

Abilità socio-emotive per Agenti Virtuali dedicati all'e-learning

Elisabetta Bevacqua*, Maurizio Mancini*, Christopher Peters*, Catherine Pelachaud*, Magalie Ochs** e Nicolas Ech Chafai**

* IUT de Montreuil Université de Paris 8

e.bevacqua, m.mancini, c.peters, c.pelachaud@iut-univ.paris8.fr

<http://www.iut.univ-paris8.fr/greta>

** France Télécom, Division R&D, France

ABSTRACT

In questo capitolo ci proponiamo di descrivere il lavoro necessario alla creazione di agenti virtuali multimodali dotati di significative capacità comunicative e sociali.

Ci basiamo sulla nostra estesa conoscenza degli Agenti Conversazionali per implementare personaggi autonomi e dall'aspetto umano capaci di esibire comportamenti sociali credibili ed espressivi e di rivestire il ruolo di tutor in sistemi dedicati all'e-learning.

Siamo dell'idea che, per raggiungere il nostro scopo, dobbiamo dotare l'agente di capacità simili a quelle possedute dagli esseri umani. L'agente virtuale dovrebbe essere in grado di percepire il mondo reale e quello virtuale in cui opera e riconoscere le intenzioni dell'utente nel ruolo di studente e degli altri agenti se presenti. Dovrebbe, quindi, essere capace di intraprendere o lasciarsi coinvolgere in una conversazione, di portare avanti un'interazione comunicando sia attraverso la parola che l'uso di segnali non-verbali; l'agente dovrebbe essere anche in grado di adottare un comportamento espressivo, mostrando emozioni attraverso il corpo e la voce.

1. INTRODUZIONE

Nei sistemi dedicati all'e-learning gli agenti virtuali interattivi stanno diventando sempre più utili aprendo nuove possibilità e nuove interessanti sfide per la valutazione e l'applicazione di tecniche d'intelligenza artificiale a livello umano. Diversi agenti pedagogici sono stati implementati recentemente con lo scopo di facilitare il processo di apprendimento.

Adele, un agente progettato per operare nell'ambito dell'insegnamento a distanza via Web, è fornito di una serie di funzioni pedagogiche chiave per la spiegazione di un argomento e il monitoraggio dello studente sollecitando domande, suggerimenti e spiegazioni. Adele è dotato di capacità multimodali per una continua comunicazione con gli studenti (Johnson *et al.*, 1998). Un altro esempio è Amico, un prototipo sviluppato per migliorare l'insegnamento della matematica attraverso l'interazione dello studente con uno o più agenti virtuali (Rasseneur *et al.*, 2002).

Nuovi ambienti pedagogici chiamati *Serious Games* (Giochi Seri) stanno cominciando a prendere piede. In un ambiente virtuale con l'ausilio di agenti virtuali, i giocatori/studenti apprendono come agire in particolari situazioni: dominare un incendio, effettuare un'operazione chirurgica, gestire le crisi umanitarie a distanza, ecc...

I *Serious Games* adattano le tecnologie, sinora dedicate ai videogames, a domini diversi (quali l'amministrazione, la sanità, la difesa, l'industria e così via) per fornire un nuovo tipo d'insegnamento simulato.

Il DARPA Tactical Language Training Project (Johnson *et al.*, 2005) è un buon esempio di *Serious Game*. Esso è composto da una serie di tools per l'insegnamento di lingue *tattiche*, un

sottoinsieme di conoscenze linguistiche e culturali e di capacità necessarie per eseguire missioni specifiche. Per aumentare il desiderio d'apprendere degli studenti e per dar loro modo di praticare ciò che hanno imparato, questo sistema d'insegnamento mette a disposizione una serie di missioni simulate: in un contesto virtuale, interagendo con avatars e agenti virtuali, gli studenti possono applicare, e quindi praticare, i nuovi vocaboli e gesti appresi.

In questo tipo di giochi volti all'apprendimento l'uso di agenti il più realistici possibili diventa essenziale.

In questo capitolo specifichiamo le capacità necessarie che un agente virtuale deve possedere per essere un valido ausilio nel campo dell'e-learning. Avere un aspetto umano, infatti, non è sufficiente, ciò che realmente conta dal punto di vista dell'interazione è l'abilità a ragionare con gli utenti, conversando con loro servendosi di quelle capacità comunicative ed emozionali che sono proprie degli esseri umani.

Agenti di questo tipo, quindi, sono in grado di percepire ciò che li circonda, comunicare con gli utenti e gli altri avatars presenti e di reagire ai loro discorsi e agli eventi che si manifestano nell'ambiente circostante.

La comunicazione è un'interazione fondamentalmente multimodale che implica processi complessi in cui i segnali non-verbali rivestono un ruolo estremamente importante. Tali segnali posseggono diverse funzioni: analizzare l'ambiente circostante e le persone reali e non che lo popolano, indicare il proprio stato mentale ed emotivo, mandare messaggi interpretabili agli altri e persino organizzare i propri pensieri. Nell'ambito delle nostre ricerche, siamo principalmente interessati a dotare l'agente virtuale delle prime tre funzioni, il che significa renderlo capace di:

1. generare comportamenti comunicativi multimodali,
2. avere emozioni e mostrare atteggiamenti espressivi,
3. percepire l'ambiente circostante e lo sguardo degli altri agenti e/o utenti,
4. gestire la comunicazione fornendo feedback quando necessario.

