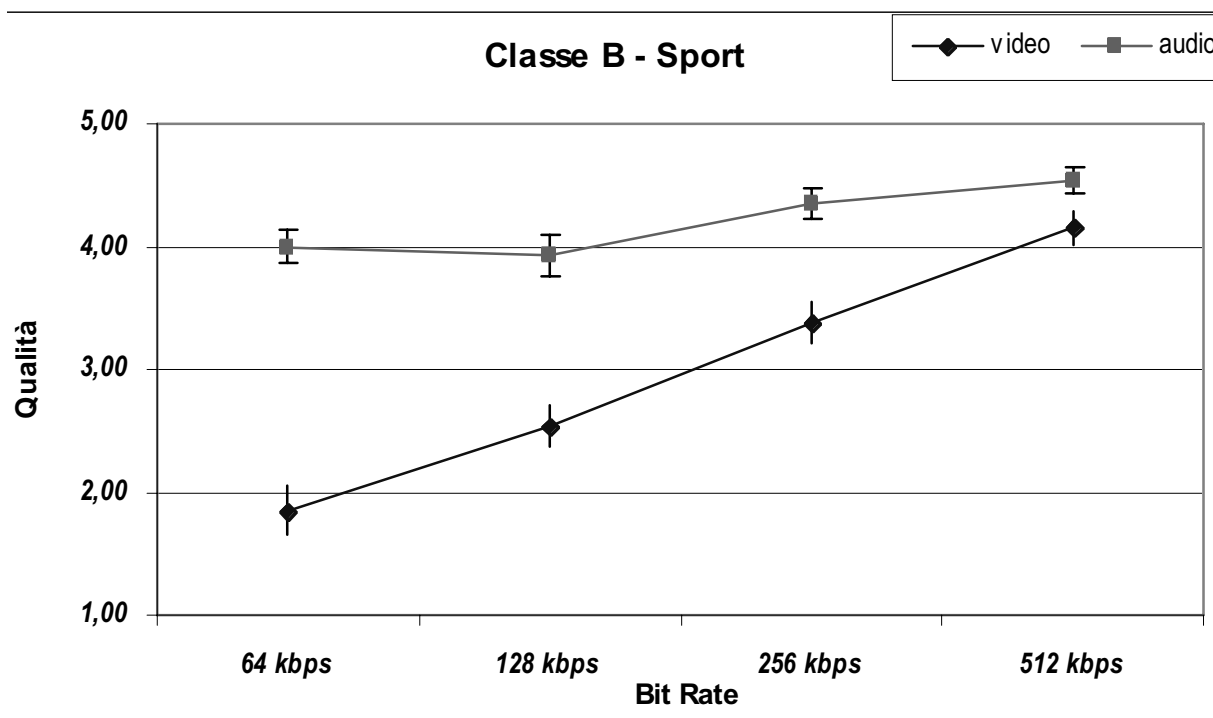


Figura 10 - Valutazione della qualità Video ed Audio per contenuto "Sport"

