



Interfacce Utente Avanzate per l'e-learning

Luca Bevilacqua¹, Nicola Capuano², Annunziata Cascone², Federico Ceccarini¹, Fabio Corvino¹, Ciro D'Apice², Ivano De Furio¹, Giovanna Scafuro¹, Gianluca Supino¹

¹Engineering.IT, ²Università di Salerno, Dipartimento di Ingegneria dell'Informazione e Matematica Applicata

luca.bevilacqua@eng.it, ncapuano@unisa.it, cascone@diima.unisa.it, federico.ceccarini@eng.it, fabio.corvino@eng.it, dapice@diima.unisa.it, ivano.defurio@eng.it, giovanna.scafuro@eng.it, gianluca.supino@eng.it

Parole chiave: multimodal, e-learning, attentive use interface, affective user interface, perceptual user interface.

Abstract

In questo articolo si delinea un approccio originale all'interazione con i sistemi di e-learning che integra i più recenti risultati della human-computer interaction. In particolare si mostra l'applicabilità delle multimodal, attentive, affective e perceptual user interface per monitorare i comportamenti dello studente durante l'interazione con un sistema di e-learning con l'obiettivo di misurarne il livello di attenzione e coinvolgimento e di fornirgli tempestivamente il necessario supporto per attuare un processo di apprendimento efficace.

1 Introduzione

L'e-learning ha seguito negli ultimi anni un'evoluzione parallela a quella tecnologica, proponendo nuovi modelli che di volta in volta ne recepiscono le innovazioni. Ne è esempio il paradigma dell'apprendimento collaborativo in cui si affermano prepotentemente le tecnologie del social networking.

Nonostante ciò l'e-learning sconta ancora tutta una serie di difficoltà di fruizione oggettive e soggettive: la lunga abitudine alla formazione in aula, le possibili esperienze negative maturate con prodotti di prima generazione, un vissuto di formazione autogestita mal organizzata, la mancanza di un rapporto emozionale e quell'interazione tipica di una lezione frontale. Tali problemi ancora irrisolti nel rapporto emozionale e nell'interazione con il docente rappresentano una delle principali cause di abbandono dei corsi on-line. Un consistente contributo alla loro soluzione e al conseguente sviluppo dell'e-learning potrebbe arrivare dall'adozione di interfacce emergenti dagli studi sulla human-computer interaction (HCI).

In particolare tre paradigmi sembrano particolarmente interessanti:

- le Perceptual User Interface (PUI) che consentono di desumere le intenzioni dell'utente dall'osservazione dei suoi comportamenti sia espliciti sia impliciti;
- le Attentive User Interface (AUI) concepite per dedurre e gestire al meglio l'attenzione dell'utente;
- le Affective User Interface (AFUI) che analizzano lo stato emotivo dell'utente per adattarsi ad esso.

L'applicazione delle PUI, AUI e AFUI nei sistemi di e-learning consente non solo di facilitare il raggiungimento di risultati migliori rispetto a strumenti tradizionali ma anche di ottenere un maggiore coinvolgimento dei discenti in tutto il processo di e-learning. Come descritto in (Bangert-Drowns *et al.*, 2001), infatti, un discente coinvolto mostra di essere pienamente ed attivamente partecipe sia dal punto di vista comportamentale che intellettuale ed emotivo nelle attività didattiche.

Tuttavia l'influenza delle emozioni nei processi di apprendimento sembra essere ancora sottostimata. Solo recentemente si stanno avviando un sempre maggior numero di ricerche (e.g. Kort *et al.*, 2001; Currin, 2003) che pongono le emozioni al centro di ogni tentativo e risultato di apprendimento specialmente in e-learning. Tali studi rivelano l'importanza degli stati emozionali del discente e, in particolar modo, la relazione tra emozioni e apprendimento efficace (e.g. Kort *et al.*, 2001).

Alcuni studi (e.g. Isen, 2000) dimostrano altresì che uno stato d'animo positivo induce un modo diverso di pensare, caratterizzato da una tendenza verso una maggiore creatività e flessibilità nel problem solving ed una maggiore

efficienza e precisione nel decision making.