E' inoltre da tener conto che per essere veritieri, più agenti virtuali nello stesso ambiente dovrebbero avere non solo un aspetto fisico differente, ma mostrare anche comportamenti differenti, reagendo e atteggiandosi in modo unico secondo un modello emotivo e comportamentale che consenta la parametrizzazione delle sue proprietà caratteriali.

2. EMOZIONI

Le emozioni sono un elemento importante in un contesto e-learning popolato da agenti virtuali: un agente capace di mostrare emozioni riesce con maggiore probabilità a mantenere l'interesse dell'utente e a suscitare in lui una reazione emotiva, evidenziando meglio, in questo modo, le sue difficoltà o i suoi successi nell'interazione con l'insegnante sintetico.

Nonostante il valore che apporterebbe un agente emozionale¹ nel campo dell'e-learning, non sono ancora stati integrati negli agenti virtuali sofisticati modelli per la generazione delle emozioni.

Il primo scopo di un simile modello è quello di generare sul volto dell'avatar un'espressione credibile che l'utente sia in grado di riconoscere ed interpretare. In questa sezione descriviamo inizialmente le capacità che deve possedere un agente emozionale e inseguito proponiamo un metodo per realizzare tali capacità.

¹ Per agente *emozionale* intendiamo un agente capace di esprimere il proprio stato emotivo tramite il comportamento.

Prima di tutto un agente virtuale dovrebbe essere in grado di adottare un comportamento specifico per trasmettere le proprie emozioni, il che consiste nel simulare tre processi (Picard, 1997): suscitare, esprimere e provare un'emozione. Tali processi consentono all'agente di comprendere le circostanze in cui un'emozione è stata scatenata, mostrare un'emozione tramite l'espressione del volto e del corpo e manifestare che tipo di influenza essa ha sul suo modo di ragionare e agire. Un agente emozionale dovrebbe anche essere capace di interpretare l'ambiente circostante secondo le sue emozioni e quelle degli altri avatar e utenti che popolano in suo mondo (Crawford, 2003).

L'integrazione di queste abilità portano al miglioramento di diverse caratteristiche degli agenti virtuali, come:

1. credibilità: un avatar capace di esprimere emozioni crea l'illusione di essere vivo aumentando il senso di coinvolgimento nell'interazione da parte dell'utente (Bates, 1994).
2. relazione emozionale: agenti capaci di generare un coinvolgimento emotivo con gli altri personaggi virtuali e/o con gli utenti. A tal fine le caratteristiche descritte in (Stern, 2003) sono fondamentali.
3. autonomia: le emozioni possono servire nella selezione di azioni appropriate o nella memorizzazione e nel recupero di determinate informazioni (Cañamero, 2003),
4. orientamento in base alle emozioni: gli agenti emozionali scelgono le loro azioni in base alle loro emozioni e a quelle che vogliono suscitare nell'utente.

L'*Appraisal Theory* (Scherer, 2000) è una delle possibili strategie utili per rendere un agente virtuale capace di interpretare una situazione a seconda delle sue emozioni. Secondo questa teoria, le emozioni sono scatenate dall'interpretazione soggettiva di un evento che dipende da fattori culturali, dalla situazione in sé e dal carattere stesso dell'individuo. Inoltre le convinzioni e gli obiettivi di una persona sono fattori fondamentali che influenzano come e quando le emozioni vengono provate.

Nei modelli già esistenti per la generazione delle emozioni (come, per esempio, l'OCC (Ortony *et al.*, 1998)), l'interpretazione di un evento deriva dalla valutazione di un insieme di variabili dette *variabili di valutazione*; particolari combinazioni di tali variabili porta allo scatenamento di emozioni. In questi modelli spesso le variabili di valutazione sono definite a priori dal programmatore (Elliot, 1992, Reilly, 1996) o basate sugli obiettivi che l'agente si prefissa di raggiungere (Gratch, 2000). Noi, invece, vogliamo creare un modello indipendente da qualsiasi dominio e che renda l'agente capace di valutare le proprie emozioni e quelle degli altri (agenti e/o utenti) senza conoscerne a priori le intenzioni.

Nel nostro modello rappresentiamo gli eventi capaci di suscitare emozioni basandoci sull'approccio BDI (*Beliefs-Decisions-Intentions*) (Rao e Georgeff, 1991). Lo stato mentale di un agente nell'architettura BDI è definito tramite un insieme di attitudini mentali come le convinzioni (*beliefs*) e le intenzioni (*intentions*) e corrisponde alla rappresentazione del mondo ad un dato istante, includendo una descrizione degli eventi percepiti nell'ambiente circostante.

Allo stesso modo, dato che un evento capace di suscitare un'emozione è anche rappresentabile tramite attitudini mentali, esso può essere anche descritto come un particolare stato mentale. Basandoci sul modello OCC (Ortony *et al.*, 1998), abbiamo definito il valore delle variabili di valutazione attraverso convinzioni e intenzioni. Per esempio, un evento desiderabile corrisponde alla convinzione che il verificarsi di tale evento porterà l'agente a raggiungere uno dei suoi obiettivi.

