

HEVC

High Efficiency Video Coding

RESPONSABILE

VITTORIO BARONCINI

Nel corso del 2010, il gruppo ISO/IEC JTC1/SC29/WG11 - MPEG, *Moving Picture Expert Group*, ha avviato il processo di valutazione di nuove proposte per una tecnologia di compressione delle immagini, che costituisca un effettivo progresso rispetto allo standard MPEG-4 AVC (*Advanced Video Coding*), sia nel campo delle applicazioni televisive, dove il codificatore ha il ruolo di maggior rilevanza (*random access mode*, ossia piena possibilità di accesso in modo casuale al contenuto video), sia in quelle dei servizi mobili, dove è richiesta una simmetria di efficienza del codificatore in ricezione e in trasmissione (*low delay mode*, ossia decodifica a basso tempo di latenza). Questo processo di valutazione formale della qualità video è propedeutico alla selezione dei nuovi algoritmi da adottare nella definizione di un nuovo standard. Come da prassi, il gruppo MPEG giunge alla definizione del processo di valutazione di nuove proposte seguendo alcuni passi formali.

Il primo passo è, in presenza di uno standard precedentemente emesso da MPEG, la verifica che le nuove tecnologie forniscano prestazioni tali da giustificare un confronto formale con la tecnologia esistente. Questo passo prevede l'emissione di una *Call for Evidence* (CfE): un documento "pubblico", ossia aperto a proposte provenienti anche da enti e aziende normalmente non operanti in ISO, che descrive i termini per rispondere a un "invito" a sottoporre nuove tecnologie che dimostrino l'"evidenza di superiorità" rispetto allo standard esistente (in questo caso MPEG-4 AVC). La certificazione dell'evidenza di superiorità di nuove tecnologie si esegue, per prassi e tradizione, immediatamente prima e durante lo svolgimento della riunione di MPEG successiva a quella in cui è stata emessa la CfE. In questa sede, un gruppo di Test prepara una sessione di valutazione da parte di esperti (*expert viewing session*), alla quale partecipano esperti del settore video (estranei agli enti che hanno inoltrato proposte per la CfE); i risultati della valutazione eseguita dagli esperti permettono a MPEG di acquisire l'evidenza formale della superiorità della nuova tecnologia.

Come secondo passo, in caso di dimostrazione di evidenza di superiorità della nuova tecnologia, il gruppo MPEG procede all'emissione di una "Call for Proposal" (CfP), e cioè un invito formale a presentare proposte per la standardizzazione di una nuova tecnologia. L'emissione di CfP avviene attraverso la stesura di un documento "pubblico" che contiene una descrizione dettagliata delle condizioni necessarie all'inoltro di una "Proposta" e le modalità di valutazione a cui essa verrà sottoposta. MPEG esegue questo processo di valutazione su due livelli che convergeranno nella formazione del giudizio:

- valutazione formale soggettiva della qualità;
- valutazione oggettiva della tecnologia proposta mediante analisi della complessità e dei principi scientifici che la sottendono.

I risultati di questa valutazione permettono la scelta delle tecnologie che comporranno il nuovo standard.

Inizia quindi la “fase di collaborazione”, in cui tutti gli algoritmi giudicati validi concorrono nella creazione del nuovo standard. La valutazione oggettiva viene seguita durante la riunione di MPEG successiva all’emissione della “Call for Proposal”, mentre la valutazione soggettiva formale viene eseguita nei mesi che precedono.

Nel contesto del processo di valutazione delle nuove tecnologie rispetto allo standard MPEG-4 AVC, poi denominate “*High Efficiency Video Coding*” - HEVC, FUB ha avuto l’incarico di “Coordinamento delle attività”, nella persona del dott. Baroncini (Presidente del sotto-gruppo di Test di MPEG). FUB ha quindi progettato e coordinato sia la “Call for Evidence”, sia la “Call for Proposal”, effettuando un lavoro scientifico di progettazione di nuove metodologie di test e un’importante opera di pianificazione e organizzazione logistica delle prove soggettive formali.

DIMENSIONI DELLA PARTECIPAZIONE

La “CfE” [1] ha visto la partecipazione di 12 proponenti (un numero estremamente elevato se confrontato con le risposte alle precedenti “CfE”), che hanno sottoposto esempi di codifica per quasi tutte le condizioni di compressione previste. Un numero di partecipanti così elevato ha reso necessario per il gruppo di Test eseguire un drastico e profondo cambiamento del programma iniziale, sia nella tempistica di esecuzione delle valutazioni, sia nella stessa procedura di valutazione, come descritto nel paragrafo seguente. A questa fase hanno partecipato 3 centri di ricerca universitari (Leibniz University, RWTH Aachen, Zhejiang University) e 8 soggetti industriali (ARC, Huawei, LG, Mitsubishi, NEC, Samsung America, Samsung Korea, Toshiba).

La “Cfp” [2] ha visto la partecipazione di un numero di proponenti mai raggiunto prima in MPEG: 4 centri di ricerca universitari (ETRI, FhG-HHI, RWTH Aachen, NCTU/ITRI), 3 Broadcasters (NHK, BBC, SK Telecom) e 23 società industriali (Renesas, Samsung, Mitsubishi, Qualcomm, RIM, Tandberg, Ericsson, Nokia, Microsoft China, Toshiba, Jujitsu, Panasonic, Huawei, LG, MediaTek, JVC, Mitsubishi, Intel, SHARP, NEC, Sony, Hitachi, Texas Instruments), che hanno contribuito (da sole o in consorzi) alla formulazione di 27 proposte di nuove tecnologie per la compressione video.

Va rilevato che lo sforzo richiesto per l’esecuzione dei test della “Call for Proposals” andava ben oltre la semplice esecuzione delle sessioni di prove soggettive. In effetti, FUB ha avuto l’onere di gestire anche l’enorme mole di file video (più di 4000) ricevuti dai Proponenti; questo lavoro ha comportato un’attenta verifica di ciascun file al fine di evitare l’inserimento nei test di materiale codificato in modo errato o incompleto, che avrebbe complicato la successiva fase di valutazione formale, sino a renderne parzialmente inutilizzabili i risultati.