In definitiva mediante la percezione continua dello stato attentivo ed emozionale dello studente è possibile raccogliere informazioni utili per migliorare l'usabilità e la fruibilità del corso. Inoltre si rendono possibili interventi diretti da un lato, al recupero o mantenimento dell'attenzione dell'utente e dall'altro, a fornire un supporto psicologico e pedagogico allo studente in difficoltà durante la fruizione del corso stesso.

Engineering.IT collabora con il Dipartimento di Ingegneria dell'Informazione e Matematica Applicata dell'Università di Salerno al progetto di ricerca di rilievo nazionale WiSe¹, con l'obiettivo di investigare il potenziale di nuovi paradigmi per il miglioramento dell'interazione uomo-macchina. Il presente contributo intende descrivere le attività legate al progetto WiSe relativamente alla realizzazione di interfacce multimodali di tipo percettivo, attentivo ed affettivo e la loro applicabilità agli scenari di e-learning.

2 Lavori correlati

Esistono diversi tentativi in letteratura per determinare lo stato attentivo ed emotivo attraverso algoritmi di elaborazione di segnali che operano su dati provenienti da sensori più o meno sofisticati (Picard *et al.*, 2001; Whang *et al.*, 2003). Tali algoritmi forniscono risultati anche piuttosto accurati ma, di contro, richiedono l'impiego di tecnologie piuttosto invasive tra cui la rilevazione della conduttanza cutanea, il monitoraggio del battito cardiaco e la misurazione delle attività cerebrali.

Se tali tecnologie sono applicabili in alcuni ambiti ristretti, esse non sono soddisfacenti in altri ambiti, tra cui l'e-learning, in cui tali sensori possono distrarre gli utenti ed interferire con le loro attività primarie. Per tali ragioni, le ricerche in ambito e-learning cercano di limitare il campo d'azione a sensori non intrusivi tra cui telecamere e microfoni.

Un metodo ampiamente utilizzato per la percezione dello stato emotivo dell'utente è l'analisi delle espressioni facciali mediante elaborazione delle immagini. Applicando tali tecniche in (Neji & Ammar, 2007), ad esempio, è stato possibile rilevare l'insorgere di emozioni tra cui gioia, tristezza, rabbia, paura, disgusto e sorpresa.

Altre ricerche si basano sull'analisi della postura (Mota and Picard, 2003). Ad esempio in (D'Mello *et al.*, 2007) è stato utilizzato un sensore per la misurazione della pressione in diversi punti della seduta e dello schienale di una sedia per determinare lo stato attentivo ed affettivo di utenti che interagivano

¹ Il progetto WiSe "Wireless Services: an User Centric Approach" è finanziato dal MIUR in base Al D.Lgs. N. 297/1999. Iniziato a Giugno 2006 con l'obiettivo di condurre un programma sistematico di studi ed esperimenti volti ad investigare il potenziale industriale di nuovi paradigmi per il miglioramento dell'interazione uomo macchina (con particolare ma non esclusivo riferimento all'e-learning), esso si concluderà a Maggio 2010.

con un sistema di e-learning. Analizzando le mappe di pressione in output dal sistema è stato possibile addestrare il sistema a riconoscere tra 5 emozioni diverse (noia, confusione, coinvolgimento, gioia e frustrazione) con un'accuratezza del 70%.

Il gruppo AutoTutor dell'Università di Memphis ha identificato in (Craig *et al.*, 2004) alcuni collegamenti tra obiettivi didattici ed emozioni postulando che l'apprendimento è sempre associato ad un episodio emotivo.

In Kort (*et al.*, 2001) viene proposto un modello di apprendimento nel quale le emozioni sono rappresentate su un piano cartesiano e il discente si muove tra emozioni contigue seguendo un percorso a spirale. In tal senso il livello di apprendimento ed il livello di affettività varia a seconda del quadrante in cui lo studente si trova in un determinato istante.

L'Affective Computing Group del MIT sta invece investigando quali siano le correlazioni tra emozione, cognizione ed apprendimento (Burlison *et al.*, 2004). Altre ricerche tra cui (Fowler & Mayes, 1999) descrivono la relazione tra performance di apprendimento e livello di eccitazione emotiva come una curva a U invertita: le persone apprendono meglio quando le loro emozioni raggiungono un livello moderato.