Seguendo l'*Appraisal Theory*, abbiamo rappresentato gli eventi che suscitano emozioni come una combinazione di valori per le variabili di valutazione, cioè come una combinazione di attitudini mentali. Tramite la formalizzazione di tali eventi, un agente virtuale può interpretare in tempo reale le proprie emozioni che derivano da una certa situazione e può comprendere anche le emozioni degli altri, quando ne conoscendo le intenzioni. Maggiori dettagli sul modello possono essere trovate in (Ochs *et al.*, 2005).

3. COMPORTAMENTO COMUNICATIVO

In contesti di interazione sociale, lo studio dei comportamenti comunicativi che gli agenti virtuali dovrebbero mettere in atto è di fondamentale importanza per ottenere un'interazione credibile e coinvolgente. Gli esseri umani comunicano attraverso una gran varietà di *canali* di comunicazione: parole, espressioni facciali, postura, sguardo, gesti, ecc. I comportamenti non verbali accompagnano il fluire delle parole e sono sincronizzati ad esse, alle sillabe più accentate e alle pause tra le frasi (Kendon, 1974, Schefflen, 1964).

Il sollevamento delle sopracciglia di solito coincide con il momento in cui si pronunciano le parole che hanno più enfasi (Ekman, 1979); la fase di “stroke” dei gesti cioè la parte di maggior forza nell'esecuzione dei gesti, avviene inoltre sulle parole più enfatizzate o appena prima (McNeill, 1992).

I comportamenti non verbali possono avere varie funzioni comunicative (Poggi, 2003, Chovil, 1991) e servono per regolare il procedere della conversazione: l'alternarsi dei turni di conversazione, il mantenimento dell'argomento principale del dialogo, ecc. Le azioni come sorridere brevemente, sollevare le sopracciglia, stropicciare il naso spesso hanno luogo in coincidenza con degli atti verbali. Questi comportamenti non verbali possono sostituire una parola che non viene pronunciata o enfatizzare quello che si sta dicendo.

I gesti possono servire ad indicare un punto nello spazio circostante o descrivere la forma di un oggetto. Le espressioni facciali sono il canale primario per esprimere il proprio stato emotivo. Esse possono anche comunicare una particolare attitudine verso quello che si sta dicendo (per esempio ironia) o verso l'interlocutore (per esempio rivelare sottomissione).

Sia i gesti che le espressioni facciali terminano di solito quando il parlato finisce o poco dopo. A causa delle differenze tra i diversi canali comunicativi, l'implementazione di un modello comunicativo realistico per personaggi virtuali è una sfida molto difficile.

3.1 Funzioni comunicative

Nel nostro modello, seguiamo la tassonomia delle funzioni comunicative proposta in (Poggi, 2003). Una funzione comunicativa è definita come una coppia (*significato*, *segnale*). A ciascuna funzione possono essere associati differenti segnali; cioè, dato un certo significato o scopo della comunicazione, ci possono essere diversi modi di comunicarlo. Per esempio per comunicare un'*enfasi* (su una parola) si potrebbe: alzare le sopracciglia, annuire, produrre un gesto di tipo “beat”, produrre una combinazione delle tre azioni nello stesso momento. Viceversa lo stesso segnale potrebbe convogliare differenti significati, ad esempio sollevare le sopracciglia può indicare sorpresa, enfasi o l'intenzione di dare un suggerimento.

Di seguito riportiamo la nostra tassonomia delle funzioni comunicative (Poggi, 2003):

1. informazione sulle convinzioni dello speaker: comportamenti che danno informazioni sulle convinzioni di chi parla, ad esempio sul grado di certezza di quello che si sta dicendo;
2. informazione sulle intenzioni dello speaker: chi parla può dare informazioni sui suoi scopi attraverso, ad esempio, la scelta dei gesti o mettendo in risalto una certa parte della frase;
3. informazione sullo stato emotivo dello speaker: chi parla comunica il proprio stato emotivo attraverso le espressioni facciali;
4. informazione metacognitiva: chi parla può comunicare il fatto che sta cercando di ricordare un'informazione.

Nel nostro modello, il comportamento dell'agente è sincronizzato col parlato (Pelachaud e Bilvi, 2003) ed è coerente con quello che l'agente sta pronunciando. Per controllare il comportamento dell'agente utilizziamo un linguaggio di rappresentazione basato su XML chiamato 'Affective Presentation Markup Language' (APML) in cui i tag sono le funzioni comunicative elencate sopra (DeCarolis *et al.*, 2004). In APML, il testo che l'agente deve pronunciare è annotato con i tag comunicativi come nel seguente esempio:

```
<turn-allocation type="take turn"> <performative type="greet">  
Buongiorno, Angela!  
</performative> </turn-allocation> <affective type="happy">
```

Il testo viene sintetizzato automaticamente e viene riprodotto durante l'animazione finale dell'agente.



Figura 1. A sinistra: un gesto di saluto istanziato a partire dall'APML dell'esempio in sezione 3. L'APML viene convertito in un'animazione facciale e gestuale per il nostro agente. A destra: il nostro agente nell'ambiente di sviluppo Torque.