Una tale massiccia adesione non poteva essere sostenuta da un solo laboratorio di test; quindi FUB, pur mantenendo la responsabilità di progettazione dei test, ha deciso di coinvolgere nella fase di esecuzione altri due laboratori di livello internazionale: la EBU (*European Broadcasting Union* - Ginevra) e la EPFL (*Ecole Polytechnique Fédérale de Lausanne*). Sono invece rimasti a carico di FUB l’elaborazione statistica dei dati e la stesura del rapporto finale dei test.

ATTIVITÀ SVOLTE

Il Progetto ha richiesto, nell'arco di sei mesi, lo svolgimento delle attività di: ricezione e verifica dei dati (dei Proponenti); progettazione dei test; svolgimento dei test in FUB e nei due laboratori esteri (EBU e EPFL); raccolta ed elaborazione dei dati; stesura del rapporto finale.

RICEZIONE E VERIFICA DEI DATI

La particolarità del test HEVC è rappresentata non solo dall'elevato numero di proposte ricevute da MPEG (in totale 27, come già detto), ma anche e soprattutto dal gran numero di variabili (*bit rate* di codifica e sequenze di test) che la fase competitiva ha richiesto.

Il testo della CfP prevedeva cinque classi di formato video (corrispondenti a diverse tipologie di applicazione), e, per ogni classe, quattro bit rate di codifica, oltre a due modalità di progetto del codificatore (*random access* e *low delay*). Da ciò consegue che ogni proponente ha dovuto fornire a MPEG un totale di 135 video clip codificate; inoltre MPEG, per ottenere un metro di paragone con la tecnologia precedente (MPEG-4 AVC) ha confezionato 2 proposte di riferimento, dette *anchors* (“ancore”), portando a 29 il numero totale di proposte da valutare; di conseguenza il numero totale di video clip (ognuno lungo 10 secondi) da valutare è salito a oltre i 3400.

Una tale mole di dati ha reso obbligatorio il ricorso alla spedizione di dischi rigidi per effettuare lo scambio dei file video fra i Proponenti, il coordinatore dei test e i laboratori di test. La logistica di questo scambio è stata affetta da numerosi imprevisti legati alle spedizioni, alla rottura dei dischi e all'errata (o mancante) presenza di file nel disco. Queste difficoltà hanno richiesto una lunga ed estenuante opera di verifica di ogni singolo file video ricevuto e, in definitiva, un nuovo consistente scambio di dati, fra i Proponenti e la FUB, in parte tramite comunicazione elettronica in parte con rispedizione di dischi rigidi. Alla fine del processo di verifica si contavano circa 4000 file esaminati e un totale di 40 dischi rigidi inviati fisicamente, oltre a numerosi altri file scambiati per via telematica.

PROGETTAZIONE DEI TEST

L'elevato numero di sequenze video da valutare ha portato alla progettazione del più grande test formale di valutazione soggettiva di video mai realizzato, non solo nella storia di MPEG, ma nella storia di tutte le attività che hanno avuto luogo in questo campo. Questa situazione ha comportato enormi difficoltà, sia dal punto di vista logistico sia, soprattutto, dal punto di vista progettuale.

Per comprendere la difficoltà progettuale è necessario ricordare che un test di valutazione soggettiva non può durare più di 20 – 25 minuti, se si vuole evitare che i volontari che partecipano al test si affaticino e, di conseguenza, producano risultato poco attendibili. Il progetto di un test prevede quindi la suddivisione di tutti i video da valutare in “sessioni”, ossia spezzoni di valutazione separati, che permettano ai volontari di essere rimpiazzati da altri volontari e di riposare prima di eseguire un altro test.

Dato che una sessione di test permette di valutare, a seconda del metodo di test, dai 20 ai 50 video, il progetto complessivo del test ha portato alla definizione di un totale di oltre 136 sessioni di test.

SVOLGIMENTO DEI TEST IN FUB, EBU ED EPFL

Al test hanno partecipato, come volontari, oltre 800 studenti provenienti dalle università di Ginevra, di Losanna e di Roma. Ognuno dei tre laboratori di test ha coinvolto un elevato numero di "assistanti". FUB si è avvalsa di circa 10 persone, per oltre cinque mesi di attività complessiva. Per dare un'idea delle difficoltà logistiche e di progettazione di questo test, è utile ricordare che il test di maggiori proporzioni mai realizzato prima era stato eseguito, sempre da MPEG, nel 1995, per la fase di avvio del processo di standardizzazione della prima versione di MPEG-4; questo test comportò un totale di 13 sessioni di test, e cioè meno di un decimo delle sessioni di test eseguite per il test HEVC.

La complessità progettuale e logistica di questo esperimento non permette di fornire in questo contesto un rapporto completo nei dettagli; una descrizione completa di tutti gli aspetti progettuali e logistici del test può essere reperita nel rapporto conclusivo del test HEVC emesso dal gruppo MPEG [3].

RACCOLTA ED ELABORAZIONE DEI DATI E RAPPORTO FINALE

Una fase molto complessa e delicata è stata la raccolta dei dati *raw* (grezzi) dai laboratori e la loro elaborazione statistica. Per ogni sessione si è reso necessario riordinare i dati grezzi secondo il tipo di immagine e il bit rate di compressione (la progettazione di un test prevede la presentazione "casuale" del materiale video).