Un ulteriore modello piuttosto diffuso per la valutazione delle emozioni in didattica è quello conosciuto come OCC (Ortony *et al.*, 1990). Questo modello specifica 22 categorie di emozioni basate sulle reazioni emotive di utenti a situazioni costruite ed è usato estensivamente per riconoscere e mappare le emozioni degli utenti durante l'interazione con giochi educativi.

3 WiSe e MAAPUI per l'e-learning

Lo studio delle PUI, AUI e AFUI nell'ambito del citato progetto WiSe ha condotto alla definizione di un modello unico che ingloba i tre paradigmi detto MAAPUI: Multimodal Attentive Affective Perceptive User Interface (Fig. 1). Il modello è del tutto generale e può essere utilizzato in differenti campi applicativi. Tra questi, di particolare interesse in WiSe, è l'e-learning (oggetto del presente lavoro).

L'approccio adottato prevede l'analisi dei pattern comportamentali dello studente acquisiti mediante l'osservazione di differenti modalità di interazione (sia implicite che esplicite) che egli realizza durante la sessione di e-learning.

Lo stato attentivo e affettivo dell'utente viene inferito:

- dall'analisi di una serie di caratteristiche estratte da segnali video, dai quali è possibile ricavare informazioni sulla postura e le espressioni del volto;
- dall'analisi dell'interazione dell'utente con la postazione di lavoro (tracking della tastiera, del mouse e delle attività su PC).

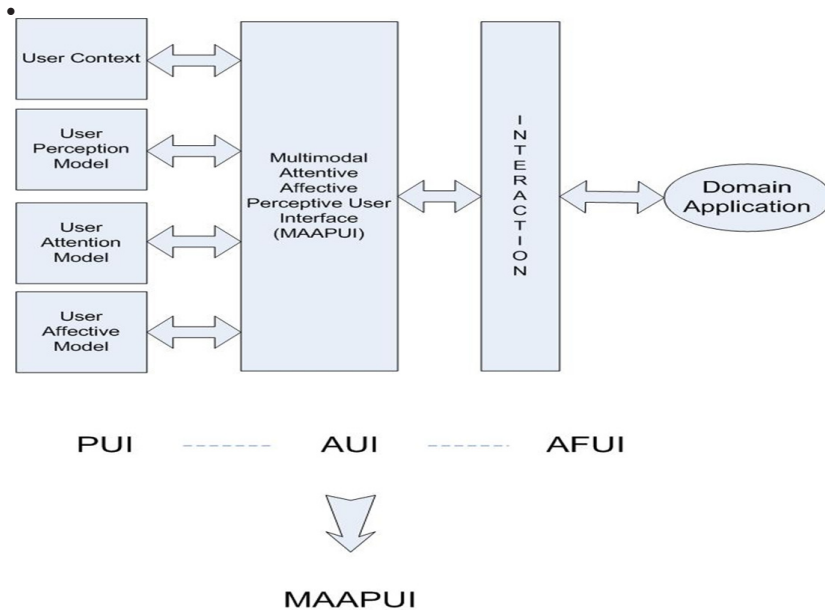


Fig. 1 Multimodal Attentive Affective Perceptual User Interface.

Tali informazioni sono integrate con quelle relative al profilo dello studente (preferenze, competenze, personalità e motivazione) al fine di alimentare un modello cognitivo dello stesso (Fig. 2).

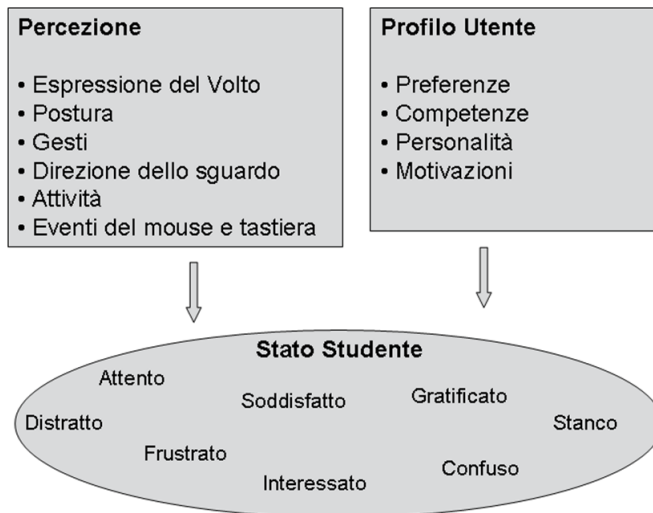


Fig. 2 Modello attentivo-emozionale dello studente.