3.2 Varietà nel comportamento

Un altro aspetto da tenere in considerazione quando si animano degli agenti virtuali è quello della simulazione di un comportamento individualizzato per ciascun agente. Cioè il comportamento dovrebbe essere coerente con la personalità, l'umore, lo stato emotivo e le caratteristiche individuali che si vogliono riprodurre.

Queste caratteristiche sono molto importanti nei sistemi di apprendimento a distanza in cui degli studenti interagiscono con un insegnante artificiale attraverso degli avatar, cioè delle rappresentazioni virtuali dei partecipanti al corso. Senza queste caratteristiche i mondi virtuali sembrerebbero popolati non da “umani virtuali” ma da androidi e zombi. Insomma i personaggi sintetici dovrebbero diversificarsi non solo per il loro aspetto esteriore (area di ricerca molto popolare nei videogiochi contemporanei) ma anche nel modo in cui si comportano (Isbister, 2006).

Come detto nelle sezioni 2 e 3, il comportamento esterno di un agente virtuale è determinato dal suo stato emotivo e dagli scopi della conversazione. In situazioni reali ci sono altre due considerazioni da fare: prima di tutto, un qualsiasi tipo di comportamento, per esempio un gesto o un'espressione facciale, verrà effettuato in modo leggermente diverso da due individui anche se il loro stato interno (che è la causa del gesto o espressione) è in teoria esattamente identico; seconda cosa, dato lo stesso stato emotivo e lo stesso scopo della conversazione due individui possono scegliere di effettuare diversi tipi di segnali non verbali. Cioè non è sufficiente fissare uno stato emotivo e uno scopo per essere sicuri del comportamento finale di una persona. Ci sono altri fattori, relativi all'individualità, che determineranno il comportamento. Considereremo più in dettaglio due aspetti del nostro modello per agenti individualizzati: il nostro modello per l'espressività e quello per il comportamento multimodale.

Modello per l'espressività Vari ricercatori (vedi, ad esempio, Wallbott e Scherer (Wallbott e Scherer, 1986), Pollick (Pollick, 2004)) hanno svolto studi sulle caratteristiche motorie del comportamento umano e le hanno spesso codificate in categorie come lento/veloce, contratto/espanso, debole/energico, sgradevole/piacevole. Definiamo l'*espressività* del comportamento come la “qualità” dell'esecuzione motoria di un gesto, un'espressione o più in generale di un qualsiasi comportamento fisico.

Partendo dai risultati riportati in (Wallbott e Scherer, 1986) abbiamo definito e implementato l'espressività (Hartmann *et al.*, 2005) come un insieme di parametri che influenzano le seguenti qualità di esecuzione di un comportamento non verbale: velocità di braccia/testa, volume dello spazio utilizzato dalle braccia/testa, energia e fluidità delle braccia/testa, numero di ripetizioni dello stesso comportamento.

Modello per la generazione multimodale del comportamento Un agente virtuale dovrebbe essere capace di interagire con l'utente attraverso tutte le modalità coinvolte di solito nella comunicazione tra umani: parlato, gesti, sguardo, espressioni facciali, movimenti del corpo, postura (Pelachaud *et al.*, 2002, Ruttkay e Pelachaud, 2002).

Nel nostro modello abbiamo implementato un meccanismo per scegliere come e cosa comunicare su queste modalità durante la conversazione. Il nostro modello si basa su una gerarchia tra le modalità in cui ad ogni modalità è associato un valore che rappresenta il suo livello di preferenza. Per esempio un agente che tende ad usare i gesti durante la comunicazione avrà un valore più alto sulla modalità gestuale. I valori di questa gerarchia possono essere determinati da studi sulle diverse culture (gli italiani ad esempio comunicano

in genere utilizzando molto i gesti, ecc), fattori contestuali (in un ambiente rumoroso si tende ad utilizzare un maggior numero di gesti mentre si parla) o comportamenti idiosincratici (una persona può avere una tendenza innata ad utilizzare spesso i movimenti della testa mentre parla). Durante la comunicazione il nostro agente sceglierà diversi segnali (gesti, espressioni facciali, ecc) a seconda della gerarchia delle modalità, il significato di quello che l'agente dice e gli scopi della conversazione: ad esempio gli stati emotivi vengono manifestati principalmente con le espressioni facciali mentre le descrizioni degli oggetti vengono di solito accompagnate da molti gesti.

4. PERCEZIONE E ATTENZIONE

L'abilità di percepire l'ambiente circostante attraverso i sensi in modo congruente con gli esseri umani è un requisito importante per agenti virtuali che devono mostrare un comportamento plausibile. Un agente che percepisce il mondo in modo diverso da una persona reale è destinato a comportarsi in maniera differente, indipendentemente da quanto sofisticate possano essere le sue capacità cognitive.

Per questo cerchiamo di rendere i sensi dell'agente più simili possibili a quelli dell'uomo, implementando per lui una visione sintetica in tempo reale, una memoria e l'abilità di spostare la sua attenzione (Peters, 2004). Queste tre caratteristiche collaborano tra di loro per raccogliere informazioni sull'ambiente circostante a cui i sensi dell'agente devono adattarsi.

Il modulo per la visione sintetica è monoculare a multipla risoluzione; opera su immagini statiche (aggiornate frequentemente) della parte del mondo visibile dall'agente.