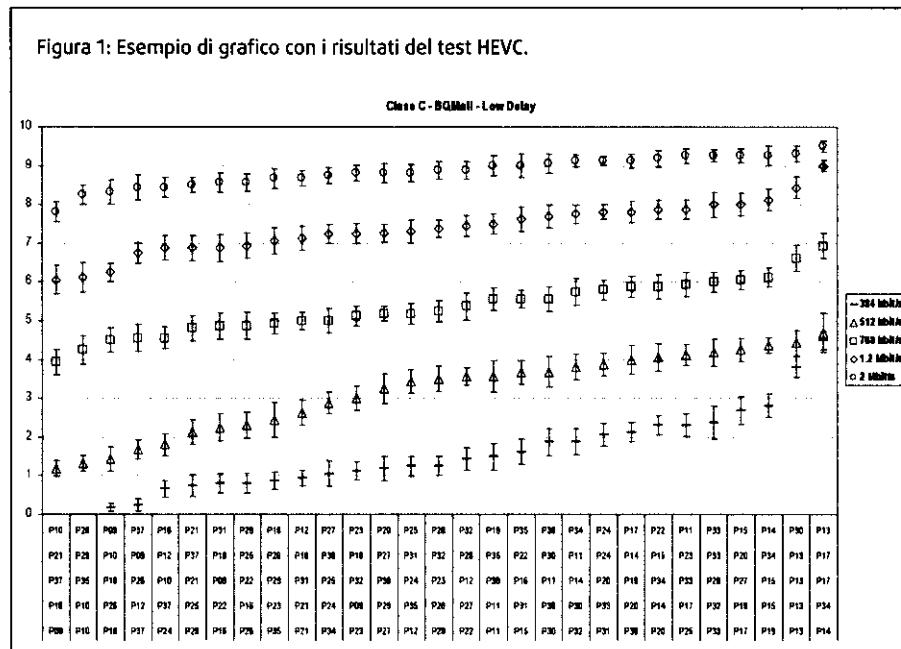
Inoltre, essendo stati utilizzati due differenti metodi di valutazione della qualità soggettiva formale (il DSCQS¹ e il DSIS²), è stato necessario rendere omogenee le diverse tipologie di dati grezzi forniti dai due metodi, in un formato unico, che ha consentito una valutazione delle proposte con una scala di qualità a undici livelli da 0 a 10 (dove il livello "0" rappresenta la qualità più bassa, mentre il livello "10" rappresenta l'eccellenza). I dati normalizzati e opportunamente raggruppati in funzione delle singole sequenze video e delle classi di servizio, hanno permesso di ottenere una rappresentazione tabellare e grafica dei risultati.

Di particolare interesse e utilità sono risultati i grafici comparativi fra la precedente tecnologia e le nuove tecnologie sinteticamente indicate con i nomi dei Proponenti. Questi non sono stati individuati in modo esplicito, ma solo attraverso i codici da P11 a P38. Solo quelli che lo hanno espressamente richiesto sono stati pubblicamente identificati, ma solo internamente a MPEG.

I grafici comparativi delle tecnologie sono stati ottenuti approntando, per ogni sequenza di test e per ogni classe di servizio, un diagramma nel quale compaiono tutti gli attori in gioco (Proponenti e *anchor*). Un esempio di grafico è riportato nella Figura 1, che si riferisce al caso della sequenza di test "BQMall", per la Classe C (WVGA) e nel caso di condizione "low delay" (codificatore e decodificatore a complessità bilanciata). Le cinque curve rappresentate si riferiscono ai differenti *bit rate* di compressione del video (da 384 Kbps fino a 2 Mbps); i codici P10 e P09 indicano la prestazione fornita dagli *anchor*, ossia dalla precedente tecnologia (MPEG-4 AVC).

¹ ITU-R (ITU Radiocommunications Sector). Methodology for the subjective assessment of the quality of television pictures. Raccomandazione ITU-R BT 500-11, Ginevra, settembre 2009.

² ITU-T (ITU Telecommunications Sector). Subjective video quality assessment methods for multimedia applications. Raccomandazione ITU-T P.910, Ginevra, aprile 2009.



Anche se il grafico illustrato costituisce solo un esempio, esso permette di avere una visione immediata di come la tecnologia precedente (MPEG-4 AVC), individuata dai codici P09 e P10, si collochi sempre ai limiti inferiori della scala di qualità, risultando ampiamente inferiore alle prestazioni delle nuove tecnologie che concorreranno alla creazione dello standard HEVC.

PUBBLICAZIONI

Nell'ambito del Progetto di test di sequenze video HEVC, FUB ha coordinato la produzione dei seguenti documenti:

- [1] MPEG Test, Video and JCT-VC. *Call for Evidence on High-Performance Video Coding* (HVC), Doc. ISO/MPEG N10553, Maui, USA, aprile 2009.
- [2] MPEG Test, Video and JCT-VC. *Draft Call for Proposals on High-Performance Video Coding* (HVC), Doc. ISO/MPEG N10926, Xi'an, Cina, ottobre 2009.
- [3] MPEG Test, Video and JCT-VC. *Report of Subjective Test Results of Responses to the Joint Call for Proposals (CfP) on Video Coding Technology for High Efficiency Video Coding* (HEVC), Doc. ISO/MPEG N11275, Dresden, Germania, 2010.

PROGETTI FINANZIATI DA INIZIATIVE NAZIONALI O REGIONALI

- **IRMA**
Intelligent Retrieval in Multimedia Archives
- **Speaky Acutattile**
- **Studio, sperimentazione e verifica delle funzionalità
dei decoder**

IRMA

Intelligent Retrieval in Multimedia Archives

RESPONSABILI

ANDREA BERNARDINI
FEDERICA MANGIATORDI

Il Progetto “IRMA - Intelligent Retrieval in Multimedia Archives” rientra tra gli interventi di sostegno pubblico agli investimenti da parte delle imprese previsti dal Bando POR FESR Lazio 2007-2013 nell’ambito del settore-filiera dell’audiovisivo per favorire la crescita armonica e sostenibile del territorio. La valutazione, il monitoraggio e il cofinanziamento del Progetto sono di competenza di Sviluppo Lazio S.p.A., società di scopo della Regione Lazio per il sostegno all’innovazione tecnologica. Il Progetto ha avuto inizio con la pubblicazione delle graduatorie in esito al suddetto Bando sul Bollettino Ufficiale della Regione Lazio (BURL) n.33 del 07/09/2010-Supplemento ordinario n.155, e dovrà essere realizzato in due anni. Nella realizzazione del Progetto, saranno coinvolte varie figure professionali (ricercatori, tecnici di ricerca, analisti e sviluppatori di software), per un impegno complessivo pari a 491 mesi/persona. L’impegno FUB è di 122 mesi/persona.