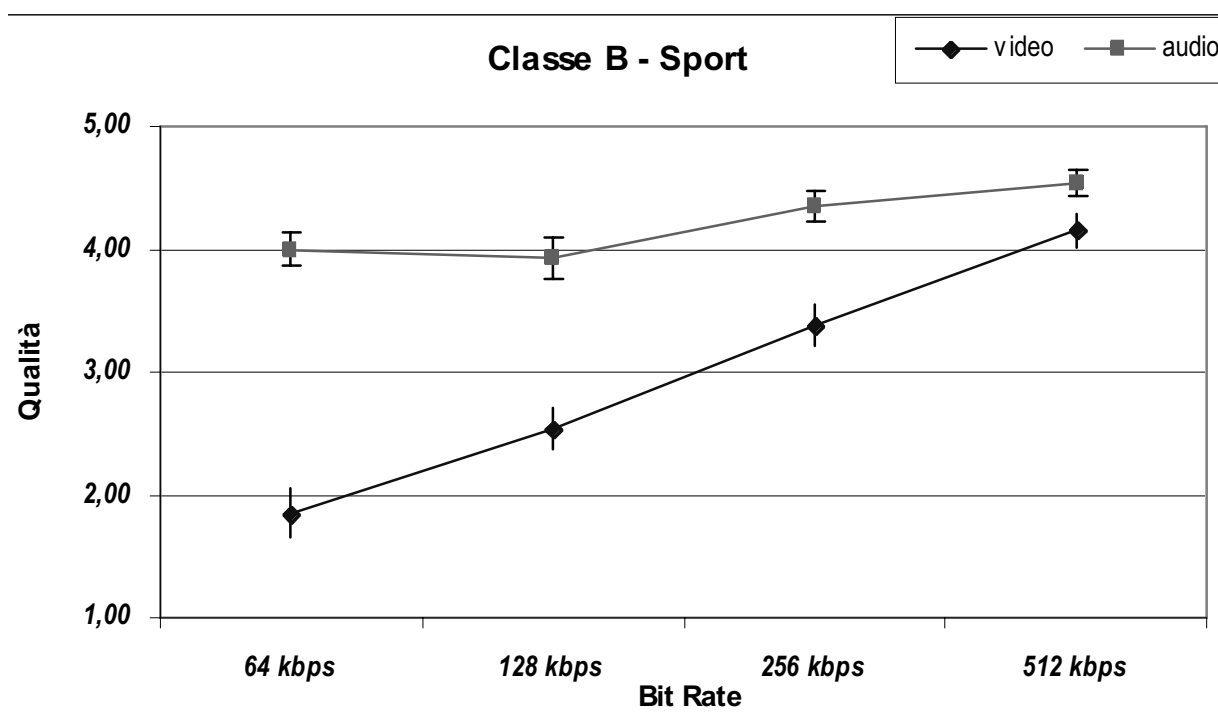


Figura 11 - Valutazione della qualità Video ed Audio per contenuto "Movie"

assenza di degrado.

In effetti questo conferma quanto atteso; ovvero se si vuole avere un attendibile giudizio del segnale audio, è necessario eseguire un "addestramento" più accurato dei soggetti per quel tipo di valutazione.

In ogni caso, il giudizio espresso per l'audio è

rientrato nei normali valori di oscillazione previsti per valutatori non particolarmente esperti e non vincolati da particolari "commitment" (ovvero non pagati per eseguire il test).

Esperienze con valutatori retribuiti hanno dato in passato una maggiore stabilità di risultati.

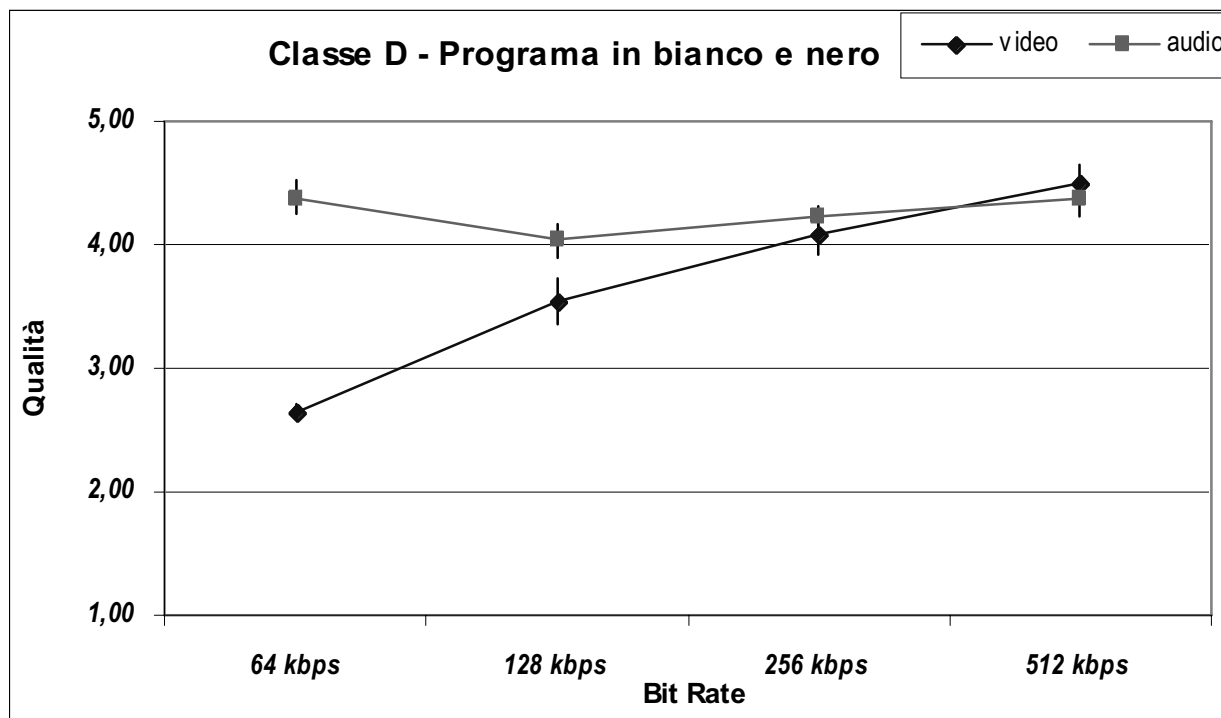


Figura 12 - Valutazione della qualità Video ed Audio per contenuto "Programa in bianco e nero"