Ogni volta che l'agente percepisce il mondo circostante, l'immagine catturata viene ridisegnata usando dei falsi colori. In questo tipo di visualizzazione gli oggetti presenti nella scena vengono sostituiti da sagome con un colore preassegnato che vengono confrontate con quelle contenute in un database. Sebbene questo tipo di approccio sia abbastanza dispendioso in termini di tempo di calcolo rispetto a metodi geometrici (Funge, 2004) esso è di fondamentale importanza quando si desiderino avere capacità di attenzione sia a livello di oggetti che di spazio; questo approccio dipende intrinsecamente dalle condizioni di luce e da eventuali effetti sulla qualità di disegno della scena.

Lo scopo del modulo di attenzione visuale è quello di scegliere un insieme di oggetti o punti visibili nella scena per poi permettere all'agente di effettuare un'elaborazione avanzata ed è basato su un modello realtime molto noto in neurologia computazionale (Itti, 2002)

Noi lo usiamo con delle scene virtuali passando l'intera scena percepita dal modulo di visione sintetica in modo da ottenere una "saliency map": una classificazione di tutto ciò che compare nella scena fatta in due dimensioni in toni di grigio che ci informa su quali aree della scena potrebbero attrarre l'attenzione visuale. Un sistema di memoria immagazzina il livello di incertezza per ogni oggetto; questi valori vengono combinati in una mappa d'incertezza con la quale la saliency map viene combinata per fornire infine la mappa d'attenzione.

Le regioni d'interesse vengono selezionate dalla mappa d'attenzione e vengono usate per simulare il tracciato secondo cui l'agente percorre visivamente la scena che vede.

In un sistema di apprendimento a distanza, queste capacità permettono di avere un comportamento dello sguardo dell'agente più realistico.

Questo tipo di sistema sono la base per l'implementazione di meccanismi cognitivi più ad alto livello: per esempio abbiamo usato queste capacità insieme alla "Theory of mind" per controllare il momento di inizio dell'interazione tra agenti virtuali (vedi sezione 5.1).

4.1 Theory of Mind

In parallelo con la percezione, attenzione e memoria per gli agenti virtuali stiamo portando avanti un progetto per incorporare nel nostro agente un modello percettivo basato sulla psicologia evolutiva e supportato da risultati neuropsicologici (Peters, 2005).

In una comunità virtuale è di vitale importanza che gli avatar e gli agenti agiscano in modo socialmente credibile, e che quindi mostrino attenzione verso gli elementi importanti del mondo virtuale. Nel nostro modello un meccanismo di percezione dell'intenzionalità (ID) filtra gli altri agenti della scena e ne segmenta le immagini in varie parti: occhi, testa, corpo. Un modulo di percezione della direzione dell'attenzione (DAD) calcola la direzione di occhi, testa e corpo degli altri agenti rispetto alla propria posizione. Queste direzioni permettono di determinare un parametro chiamato *attention level* (livello d'attenzione) che viene enfatizzato in condizioni ad esempio di mutua attenzione, segnalate da un altro modulo chiamato MAM. Il risultato di questa elaborazione viene immagazzinato in una memoria a breve termine.

Un agente può integrare un modello del genere per determinare se gli altri agenti lo stanno percependo (visivamente) nell'ambiente e qual'è il livello di interesse che questi hanno verso di lui. Il sistema è utile per generare un comportamento dello sguardo più realistico basandosi sulla percezione del comportamento degli altri.

In futuro intendiamo usare questo algoritmo per poter percepire l'attenzione dell'utente in un ambiente reale (Bevacqua *et al.*, 2006).

5 GESTIONE DELLA COMUNICAZIONE

La comunicazione è per natura bi-direzionale e la barriera tra locutore e ascoltatore può essere confusa e indefinita: entrambi emettono segnali e si adattano costantemente l'uno all'altro. La comunicazione implica un punto d'inizio e uno di fine, eventi che sono strettamente legati a comportamenti convenzionali, tra cui lo sguardo, il movimento del capo e i gesti che rivestono un ruolo primario: attraverso lo sguardo potremmo mostrare il desiderio di voler cominciare una conversazione, smettendo di gesticolare potremmo segnalare di aver finito di parlare e di voler cedere il turno all'altro partecipante al dialogo, mentre potremmo mostrare il desiderio di prendere la parola o d'intraprendere un'interazione attraverso una serie di gesti.

Una valida organizzazione olistica della comunicazione deve prevedere la capacità di gestire l'inizio e il mantenimento di una comunicazione tramite la valutazione e l'interpretazione dei segnali di feedback emessi dall'altro interlocutore.

5.1 Inizio di un'interazione

Un aspetto importante e raramente studiato dell'interazione tra esseri umani e/o agenti virtuali è il modo in cui l'interazione stessa comincia. Tutti i comportamenti legati all'instaurazione di un'interazione (come l'orientamento dello sguardo, il saluto con una mano, chiamare il nome della persona) sono estremamente significativi per aggiungere realismo nelle comunità virtuali dove i comportamenti realistici sono fortemente desiderabili (Isbister, 2006).