“IRMA” intende creare strumenti avanzati per modernizzare le tecniche di restauro, memorizzazione e pubblicazione del patrimonio filmico disponibile nei fondi di archivio presenti nel territorio della Regione Lazio. Gli strumenti creati permetteranno di intervenire, a più livelli, su materiale audiovisivo memorizzato con supporti obsoleti o a rischio (in particolare pellicole e primi formati elettronici), per digitalizzarli e per organizzarli ai fini di una sua migliore fruizione attraverso un portale di pubblicazione. La complessità ed eterogeneità delle finalità del Progetto richiede lo sviluppo di una piattaforma che integri avanzate tecnologie informatiche e innovative metodologie scientifiche per la metadatazione automatica, la classificazione semantica, l’indicizzazione e il recupero di dati multimediali.

Tale piattaforma si concretizza in un ecosistema IRMA caratterizzato da un approccio pluridisciplinare in cui algoritmi innovativi di metadatazione automatica e classificazione semantica per l’indicizzazione e il recupero dei dati multimediali sono integrati con avanzate tecnologie informatiche, ponendo attenzione alla definizione e ottimizzazione del processo end-to-end.

Al Progetto partecipano varie realtà del settore della ricerca e aziende con una spiccata propensione all’innovazione: RAI (CRIT – Teche), Fondazione Ugo Bordoni, Scai Connect, Eurix e Primesail. Al fine di conseguire gli obiettivi del Progetto, le Parti interessate hanno costituito il 5/11/2011 un’ATI (Associazione Temporanea di Imprese), di cui FUB è capofila. Oltre alla conseguente responsabilità di gestione del progetto, FUB ne ha in carico anche il coordinamento tecnico.

QUADRO DI RIFERIMENTO

L'attuale evoluzione dei paradigmi di fruizione e generazione dei contenuti multimediali attraverso i canali digitali dell'informazione sta portando alla luce, oltre che interessanti opportunità di business per gli operatori delle infrastrutture di distribuzione dei contenuti, alcune criticità fondamentali. Il fenomeno al quale si assiste è che, all'aumento progressivo delle capacità trasmissive e della capillarità dell'infrastruttura di telecomunicazioni, non corrisponde altrettanta capacità di collocare sul mercato dell'informazione multimediale contenuti adeguati alle aspettative degli utenti. Nonostante negli ultimi anni sia notevolmente cresciuto, a livello nazionale e internazionale, l'interesse verso queste tematiche, rimangono irrisolte alcune criticità essenziali per il pieno sfruttamento di questi valori sul mercato dell'informazione.

Il principale ostacolo è rappresentato dalla parziale o totale inaccessibilità dei contenuti stessi, dovuta alle tecnologie di conservazione e memorizzazione obsolete o comunque non adatte ad essere integrate economicamente e tecnicamente nei moderni processi produttivi, completamente basati sugli sviluppi e sui componenti della tecnologia dell'informazione.

In secondo luogo, anche dando per scontata la raggiungibilità del pieno accesso all'oggetto d'archivio rappresentato in forma digitale, cioè l'accesso all'oggetto che contiene il valore, un'altra non meno importante problematica è rappresentata dall'inaccessibilità al contenuto espresso dagli oggetti d'archivio, cioè dalla mancata o insufficiente conoscenza di che cosa è contenuto e, in ultima analisi, di che cosa è il valore. Questo secondo aspetto, sebbene in prima istanza collocabile ad un livello meno complesso dal punto di vista ingegneristico e sistemistico, è caratterizzato da un'estrema complessità concettuale e operativa, perché le dimensioni di rappresentazione del contenuto espresso, altrimenti definibile come semantica del contenuto, sono molteplici e in linea di principio incontrollabili.

OBIETTIVI DEL PROGETTO

"IRMA" intende dimostrare, in maniera complementare e non conflittuale alle correnti tendenze, come un'effettiva valorizzazione delle opportunità di business legate alla distribuzione di contenuti digitali, possa passare anche attraverso la valorizzazione dei patrimoni audiovisivi a elevato contenuto culturale, le cui potenzialità in termini di valore aggiunto per i nuovi servizi non sono state ancora pienamente valutate e sfruttate.

Collocandosi in questo contesto, "IRMA" mette in opera tecnologie e infrastrutture ad alto contenuto innovativo per la valorizzazione del patrimonio audiovisivo nella moderna era dell'informazione digitale, contribuendo in maniera sostanziale a risolvere le problematiche di accessibilità degli oggetti d'archivio sia dal punto di vista della rappresentazione, sia del contenuto espresso.

Nel Progetto è prevista la creazione di una catena completa di lavoro che recupera e valorizza il materiale audiovisivo registrato su supporti obsoleti tramite digitalizzazione e memorizzazione in sistemi ad alta affidabilità e capacità. I filmati audiovisivi, annotati tramite sistemi di metadatazione automatica e contestualizzazione manuale sono poi organizzati e aggregati per renderli fruibili attraverso un portale di pubblicazione. La creazione del portale integra tipologie di ricerca distribuita su diversi livelli informativi: ricerca anagrafica, di contenuto

testuale e di contenuto multimediale. La ricerca testuale è implementata utilizzando sia il paradigma di indicizzazione full text, sia paradigmi di interrogazione basati sull'elaborazione computazionale del linguaggio naturale.