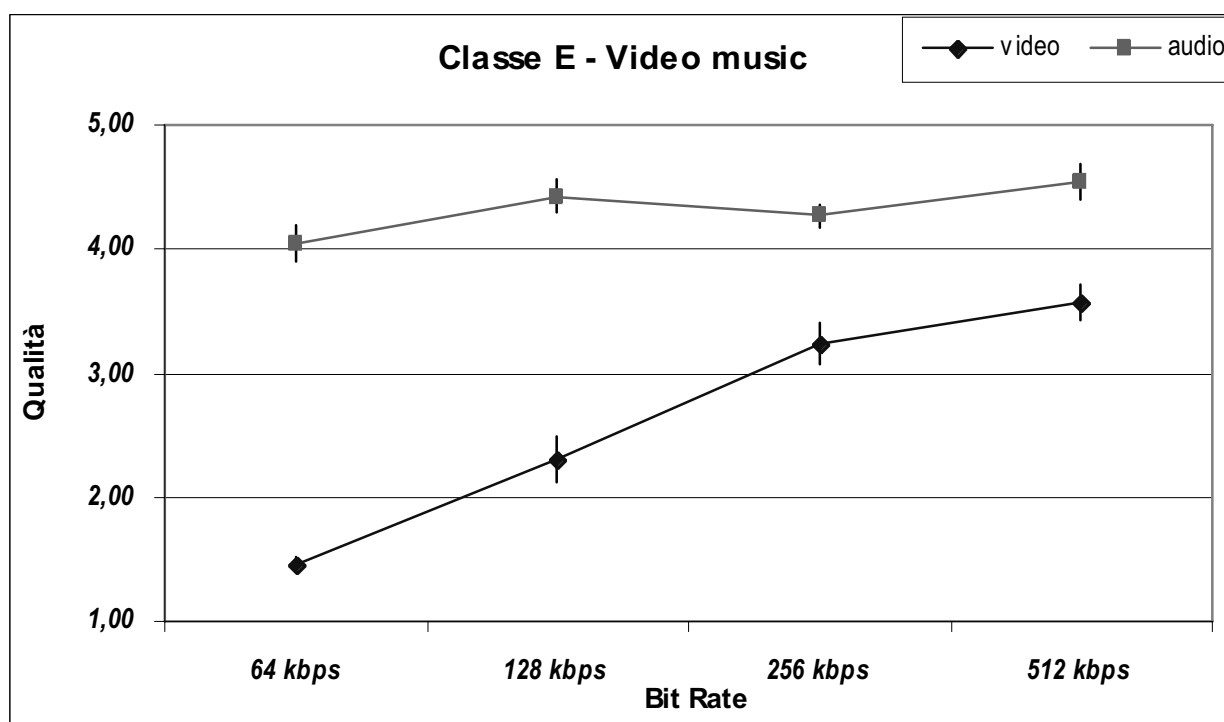


Figura 13 - Valutazione della qualità Video ed Audio per contenuto "Video Music"

CONCLUSIONI

I risultati del lavoro svolto hanno permesso di individuare la tipologia di connessione per utenza di tipo "Consumer" che permette di ottenere una qualità accettabile in funzione della banda disponibile all'utente e per raggiungere un servizio Multimediale attraverso una rete IP.

Inoltre è stato possibile individuare, all'interno di una valutazione di tipo globale, quali categorie di servizio presentano potenziali criticità in funzione della semantica del segnale ad ognuna di esse afferente.

Più in dettaglio categorie di servizio di tipo "Sport" risultano essere quelle maggiormente critiche anche per collegamenti ad alta capacità di canale, mentre collegamenti del tipo di quelli disponibili utilizzando una linea telefonica commutata permettono accessi "accettabili" solo ad alcune categorie di servizio.

Va infine messo in evidenza che il presente lavoro prescinde da tutte le tematiche connesse

con l'accesso di rete WEB, ovvero con tutti i possibili degradi o difficoltà di accesso ad un servizio "Consumer" afferibili a problemi di rete (accesso al server servizi, traffico in rete ecc.). Esso si limita quindi ad una individuazione delle condizioni ideali di accesso ad un servizio per il quale sia effettivamente disponibile la banda individuata nell'esperimento.

La tabella riportata qui di seguito indica per sommi capi i risultati dell'esperimento classificati in funzione della banda e della categoria di servizio. Per ogni punto di incrocio della tabella un numero, da 1 a 5, indica il livello di qualità ottenuto mediante le prove soggettive formali; valori pari o prossimi all'1 indicano un livello di qualità inaccettabile, mentre il valore 5 indica una situazione in cui il segnale viene percepito come esente da alcun tipo di degrado. Valori intermedi indicano situazioni intermedie di qualità, grossolanamente approssimabili con gli aggettivi "Mediocre" per il valore 2, "Accettabile" per il valore 3 e "Buono" per il valore 4.