Nella realtà, le persone devono di tanto in tanto impiegare i loro sensi per verificare se c'è la possibilità di instaurare di un contatto con le persone circostanti, devono conquistare l'attenzione di chi vogliono coinvolgere nell'interazione, devono segnalare il loro stesso interesse sia tramite segnali verbali che non verbali e devono cercare la cooperazione dell'altra persona per portare avanti la conversazione.

A volte, dal nostro punto di vista per lo studio di comportamenti plausibili, non è importante la conversazione in sé, quanto le azioni necessarie a cominciare una conversazione. Per questa fase di un'interazione abbiamo implementato un sistema automatico non verbale basato principalmente sulla direzione dell'attenzione e sulla *Theory of Mind* (vedere sezione 4) con cui guidiamo una macchina gerarchica a stati finiti attraverso vari fasi di un'interazione: come, per esempio, monitorare l'ambiente circostante, attrarre l'attenzione dell'altro e valutarne la sua reazione. Quindi, a differenza di altri sistemi, il comportamento dell'agente è basato non solo sui suoi obiettivi, ma anche su ciò che suppone siano le intenzioni dell'altro, per evitare l'imbarazzante situazione di cominciare un'iterazione con chi non ne alcuna intenzione (Goffman, 1963). Le nostre ricerche sulla definizione dei comportamenti necessari per l'instaurazione di una conversazione, suggeriscono che gli utenti possono essere tanto sensibili ai sottili segnali non verbali nella realtà quanto in un ambiente virtuale (Peters, 2006).

5.2 Feedback

Durante una conversazione due flussi di informazioni si instaurano tra la persona che parla (*locutore*) e quella che ascolta (*interlocutore*), il primo riguarda il contenuto effettivo del discorso del locutore e il secondo la reazione dell'interlocutore a tale discorso. Dalla persona che ascolta, durante una conversazione, ci si aspetta che fornisca una serie di segnali per informare il locutore del successo dell'interazione (Allwood *et al.* 1993, Poggi, 2005). Questi segnali prendono il nome di *feedback*; senza di loro una comunicazione diventa difficile e frustrante poiché la persona che parla non ha modo di capire se l'ascoltatore sta seguendo il discorso e che tipo di effetto le sue parole stanno avendo.

Quindi il comportamento dell'interlocutore è un aspetto fondamentale che dobbiamo prendere in considerazione se vogliamo implementare agenti credibili che possano essere usati nel campo dell'e-learning; un personaggio sintetico credibile non rimane immobile una volta che smette di parlare e cede il turno all'altro partecipante alla conversazione, ma continua a muoversi emettendo segnali per far vedere che è ancora parte attiva nell'interazione.

I segnali di feedback dipendono fortemente dal contesto, l'interlocutore, infatti, emette principalmente segnali a seconda di ciò che il locutore dice, per mostrare cosa pensa del contesto del discorso: se concorda o non concorda, se crede o non crede e così via; oppure per provocare un effetto particolare nella persona che parla, per esempio può fissarlo con gli occhi spalancati per mostrare stupore e incredulità chiedendo indirettamente una conferma. Ci riferiamo a questo tipo di feedback come *feedback cognitivo*.

D'altro canto l'ascoltatore può anche emettere segnali inconsciamente, senza riflettere (Allwood *et al.*, 1993); infatti, durante una conversazione, deve prendere tante piccole decisioni comportamentali in un tempo talmente breve che spesso agisce senza neanche rendersene conto. L'interlocutore risponde d'istinto al comportamento del locutore, per esempio spostando lo sguardo nella stessa direzione di quello della persona che parla o sorridendo quando il locutore sorride. Questo tipo di feedback è detto *feedback reattivo*.

Di conseguenza un solo modello per la generazione del feedback in un agente virtuale, non è sufficiente e sono necessari due modelli computazionali, rispettivamente un modello cognitivo e uno reattivo. Il primo, comunque, è piuttosto difficile da implementare, e soprattutto da rendere *real-time*: per elaborare dei segnali sensati di feedback non solo è necessario estrapolare il valore semantico del discorso (cosa ancora complessa) ma bisogna avere anche informazioni sulla personalità dell'interlocutore.

Il secondo modello, invece, è di più facile realizzazione dato che spesso i feedback istintivi dell'ascoltatore sono suscitati dal comportamento del locutore. Un insieme di regole per generare automaticamente questo tipo di segnali è stato proposto in (Maatman, 2005, Friedman e Gillies, 2005). Per esempio, a partire da un corpus di dati reali, Maatman ha osservato che i segnali *continuers* (come annuire col capo o brevi risposte verbali atte ad incitare il locutore a proseguire) appaiono solitamente quando la voce della persona che parla varia di tono; aggrottare le sopracciglia e spostare il corpo e lo sguardo spesso sono la conseguenza dell'incertezza mostrata dal locutore; infine, le espressioni facciali si allineano a quelle di colui che parla per empatia o per evidenziare che l'interlocutore sta seguendo e capendo il discorso. Quest'ultimo atteggiamento dell'ascoltatore è detto *mimicry* (mimica).

5.3 Mantenimento dell'interazione

Come abbiamo già menzionato, la gestione di una comunicazione implica che l'agente locutore si comporti in un modo specifico per instaurare una conversazione e per valutare continuamente il comportamento dell'interlocutore in modo da comprenderne la reazione alle sue parole. Un altro aspetto dell'interazione da non dimenticare consiste nella gestualità del locutore. L'uso dei gesti, infatti, apporta una quantità maggiore di informazioni durante una comunicazione.