Il Progetto comprende le seguenti tematiche di ricerca: l'individuazione automatica di specifici concetti semanticci, sulla base delle caratteristiche estratte dall'analisi numerica del contenuto video, nell'ambito della metadattazione automatica; l'integrazione degli indici di contenuto e del sistema di metadattazione automatica; l'integrazione degli indici con contenuto multimediale e degli indici relativi alle interrogazioni sottomesse al sistema; l'integrazione degli indici di contenuto e del sistema di metadattazione automatica.

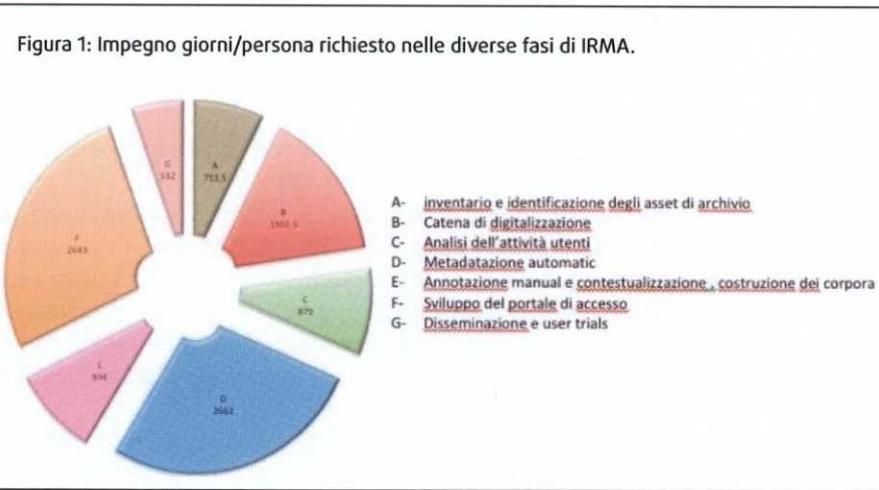
L'obiettivo di modernizzare le tecniche di restauro, memorizzazione e pubblicazione del patrimonio filmico disponibile è conseguito ricorrendo ad architetture complesse che richiedono elevati livelli di integrazione tra componenti software e pongono problemi di interoperabilità. "IRMA" fronteggia queste problematiche con l'adozione di metodologie, protocolli e formati standard. Tale pratica costituisce un punto d'innovazione del Progetto per quanto riguarda i seguenti aspetti:

- verifica e messa in opera di componenti/servizi/sottosistemi aderenti a standard internazionali e conseguente *benchmarking* e valutazione degli standard adottati rispetto alle esigenze applicative concrete del Progetto;
- costruzione di *framework* basati su metodologie, protocolli e formati standard che rappresentano embrioni di prodotti da collocare sul mercato nazionale e internazionale;
- conservabilità e accessibilità a lungo termine dei contenuti, assicurata dall'adozione di formati di codifica e incapsulamento standard.

APPROCCIO TECNICO

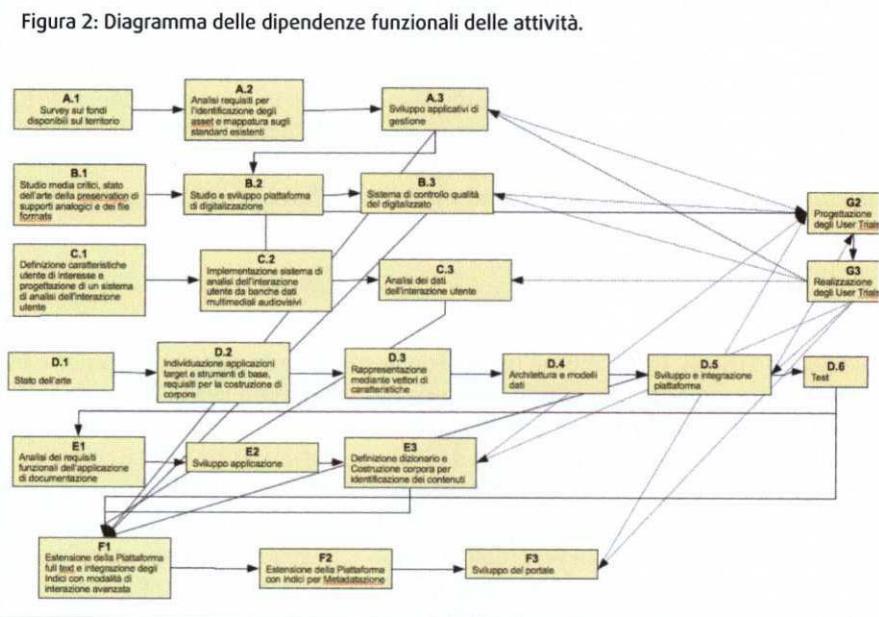
Per la realizzazione del Progetto sono state definite varie fasi:

- A) inventario e catalogazione anagrafica delle opere audiovisive e dei fondi d'archivio, con particolare attenzione ai contenuti d'archivio che versano in stato di conservazione critico (ad es. pellicole, primi formati elettronici);
- B) recupero del contenuto attraverso digitalizzazione e memorizzazione in sistemi ad alta capacità e affidabilità;
- C) analisi dell'interazione utente attraverso lo studio dei log delle interrogazioni del sistema da parte degli utenti durante la ricerca di contenuti multimediali negli archivi audiovisivi;
- D) applicazione di tecniche di elaborazione automatica del contenuto tese a identificare categorie semantiche e concetti, entità informative e documentali di interesse attraverso l'analisi numerica del contenuto digitalizzato;
- E) progetto e sviluppo di interfacce di annotazione del contenuto audiovisivo finalizzate all'integrazione e contestualizzazione delle informazioni estratte dal processo di elaborazione automatica;
- F) messa in opera di strumenti avanzati di pubblicazione delle informazioni multimediali ottenute dal processo di digitalizzazione e documentazione attraverso la progettazione e lo sviluppo di portali di accesso al contenuto multipiattaforma;
- G) disseminazione, progettazione e attuazione di test utenti.