Tabella riassuntiva dei risultati

Classe A									
	Video					Audio			
Bit rate	64	128	256	512	Bit rate	128	512	64	256
Qualità	1,54	1,96	2,85	4,08	Qualità	3,50	3,81	4,35	3,88

Classe B									
	Video					Audio			
Bit rate	64	128	256	512	Bit rate	128	512	64	256
Qualità	1,85	2,54	3,38	4,15	Qualità	4,00	3,92	4,35	4,54

Classe C									
	Video					Audio			
Bit rate	64	128	256	512	Bit rate	128	512	64	256
Qualità	1,12	2,23	3,23	4,38	Qualità	3,92	4,12	4,69	4,12

Classe D									
	Video					Audio			
Bit rate	64	128	256	512	Bit rate	128	512	64	256
Qualità	2,65	3,54	4,08	4,50	Qualità	4,38	4,04	4,23	4,38

Classe E									
	Video					Audio			
Bit rate	64	128	256	512	Bit rate	128	512	64	256
Qualità	1,46	2,31	3,23	3,58	Qualità	4,04	4,42	4,27	4,54

APPENDICE I – “FOGLIO DI VALUTAZIONE”

Vote 1		Vote 2		Vote 3		Vote 4	
Video	Audio	Video	Audio	Video	Audio	Video	Audio
Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>
Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>
Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>
Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>
Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>

Vote 5		Vote 6		Vote 7		Vote 8	
Video	Audio	Video	Audio	Video	Audio	Video	Audio
Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>
Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>
Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>
Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>
Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>

Vote 9		Vote 10		Vote 11		Vote 12	
Video	Audio	Video	Audio	Video	Audio	Video	Audio
Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>	Excellent <input type="checkbox"/>
Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>	Good <input type="checkbox"/>
Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>	Fair <input type="checkbox"/>
Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>	Poor <input type="checkbox"/>
Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>	Bad <input type="checkbox"/>

APPENDICE II – “CATEGORIE DI SERVIZI”

Le principali categorie di servizi utilizzate per la sperimentazione si possono schematizzare nel modo seguente:

- **News:** fornisce i servizi di informazione su eventi di carattere generale. I clienti possono fruire dei contenuti mediante meccanismo di download o streaming;
- **Intrattenimento:** permette di soddisfare il cliente nei momenti di svago. In questa categoria si includono ad esempio i video clip musicali e le informazioni su tematiche specifiche.
- **Informazioni Location Based:** in questa categoria rientrano gli avvisi di incidenti e di traffico, le previsioni meteo locali, la pubblicità, etc.

Le applicazioni con impatti video ritenuti interessanti sono riassunte nei seguenti gruppi:

- **videochiamata:** due utenti mediante telecamere collegate al terminale intergiscono a livello audio e video in tempo reale;
- **videoconferenza:** più utenti mediante telecamere collegate a terminale intergiscono a livello audio e video in tempo reale;
- **videomessaggio:** un utente riceve un videomessaggio;
- **videonews:** un utente riceve informazioni mediante video streaming i.e. telegiornale;
- **videonews finanziarie:** un utente riceve informazioni di carattere finanziario corredate da grafici;
- **videoguide:** un utente riceve informazioni turistiche mediante video streaming mentre visita una città.
- **videoclip musicali:** un utente scarica brani e video musicali (download) da fruire su terminale mobile;
- **Trailers:** a scopo pubblicitario alla possibilità di inviare brevi trailer cinematografici (circa 10 sec);
- **videoclip sportive:** un utente vede gli aspetti salienti di un evento sportivo o l'intero evento (gol di partite di calcio, incidenti-sorpassi di F1, ...);
- **localizzazione:** corredata da informazioni: punti di interesse turistico, farmacie, cinema, distributori di benzina, banche, ristoranti,...);

APPENDICE III – “ELENCO DEGLI SPOT”

Categoria di servizi	Bit rate originale
A = Cartoon	1500 bit/s
B = Sport	1500 bit/s
C = movie	1500 bit/s
D = bianco/nero	1500 bit/s
E = video music	1500 bit/s

APPENDICE IV – “CONDIZIONI DI CODIFICA”

Categoria di servizi	Bit rate	Bit rate	Bit rate	Bit rate	Compressione audio
A = Cartoon	64 Kbit/s	128 Kbit/s	256 Kbit/s	512 Kbit/s	1500 Kbit/s
B = Sport	64 Kbit/s	128 Kbit/s	256 Kbit/s	512 Kbit/s	1500 Kbit/s
C = movie	64 Kbit/s	128 Kbit/s	256 Kbit/s	512 Kbit/s	1500 Kbit/s
D = bianco/nero	64 Kbit/s	128 Kbit/s	256 Kbit/s	512 Kbit/s	1500 Kbit/s
E = video music	64 Kbit/s	128 Kbit/s	256 Kbit/s	512 Kbit/s	1500 Kbit/s