I gesti sono la rappresentazione delle intenzioni del locutore (Calbris, 2003) e non solo lo aiutano a formulare i suoi pensieri ma mostrano anche i suoi sforzi comunicativi. Per implementare agenti capaci di mantenere un ascoltatore attivo ed interessato, è importante non dimenticare questa dimensione pragmatica dei gesti. In alcune precedenti ricerche condotte da (Gullberg e Holmqvist, 1999, Barrier *et al.*, 2005) sono stati raccolti dati oggettivi sui movimenti fatti da un locutore; questi studi hanno evidenziato quali gesti attirano l'attenzione: per esempio, quando un locutore usa un gesto per descrivere o per puntare un oggetto, l'interlocutore tende a spostare lo sguardo verso le mani del locutore o verso l'oggetto in questione. Quindi, i gesti sono parte del processo di comunicazione.

Nelle nostre ricerche, abbiamo analizzato una serie di cartoni animati contenenti interazioni conversazionali. Da questo corpus di dati abbiamo ricavato informazioni sull'espressività dei movimenti dei personaggi animati, osservando che i parametri dell'espressività dei gesti (fluidità, forza, espansione spaziale e ripetitività) non sono solo l'espressione della personalità del personaggio o del suo stato emotivo ma che riflettono anche le sue intenzioni pragmatiche. Se vogliamo creare insegnanti virtuali capaci di mantenere l'interesse dell'utente in un'interazione conversazionale, dovremmo tenere presente il valore pragmatico dei gesti e considerare come siano in grado di esaltare delle parti salienti di un discorso.

6 CONCLUSIONI

In questo capitolo abbiamo presentato una serie di capacità che un agente virtuale e *real-time* dovrebbe avere per migliorare le sue prestazioni nell'ambito dell'e-learning. Una sfida interessante è quella di estendere il nostro modello al di là degli scenari che consentono solo interazioni tra due parti, per simularne altri popolati da un maggior numero di agenti e utenti. Tale sfida non è difficile solo a causa della conseguente complessità computazionale che ne deriva, ma anche a causa del fatto che la stragrande maggioranza delle teorie esistenti per l'interazione tra agenti sono sviluppate su piccoli gruppi, il che complica notevolmente la formulazione di modelli appropriati.

Stiamo, inoltre, lavorando per integrare le abilità sopraelencate in un sistema per agenti real-time in un ambiente virtuale.

Usando il sistema *Torque* (<http://www.garagegames.com>) abbiamo già integrato, in un agente sintetico, capacità di percezione, attenzione e interazioni, come descritto nelle sezioni 4 e 5.1 (vedere Figura 1).

REFERENZE

Allwood J., Nivre J., Ahlsn E. (1993). On the semantics and pragmatics of linguistic feedback. In *Semantics*, 9(1).

Barrier G., Caelen J., Meillon B. (2005). La visibilité des gestes: Paramètres directionnels, intentionnalité du signe et attribution de pertinence. *Atti del convegno "Workshop Français sur les Agents Conversationnels Anims"*. (Grenoble, France). 113–123.

Bates J. (1994). The role of emotion in believable agents. In *Communications of the ACM*, 37(7):122–125.

Bevacqua E., Raouzaïou A., Peters C., Caridakis G., Karpouzis K., Pelachaud C., Mancini M. (2006). Multimodal sensing, interpretation and copying of movements by a virtual agent. *Atti del convegno "Perception and Interactive Technologies"*. (Kloster Irsee, Germany, June 19-21).

Calbris G. (2003). L'expression gestuelle de la pense d'un homme politique. In du CNRS (Ed.), *Collection Communication*.

Cañamero, L. D. (2003). Designing emotions for activity selection in autonomous agents. In R. Trappl R., Petta P., Payr S. (Ed.), *Emotions in Humans and Artifacts*. Cambridge, MA: The MIT Press. 115-148.

Chovil N. (1991). Social determinants of facial displays. In *Journal of Nonverbal Behavior*, 15(3):141-154.

Crawford C. (2003). *On game design*. New Riders Games Press.

DeCarolis B., Pelachaud C., Poggi I., Steedman M. (2004). Apml, a mark-up language for believable behavior generation. In Prendinger H., Ishizuka M. (Ed.), *Life-like Characters. Tools, Affective Functions and Applications*. Springer Press.

Ekman P. (1979). About brows: Emotional and conversational signals. In von Cranach M., Foppa K., Lepenies W., Ploog D. (Ed.), *Human ethology: Claims and limits of a new discipline: contributions to the Colloquium*. Cambridge, England: Cambridge University Press.

Elliot C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system*. PhD Thesis, Northwestern University Evanston, Illinois.

Friedman D., Gillies M. (2005). Teaching virtual characters to use body language. *Atti del convegno "Intelligent Virtual Agent, Lecture Notes in Artificial Intelligence"*. Springer-Verlag Press.

Funge J. (2004). *Artificial intelligence for computer games: An introduction*. Wellesley, MA: A K Peters.