In Figura 1 è indicata la distribuzione dell'impegno di risorse umane nelle diverse fasi attuative del Progetto in termini di giorni/persona.

Il programma attuativo di “IRMA” comprende, nel corso di ogni fase, numerosi momenti di verifica e di condivisione dei risultati sotto forma di rilasci software e documentazione tecnica delle attività di ricerca e sviluppo al fine di assicurare una perfetta integrazione dei diversi componenti e assicurare l’aderenza al progetto presentato. Ogni fase realizzativa è suddivisa in diversi task, per un totale di 23, logicamente e funzionalmente dipendenti come descritto nel diagramma del Progetto (Figura 2) .



ATTIVITÀ SVOLTA NEL 2010

I lavori del Progetto “IRMA” hanno interessato, come già detto, solo l’ultimo quadrimestre del 2010 e hanno riguardato:

- la gestione e il coordinamento dei procedimenti necessari e propedeutici alla corretta pianificazione del progetto;
- la predisposizione di strumenti di analisi dei diversi fondi presenti negli archivi audiovisivi disponibili sul territorio (Teche RAI, Archivi Vaticani, Discoteca di Stato, Cineteca Nazionale, Istituto Luce, Archivio di Stato), al fine di censire i materiali, individuare le criticità dei supporti e le priorità di recupero;
- l’analisi e lo studio delle procedure di catalogazione degli standard per la gestione dei diritti d’autore e l’identificazione e catalogazione del materiale negli archivi fisici;
- la definizione delle caratteristiche di utente per progettare un sistema di analisi dell’interazione utente. In questo ambito sono state esplorate le più recenti tecniche di analisi e profilazione delle attività degli utenti attraverso le interrogazioni presentate ai motori di ricerca.
- l’analisi e la presentazione delle funzionalità e delle problematiche tecnico-scientifiche correlate con la creazione e la gestione di archivi digitali multimediali.

Lo scopo è migliorare le prestazioni del sistema di navigazione dei contenuti e individuare tendenze generali o relative a nicchie di utenti. Quest’analisi si avvale anche del ricco patrimonio di dati presente nei sistemi di consultazione degli archivi RAI (Teche RAI). Va precisato che, più è esteso il numero e la ricchezza di dettagli dei dati di navigazione, maggiore è la possibilità di utilizzarli per una accurata predizione degli interessi e degli obiettivi degli utenti. La distribuzione delle ricerche consente di:

- riconoscere nuovi interessi dell’utente;
- rilevare comportamenti usuali o sconosciuti;
- identificare tendenze frequenti o nuove nella ricerca di contenuti;
- individuare nuovi segmenti di mercato.

Particolare attenzione è stata prestata alle metodologie per l’analisi numerica del segnale multimediale diffusamente utilizzate in contesti internazionali. Sono stati presentati in sintesi: il quadro strumentale matematico (*wavelet circolari armoniche*); i principi teorici entro cui si muove l’attività di ricerca scientifica finalizzata alla metadatazione automatica nell’ambito del Progetto “IRMA”.

PROSPETTIVE

L’ecosistema IRMA, inteso come unione coerente di processi, competenze e infrastrutture tecnologiche scalabili, prevede di consolidare la sua esperienza instaurando un paradigma di prestazione di servizio integrata *end-to-end*, costruito sulla base delle applicazioni da realizzare e delle tipologie di materiale da trattare e adattabile ai requisiti del singolo cliente. La struttura organizzativa per l’implementazione e la pianificazione delle attività prefigura la costituzione di un centro di competenza pluridisciplinare, che gestisca complessivamente tutte le fasi del processo di digitalizzazione e valorizzazione informativa dei contenuti rendendoli disponibili sul territorio. La sinergia nascente dall’interazione e dalla crescita professionale delle competenze umane sui vari servizi gestiti dal

centro di competenza, e l'evoluzione tecnologica dei processi e delle tecnologie offerte dal servizio (grazie alla continua applicazione sui più svariati campi), costituiranno un elemento distintivo d'innovazione rispetto al tradizionale paradigma basato sulla cessione in licenza di un prodotto. Questa soluzione rappresenta un aspetto decisivo per il trasferimento attivo dei risultati del Progetto oltre i limiti temporali dello stesso.

SPEAKY ACUTATTILE

RESPONSABILE

ANDREA PAOLONI

Il Progetto muove dall'idea di contribuire al superamento del divario digitale (*digital divide*), che costituisce, per molte persone, un ostacolo al pieno accesso al mondo moderno digitale. Per consentire a queste persone (disabili, non vedenti o semplicemente anziani) di superare le barriere della tecnologia digitale, si propone un'interfaccia denominata Speaky Acutattile (nel seguito, semplicemente Speaky): un assistente intelligente vocale multimodale. Si tratta di una piattaforma digitale di sistemi e servizi innovativi mirante a realizzare una nuova interfaccia con il mondo digitale che sia realmente inclusiva, volta cioè a facilitare significativamente l'accesso ai contenuti e ai servizi digitali, in modo da includere nel target di tali servizi e contenuti 'tutti', anche coloro che per diversi motivi non sono dotati di competenze informatiche. Questo programma prevede sia attività di ricerca industriale, sia attività di sviluppo sperimentale di una nuova piattaforma tecnologica informatica che abiliti ai nuovi contenuti e servizi digitali, orientata in particolare ad applicazioni di domotica e con particolare attenzione alle necessità dei diversamente abili. Sarà una piattaforma aperta a potenziali interlocutori che potranno sviluppare contenuti e applicativi made in Italy per un mercato globale anche dal punto di vista linguistico, in quanto sarà una piattaforma localizzabile in oltre venti lingue. La piattaforma produrrà nuovi servizi basati su un'infrastruttura hardware e software di tipo client/server e garantirà l'accesso sia dall'interno della casa o dell'ufficio, sia in mobilità. I sottosistemi server e client saranno strutturati in modo da configurare, dal punto di vista del servizio, un tutt'uno trasparente all'utente. Il front-end è un avatar che colloquia con l'utente e ne accoglie le richieste, per soddisfare le quali utilizza risorse e contenuti locali o remoti a seconda della specifica necessità.