I vantaggi di tale approccio sono numerosi e consentiranno di avvicinare l'esperienza dell'e-learning a quella classica dell'apprendimento frontale. Ad esempio, la percezione in tempo reale di una postura distratta o annoiata da parte dello studente può consentire al sistema di attuare qualche attività tesa al recupero dell'attenzione. Inoltre il monitoring continuo del processo di selezione delle risposte a un test di verifica dell'apprendimento, fornisce informazioni maggiori di quelle della sola analisi a posteriori di correttezza delle risposte.

Infatti, a prescindere dalla risposta fornita a un quesito, l'osservazione di come si sia arrivati alla soluzione (se frutto di ripensamenti, di tempi di valutazione elevati, di selezione casuale) in relazione agli skill dello studente, può fornire informazioni sull'esigenza di approfondimenti da parte dello studente o anche di miglioramento della forma espositiva del test.

Ancora, la verifica che lo studente stia effettuando ricerche su internet su temi inerenti l'argomento del modulo didattico può suggerire al sistema che lo studente necessita di dettagli o informazioni addizionali. Questo può migliorare il livello di soddisfazione dello studente o ridurre la sua frustrazione.

Come proposto nel citato modello di apprendimento a spirale (Kort *et al.*, 2001), durante il processo di apprendimento lo studente esperisce differenti emozioni e stati mentali. Il sistema si pone come obiettivo quello di monitorare lo studente operando opportuni interventi quando la permanenza di stati emotivi negativi (difficoltà di apprendimento, noia, etc.) possa causare il fallimento del processo di apprendimento.

La realizzazione di un tale sistema richiede la soluzione di una serie di problemi legati all'acquisizione dei segnali, all'identificazione delle features caratteristiche, alla loro classificazione ed al loro utilizzo all'interno di un modello cognitivo che consenta di inferire con un grado di affidabilità accettabile lo stato attentivo ed emotivo dello studente.

Per l'acquisizione dei segnali e nella scelta delle features caratteristiche il nostro approccio, sulla base delle considerazioni riportate nella sezione 2, è stato necessario concentrarci su dispositivi non invasivi e di larga diffusione quali webcam o microfoni. Nel seguito si analizzano in maggior dettaglio alcune delle tecniche adottate per l'estrazione delle features.

3.1 Analisi della Postura

La postura, cioè la posizione che ciascuno assume con il corpo, è un segnale non verbale involontario, meno controllabile del volto o del tono della voce e rappresenta un utile indizio sullo stato emotivo ed attentivo di un soggetto.

Svariate sono le posture che il corpo può assumere, ma se si considerano le posture sedute, assunte da un utente durante l'interazione con una postazione fissa, il campo si restringe notevolmente ed è possibile effettuare un'analisi

statistica tesa a correlare la postura seduta con lo stato emotivo ed attentivo come mostrato in (Mota & Picard, 2003). Partendo dai risultati di questi autori, in particolare, si sta implementando un modulo di percezione della postura che utilizza l'elaborazione di segnali video dell'utente ripresi da una webcam frontale.

Il problema di rilevare la postura viene di solito affrontato con tecniche di acquisizione automatica della forma di un oggetto. Tali tecniche risultano la scelta migliore quando si vuole ad esempio discernere tra posture erette, sedute, accovacciate, in cui i rapporti di forma sono notevolmente diversi. Nel caso di studente impegnato in una sessione di e-learning, invece, essendo la postura sempre seduta, assumono maggior peso la distanza dallo schermo e l'orientamento del capo rispetto al rapporto di forma. Pertanto si è scelto di derivare la postura mediante un modulo di head-tracking.

Esistono diverse tecniche per affrontare il problema dell'head-tracking. L'approccio adottato è tracking-based e fornisce un buon compromesso tra precisione e complessità computazionale. I metodi di tracking (metodi basati sull'inseguimento) operano seguendo il movimento relativo della testa tra frame consecutivi in una sequenza video.