Goffman E. (1963). *Behaviour in public places: Notes on the social order of gatherings*. New York: The Free Press.

Gratch J. (2000). Emile: Marshalling passions in training and education. *Atti del convegno "The Fourth International Conference on Autonomous Agents"*. (Barcelona, Catalonia, Spain) 325–332.

Gullberg M., Holmqvist K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. In *Journal of Pragmatics and Cognition*, 7:35–63.

Hartmann B., Mancini M., Pelachaud C. (2005). Towards affective agent action: Modelling expressive ECA gestures. *Atti del convegno "IUI Workshop on Affective Interaction"* (San Diego, CA, January).

Johnson W. L., Shaw E., Ganeshan R. (1998). Pedagogical agents on the web. *Atti del convegno "Workshop on WWW-based Tutoring, ITS"* (San Antonio, Texas).

Johnson W.L., Vilhjálmsson H. H., Marsella S. (2005). *Serious games for language learning: How much game, how much AI?*. *Atti del convegno "Artificial Intelligence in Education Conference"* (Amsterdam, July 18-22).

Isbister K. (2006). *Better game characters by design: A psychological approach*. Elsevier Science and Technology Books.

Itti L. (2002). *Models of bottom-up and top-down visual attention*. PhD thesis, California Institute of Technology.

Kendon A. (1974). Movement coordination in social interaction: Some examples described. In Weitz S. (Ed.), *Nonverbal Communication*. Oxford: Oxford University Press.

Maatman R. M., Gratch J., Marsella S. (2005). Natural behavior of a listening agent. *Atti del convegno "5th International Conference on Interactive Virtual Agents"*. (Kos, Greece).

McNeill D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Ochs M., Niewiadomski R., Pelachaud C., Sadek D. (2005). Intelligent expressions of emotions. In Tao J., Tan T., Picard R. W. (ED.), *The 1st International Conference on Affective Computing and Intelligent Interaction*. Beijing, China: Springer Press. 707-714

Ortony A., Clore G. L., Collins A. (1988). *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press.

Pelachaud C., Bilvi M. (2003). Computational model of believable conversational agents. In Huget M.P. (Ed.), *Communication in Multiagent Systems*, volume 2650 of *Lecture Notes in Computer Science*. Springer-Verlag Press. 300–317.

Pelachaud C., Carofiglio V., De Carolis B., de Rosis F. (2002). Embodied contextual agent in information delivering application. *Atti del convegno “First International Joint Conference on Autonomous Agents and Multi-Agent Systems”* (Bologna, Italy, July).

Peters C. (2006). Evaluating perception of interaction initiation in virtual environments using humanoid agents. *Atti del convegno “17th European Conference on Artificial Intelligence”* (Riva Del Garda, Italy, August). 46-50.

Peters C. (2005). Direction of attention perception for conversation initiation in virtual environments. *Atti del convegno “International working conference on Intelligent Virtual Agents”* (Kos, Greece, September). 215-228.

Peters C. (2004). *Bottom-up visual attention for autonomous virtual human animation*. PhD thesis, Department of Computer Science, Trinity College Dublin.

Picard R. W. (1997). *Affective Computing*. MIT Press.

Poggi I. (2005). Backchannel: From humans to embodied agents. *Atti del convegno “Artificial Intelligence and the Simulation of Behaviour”* (Hatfield, England, April 12-15).

Poggi I. (2003). Mind markers. In Trigo N., Rector M., Poggi I. (Eds.), *Gestures. Meaning and use*. Oporto, Portugal: University Fernando Pessoa Press.

Pollick F. E. (2004). The features people use to recognize human movement style. In Camurri A., Volpe G. (Eds.), *Gesture-based communication in human-computer interaction*, volume 2915:10-19. Springer Press.

Rao S., Georgeff M. P. (1991). Modeling rational agents within a BDI-architecture. *Atti del convegno “International Conference on Principles of Knowledge Representation and Reasoning”* (San Mateo, CA, USA). 473-484.

Rasseneur D., Delozanne E., Jacoboni P., Grugeon B.(2002). Learning with virtual agents: Competition and cooperation in AMICO. In Cerri S., Gouardères G., Paraguaçu F. (Eds.), *Proceeding of ITS*. Biarritz, France: Springer-Verlag. 61-70.

Reilly S. (1996). *Believable Social and Emotional Agents*. PhD Thesis, Computer science. University of Carnegie Mellon.

Ruttkay Z., Pelachaud C. (2002). Exercises of style for virtual humans. In *Symposium of the AISB’02 Convention*, volume Animating Expressive Characters for Social Interactions, London, 2002.

Schefflen A. E. (1964). The significance of posture in communication systems. In *Journal of Psychiatry*, 27.

Scherer K. R. (2000). Emotion. In Hewstone M., Stroebe W. (Eds.). *Introduction to social psychology: A European perspective* (3rd ed.). Oxford: Blackwell. 151-191

Stern A. (2003). Creating emotional relationships with virtual characters. In R. Trapp R., Petta P., Payr S. (Eds.), *Emotions in Humans and Artifacts*. Cambridge, MA: The MIT Press.

Wallbott H. G., Scherer K. R. (1986). Cues and channels in emotion. In *Journal of Personality and Social Psychology*, 51(4):690–699.