Al Progetto, che ha durata triennale e complessivamente impegna oltre 90 anni persona di 13 organizzazioni, partecipano 8 industrie (Konvergence, Cultorale, HP, GESI, Mediavoice, Netlearn, Loquendo, Kir) e 5 tra università ed enti di ricerca (Fondazione Ugo Bordoni, Università di Palermo, Università di Roma "La Sapienza", CNR-ISTC, CSP).

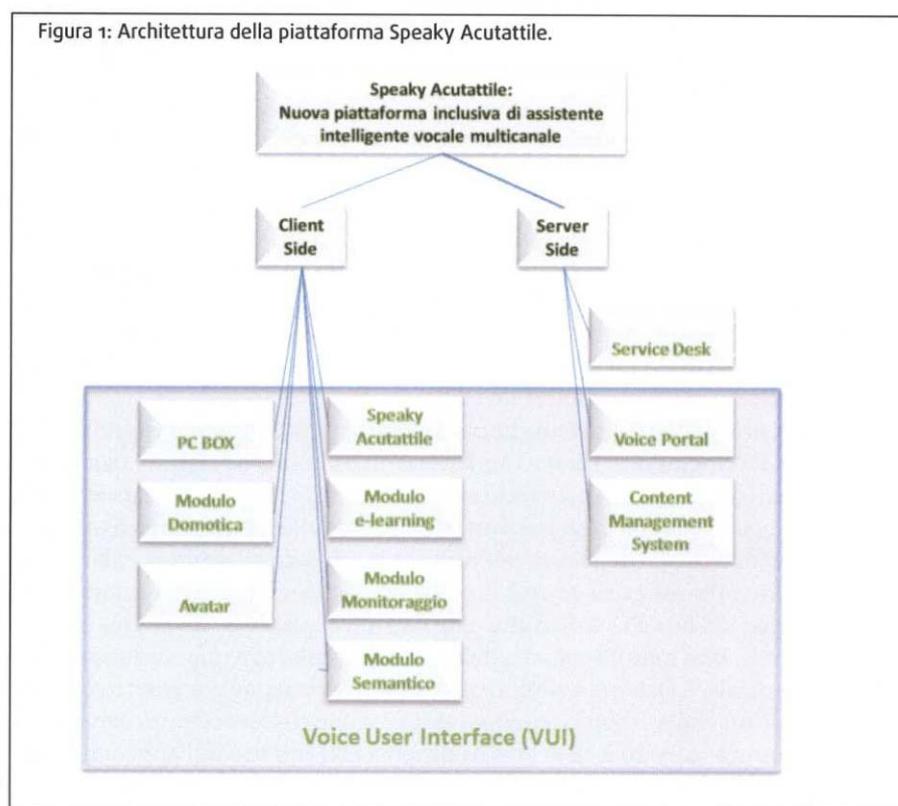
Obiettivi più specifici del Progetto, funzionali al macro obiettivo, sono:

- usabilità: costruire una piattaforma che sia potente e flessibile, su cui sia facile sviluppare nuovi moduli applicativi e contenuti, ma che sia soprattutto usabile, perché questo è uno dei principali problemi della moderna ingegneria;
- inclusività: pensare e realizzare una piattaforma che sia realmente abilitante ai nuovi servizi, a tal fine la stessa dovrà essere innovativa.

I partner del Progetto sono aziende italiane che hanno realizzato, e in diversi casi brevettato, i prodotti innovativi che sono i moduli costitutivi della nuova piattaforma.

APPROCCIO TECNICO E RISULTATI ATTESI

La nuova piattaforma consentirà la costruzione e fruizione di nuovi contenuti e servizi digitali; inoltre renderà disponibili nuove modalità di accesso ai servizi esistenti; la piattaforma garantirà l'accesso sia dall'interno della casa/ufficio, sia in mobilità. Sarà composta da diversi moduli hardware e software e avrà un'architettura di tipo client/server come può osservarsi nella Figura 1.



Lato client. Vi sarà un box *PC like* dotato di un Avatar con la più avanzata tecnologia di riconoscimento e sintesi vocale, accessibile a voce attraverso uno speciale e innovativo dispositivo wireless multifunzione, cuore della nuova piattaforma, che funge da telecomando/telefono. Questo dispositivo di input/output *mouse like*, facilita l'interazione a tutti, in particolare a ipovedenti e non vedenti. Collegato al box-PC vi sarà un avanzato sistema di sensori e attuatori domotici per l'interazione con l'ambiente casa (o ufficio), per gestione e controllo di impianti elettrici, illuminazione, riscaldamento, condizionamento, produzione di acqua calda sanitaria; sistemi per la generazione di energia, sistemi antintrusione e di riconoscimento; sicurezza e allarmi; automazione e movimentazione di cancelli e chiusure. Vi sarà inoltre un innovativo sistema di monitoraggio posturale: il sistema seguirà gli utenti in casa e, monitorandone la postura, potrà segnalare le situazioni critiche in cui la persona monitorata perde i sensi o cade. Nel sistema sono inoltre presenti dei sensori biometrici per la misurazione dei parametri vitali ai fini della telemedicina e teleassistenza.

Lato server. Sarà realizzato un completo centro servizi dotato di importan-

ti risorse di elaborazione e di archiviazione, connesso in banda larga con i vari box-PC client e dotato di un call center di operatori umani come secondo livello rispetto al sistema automatico, ove questo non fosse sufficiente per rispondere alle richieste degli utenti (si consideri ad esempio la telemedicina). Il centro servizi fungerà anche da centro di elaborazione dati e da Content Management System (CMS). Esso, inoltre, comprenderà un *content aggregator* per tutto il mondo multimediale; un sistema semantico che permette l'interazione in linguaggio naturale; un sistema di e-learning con tutto il *courseware*, i *learning objects* che possono essere richiesti dagli utenti; il gestore del modulo audiolibro con una grande quantità di testi digitali gratuiti o a pagamento, con il relativo sistema di e-commerce e payment e di gestione dei diritti digitali DRM (*Digital Rights Management*).