Una prima fase di inizializzazione individua la faccia all'interno di un frame. In seguito, si effettua la ricerca di punti caratteristici e delle principali componenti che costituiscono il viso (naso, bocca e occhi). Individuati tali punti si utilizza un algoritmo di tracking (detto Optical Flow) per seguire la posizione di questi nei frame successivi. La conoscenza della posizione dei punti nell'immagine e la conoscenza della loro posizione su un volto consente di ricavare la stima della posa della testa.

3.2 Analisi dello sguardo

Il modo in cui si esamina visivamente una scena, il tempo utilizzato per analizzarla completamente, la durata della fissazione di un particolare, alcuni movimenti involontari dell'occhio, sono tutti aspetti legati allo stato cognitivo della persona. I movimenti oculari di un utente, quindi, possono essere utilizzati per studiarne l'esplorazione e l'attenzione visive.

L'interazione con l'ambiente circostante e, più in particolare, con un'interfaccia grafica proposta dallo schermo di un computer, rappresenta una comune situazione in cui si svolge un'esplorazione visiva. Fatte queste premesse appare chiaro come un'analisi basata sui movimenti oculari possa dare riscontri oggettivi e quantitativi nello studio del comportamento e in particolare del livello di attenzione di una persona impegnata in un'attività a videoterminale.

Il sistema di analisi dello sguardo (gaze tracking) messo a punto non richiede strumentazioni costose né invasive ed è basato sulla luce visibile. Esso si

fonda sull'utilizzo di una tecnica nota come Active Appearance Model (AAM) per l'individuazione e l'isolamento dei punti caratteristici di volti in foto e filmati.

La sperimentazione condotta è stata rivolta all'individuazione della posizione della pupilla rispetto all'occhio, allo scopo di determinare la direzione dello sguardo di un utente posto di fronte ad un personal computer. In particolare si sono analizzati i pattern caratteristici di un utente impegnato in un'attività di lettura di un documento.

3.3 Analisi delle espressioni del volto

Gli insegnanti esperti sono in grado di interpretare senza particolare sforzo le necessità dello studente, le sue sensazioni ed i suoi bisogni semplicemente cercando il loro sguardo e leggendo le loro espressioni. Di conseguenza essi sono in grado di intraprendere le azioni necessarie al recupero dell'attenzione oppure rivolte alla riduzione del senso di frustrazione del discente.

Per questo motivo è stata posta grande attenzione, all'interno del progetto WiSe, all'analisi e dall'interpretazione delle espressioni del volto umano. Il sistema proposto per l'identificazione delle espressioni del volto combina tre approcci:

- un'analisi olistica di singole immagini basata su tecniche di Principal Component Analysis, come strumento per la estrazione di features e riduzione di complessità, e sulle Support Vector Machine (SVM), come sistema per la classificazione;
- un'analisi spazio temporale basata sulle Motion Energy Maps, che utilizza anch'essa le SVM per classificare le mappe energetiche caratteristiche delle diverse emozioni;
- un'analisi delle espressioni mediante l'individuazione dei punti caratteristici del volto con AAM e la loro classificazione in termini del Facial Action Coding System.

Tutti e tre i modelli utilizzano tecniche di face detection ed un algoritmo per l'identificazione delle pupille per consentire un più fine allineamento e normalizzazione delle immagini. Si ritiene, e i risultati preliminari sono confortanti, che l'utilizzo simultaneo ed integrato di tali sistemi sia in grado di consentire il raggiungimento di un discreto livello di affidabilità per il riconoscimento delle espressioni del volto.

3.4 Analisi delle attività sulla postazione

Uno studente impegnato in una sessione di e-learning mette in atto una serie di interazioni implicite che possono essere raccolte attraverso un esame

dettagliato delle attività svolte sulla postazione di lavoro. Tali informazioni sono comunicate in maniera non intenzionale dallo studente, ma possono avere rilevanza nell'inferire il suo stato attentivo ed emozionale. Reattività, lucidità, stanchezza, tensione sono stati che influenzano la maniera in cui l'utente si pone dinnanzi al compito da eseguire.