I sottosistemi server e client sono strutturati in modo da configurare, dal punto di vista del servizio, un'unica entità per l'utente: l'Avatar che parla lato client con l'utente, ne accoglie le richieste, per soddisfare le quali utilizzerà risorse e contenuti locali o remoti a seconda delle esigenze. La piattaforma sarà accessibile a voce anche da remoto via telefono per cui, sempre grazie a riconoscimento e sintesi vocale, potrà ricevere via telefono i comandi, come se l'utente stesse parlando direttamente al telecomando acutattile. L'utente si profilerà a voce in modo molto semplice e multimodale (con interfaccia adatta alle diverse disabilità); sarà quindi conosciuto dal sistema, che organizzerà in maniera adattiva l'interfaccia a seconda che si tratti di un utente ipovedente/non vedente, sordo e/o muto, disabile motorio, anziano o normodotato non anziano.

L'autorganizzazione del sistema implica l'uso di modalità di interazione diverse a seconda della tipologia di utente che lo usa. Ad esempio: se si tratta di non vedente, si organizzerà verso un massiccio uso del dispositivo acutattile; se ipovedente, utilizzerà il *device* acutattile insieme a un ingranditore video; se si tratta di un utente con difficoltà motorie degli arti, per attivare la sessione di input vocale, utilizzerà un macro pulsante *push-to-talk* e/o un pedale e lo specifico microfono direzionale a bordo del box-PC; se si tratta di un utente anziano, utilizzerà un ingranditore insieme a una minore velocità della sintesi vocale, con un'interfaccia semplice ed essenziale. Il sistema è sempre in grado di attivare, nei vari casi in cui fosse necessario, un collegamento con operatore umano presso il centro servizi.

L'architettura software e di servizi è composta da vari moduli applicativi verticali innestati su un *framework kernel* generale di interfaccia vocale contenente il motore intelligente in linguaggio naturale, l'avatar, il riconoscimento e la sintesi vocale, il sistema di interazione con la telefonia (*Computer Telephony Integration*) e il *Software Development Kit* (SDK), grazie al quale si potranno rendere accessibili via voce, in modo estremamente semplice e rapido, nuovi contenuti e applicativi 'vocalizzati', accessibili anche nella nuova 'modalità acutattile'.

STRUTTURAZIONE OPERATIVA

Allo scopo di fornire ulteriori approfondimenti sui risultati attesi, si descrivono brevemente le attività (dette *moduli*) di Progetto.

MODULO VOICE USER INTERFACE

L'idea di un computer capace di comprendere il parlato non è nuova: sono in molti a sostenere che solo un'interfaccia di questo tipo permetterebbe ai computer di entrare nella vita di tutti. Il Progetto, almeno per un limitato dominio di

attività, è volto all'implementazione di questa eccitante idea. L'obiettivo è di fornire al computer le competenze necessarie a permettergli di comunicare a voce con l'uomo. L'interfaccia dovrà elaborare il linguaggio parlato utilizzando i sistemi di riconoscimento (ASR) e sintesi della voce, ma anche un sistema di gestione del dialogo e un sistema di comprensione del linguaggio naturale. Per questo modulo, sul quale è concentrato in modo prevalente l'impegno FUB, si specificano i compiti qui approssimativamente descritti.

Task 1: Stato dell'Arte. Verrà effettuata un'analisi accurata dello stato dell'arte dei sistemi di dialogo o Agenti Conversazionali. Verranno studiati sia i modelli di sistema di dialogo attualmente in uso, sia quelli in fase di sviluppo presso i centri di ricerca.

Task 2: Descrizione dei sistemi e dei servizi di Speaky Acutattile. Verranno dettagliatamente descritti i sistemi e i servizi con cui Speaky deve interagire, mettendo in evidenza quali informazioni devono essere scambiate tra esso e il resto del mondo e con quale ordine devono essere fornite le informazioni. Per ciascun sistema o servizio dovranno essere formulati e verificati i modelli di funzionamento.

Task 3: Descrizione degli utenti. Verranno definiti i tipi di disabilità ai quali il sistema si rivolge per formulare modelli specifici, con l'obiettivo di migliorare in maniera significativa la qualità di vita delle persone disabili.

Task 4: Descrizione dei corpora. Verrà prodotta una collezione di corpora che sia rappresentativa di tutte le situazioni di interazione che Speaky deve gestire. Si dovranno quindi raccogliere i corpora relativi ai diversi servizi e farli poi confluire in un unico corpus dell'applicazione. Tali corpora dovranno altresì essere rappresentativi della tipologia di utenti ai quali Speaky si rivolge.

Task 5: Descrizione del "Mago di Oz". Verrà implementato, al fine di raccogliere i corpora sopra descritti, un sistema di simulazione del tipo "mago di Oz" utilizzando un numero adeguato di soggetti rappresentativi della popolazione che dovrà far uso di Speaky. In particolare è importante, dovendo riservare ai disabili alcune particolari applicazioni del sistema, disporre dei corpora vocali dei disabili.

Task 6: Descrizione dei database linguistici. Verrà studiata una descrizione delle modalità da utilizzare per la trascrizione e l'annotazione dei corpora vocali raccolti.

Task 7: Descrizione degli input non linguistici. Verranno dettagliatamente descritti tutti gli altri tipi di input che si prevede che l'interfaccia utente debba utilizzare, in particolare mettendo in evidenza le informazioni che da questi input possono essere raccolte.

Task 8: Descrizione dei protocolli di interscambio. Verrà prodotta una descrizione dettagliata dei vari protocolli di interscambio con i sistemi e i servizi che Speaky controllerà facendo uso di standard internazionali.