Alla luce di queste osservazioni è stato realizzato un sistema per il tracking degli input da mouse e tastiera, che oltre a presentare caratteristiche di trasparenza (l'esperienza dell'utente non viene in alcun modo alterata) ed indipendenza dalla piattaforma software utilizzata, permette l'annotazione di interessanti dettagli dell'attività utente. Nella sperimentazione sul campo saranno inoltre valutate le possibili inferenze attentive desumibili dall'esame dettagliato dell'attività dell'utente.

L'attendibilità di questi parametri è influenzata anche da informazioni ambientali che possono alterare l'attenzione dell'utente direttamente, come nel caso di un ambiente particolarmente rumoroso, o indirettamente, come nel caso di un ambiente a temperatura elevata che potrebbe provocare stati di stress e affaticamento con riflessi negativi sull'attenzione.

Conclusioni

Parallelamente allo sviluppo dei primi prototipi di sistemi percettivi, in collaborazione con il Dipartimento di Ingegneria dell'Informazione e Matematica Applicata dell'Università di Salerno, è in corso di allestimento un laboratorio per l'osservazione del comportamento degli studenti.

Il laboratorio sarà attrezzato in modo che per ciascuna postazione di e-learning sarà possibile effettuare la registrazione a mezzo busto della figura dello studente, la registrazione del desktop, il tracking degli eventi di sistema (ivi inclusi quelli della tastiera e del mouse).

Con l'ausilio di insegnanti esperti si procederà all'analisi ed alla annotazione del materiale raccolto, in modo da costituire una sorta di base della verità. Dall'analisi del comportamento reale scaturiranno i necessari affinamenti dei modelli comportamentali e le successive validazioni dei modelli proposti.

BIBLIOGRAFIA

- Bangert-Drowns R.L., Pyke C. (2001), *A taxonomy of student engagement with educational software: An exploration of literate thinking with electronic text*. Journal of Educational Computing Research, 24(3), 213-234.
- Bull E. P. (1987), *Posture and Gesture*, Pergamon Press.
- Burleson W., Picard R. W., Perlin K., Lippincott J. (2004), *A platform for affective*

- agent research*, Proceedings of the 7th international Conference on Multimodal Interfaces, New York, ACM Press, 108-115.
- Craig S. D., Graesser A. C., Sullins J., Gholson B. (2004), *Affect and learning: An exploratory look into the role of affect in learning with Auto Tutor*, Journal of Educational Media, 29 (3), 241-250.
- Curran L. (2003), *Feelin' groovy: Experts now believe that e-learning must elicit positive emotions to succeed*, eLearn Magazine October 2003.
- D'Mello S. K., Picard R., Graesser A. C. (2007), *Toward an affect-sensitive Auto Tutor*, IEEE Intelligent Systems, 22(4), 53-61.
- Fowler C. J. H., Mayes J. T. (1999), *Learning relationships from theory to design*, Association for Learning Technology Journal, 7 (3), 6-16.
- Kort B., Reilly R., Picard R.W. (2001), *An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion*, Proceedings of International Conference on Advanced Learning Technologies (ICALT 2001), Madison Wisconsin, August 2001.
- Isen, A. M. (2000), *Positive affect and decision making*, In M. Lewis & J. Haviland (Eds.), Handbook of emotions (pp. 720), Guilford, New York: The Guilford Press.
- Mota S., Picard R. W. (2003), *Automated Posture Analysis for Detecting Learner's Interest Level*, Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI.
- Neji M., Ben Ammar M. (2007), *Agent-based Collaborative Affective e-Learning Framework*, The Electronic Journal of e-Learning 2 (5), 123-134.
- Ortony A., Clore G. L., Collins A. (1990), *The cognitive structure of emotions*, Cambridge, Cambridge University Press.
- Picard R.W., Vyzas E., Healey J. (2001), *Toward Machine Emotional Intelligence: Analysis of Affective Physiological State*, IEEE Transactions Pattern Analysis and Machine Intelligence 23 (10), 1175-1191.
- Whang M. C., Lim J. S., Boucsein W. (2003), *Preparing Computers for Affective Communication: A Psychophysiological Concept and Preliminary Results*, Human Factors, 45 (4), 623-634.
- Wilson, G. F., Russell, C. A. (2003b), *Real-time assessment of mental workload using psychophysiological measures and artificial neural networks*. Human Factors, 45, 635-643.