

## *Capitolo sesto*

# **L'espressività nella comunicazione non-verbale degli agenti conversazionali**

MAURIZIO MANCINI, BJÖRN HARTMANN, CATHERINE PELACHAUD

### **1. Introduzione**

Gli agenti conversazionali (o ECA, dall'inglese *Embodied Conversational Agents*) sono un tipo di interfaccia utente, atta a riprodurre la completezza della comunicazione uomo-uomo nell'interazione uomo-macchina. Gli ECA sono delle rappresentazioni grafiche realistiche di esseri umani e possono comunicare con l'utente (o altri ECA) attraverso numerose modalità: il parlato, le espressioni facciali, lo sguardo, la gestualità, la postura del corpo.

Per migliorare la credibilità e la verosimiglianza degli Agenti Conversazionali, stiamo spostando la nostra attenzione, partendo da un agente generico, verso un modello di agente che permetta al contrario di simulare un comportamento più "individualizzato". Questo tipo di caratterizzazione di un agente dipenderà non solo da fattori "personali" (come il livello culturale, il ruolo professionale e sociale, la personalità, le esperienze personali, ecc) ma anche da un determinato insieme di fattori dipendenti dal contesto (Poggi 2003; 2006) (come l'identità dell'interlocutore, il tipo di relazione sociale tra chi parla e chi ascolta, ecc.) o variabili durante la conversazione (come l'obiettivo stesso della conversazione, il coinvolgimento emotivo, ecc.). Chiameremo tutti questi fattori "influenze comportamentali" nel senso che agiscono in qualche modo sul comportamento dell'agente. Lo stesso tipo di informazione finale potrà quindi essere comunicata in modi completamente differenti dipendenti dal tipo di influenze comportamentali in atto in un certo momento. Inoltre queste influenze agiranno su livelli separati: contribuiranno nella scelta di una o più modalità comunicative tra quelle disponibili (faccia, gesti, sguardo,

postura, ecc); determineranno il tipo di comportamento non-verbale da attivare su ciascuna modalità (distogliere lo sguardo, sorridere, cambiare postura, ecc); modificheranno le caratteristiche fisiche e temporali dei vari comportamenti non-verbali. Chiameremo questa caratterizzazione del comportamento non-verbale "espressività".

In questo articolo prima di tutto forniremo una descrizione generale del concetto di espressività nell'interazione tra gli esseri umani e spiegheremo come abbiamo deciso di modellare queste caratteristiche. Poi presenteremo una tassonomia delle influenze comportamentali e successivamente spiegheremo il funzionamento del nostro sistema, a partire dal calcolo dei parametri per l'espressività dell'agente conversazionale Greta fino ad arrivare alla generazione dell'animazione finale.

## 2. L'espressività degli esseri umani

Prima di essere studiato dai ricercatori sugli ECA, il problema di caratterizzare le espressioni facciali e i movimenti del corpo è stato preso in esame principalmente da scienze come la psicologia sociale e dalla biomeccanica. Cercheremo ora di riassumerne velocemente i risultati che ci riguardano più da vicino in modo da poter poi giustificare l'introduzione della nostra definizione di espressività per gli agenti conversazionali.

### 2.1 L'espressività delle espressioni facciali

Lo studio delle espressioni facciali è una vasta e complessa area di ricerca. Una parte di essa riguarda i metodi di identificazione automatica dell'utente a partire dall'immagine facciale (Young, Bruce 1991), il riconoscimento delle espressioni facciali, la lettura del movimento labiale (Massaro, Cohen 1990; Summerfield 1992), il riconoscimento dello stato emotivo a partire dall'espressione facciale (Ekman 1982). Altri studi riguardano argomenti come l'estrazione dell'immagine della faccia da uno sfondo casuale, la comprensione del processo di riconoscimento di facce familiari (Young, Bruce 1991), il modo in cui le emozioni sono percepite – quali sono i ruoli dei "segnali conversazionali" e dei "*backchannels*" (Poggi 2007).

Concentreremo la nostra attenzione in particolare sullo studio delle espressioni facciali e della percezione delle emozioni attraverso di esse. Le ricerche condotte in questo campo partono spesso da alcune prove

empiriche effettuate sempre nelle stesse condizioni: ad un gruppo di persone viene mostrato un insieme di fotografie di facce che esprimono certe emozioni riprese precedentemente usando degli attori (recentemente sono state utilizzate anche fotografie di espressioni "spontanee"). Dai risultati si può constatare che i soggetti riconoscono più facilmente le espressioni che non coinvolgono gli stessi tipi di movimenti facciali. Ekman e Friesen (1982) hanno anche identificato le principali differenze tra le espressioni di emozioni realmente provate e quelle recitate. Quelle false di solito sono più asimmetriche e le loro caratteristiche temporali sono diverse dalle corrispondenti espressioni reali. Inoltre è stato verificato che il riconoscimento delle espressioni facciali può essere effettuato a partire da un'astrazione della faccia che, trascurando la morfologia particolare della persona presa in esame, ne riassume le sole caratteristiche facciali generali (potremmo dire che le espressioni facciali possono essere considerate "*model-independent*").

Molti ricercatori suggeriscono, invece di cercare di dedurre direttamente l'espressione facciale che viene mostrata, di analizzare lo stato delle varie componenti della faccia (sopracciglia, occhi, angoli della bocca, ecc) in modo da ottenere l'espressione globale della faccia dalla loro sovrapposizione (Cohn *et al.* 1998; Donato *et al.* forth). Questo metodo è basato sul *Facial Action Coding System* (FACS: Ekman, Friesen 1978), che permette di misurare i segnali facciali usando delle unità elementari minimali dette *Aus* (*Action Units*).

## 2.2 L'espressività del corpo

Il nostro studio dell'espressività del corpo parte da considerazioni di carattere percettivo – come l'espressività è percepita dagli altri; e non quale tipo di attivazione dei muscoli ha provocato la produzione di certi segnali. I ricercatori nel campo della psicologia sociale (Wallbot, Scherer 1986; Gallaher 1992; Wallbot 1998) hanno analizzato la percezione dei movimenti del corpo, principalmente attraverso degli strumenti di misura ad hoc, realizzati riducendo progressivamente un'estesa lista di possibili comportamenti tra i quali gli annotatori potevano effettuare una scelta.

Wallbot e Scherer (1986) hanno codificato le loro impressioni sui principali tipi di movimenti del corpo in cinque categorie: lento/veloce, ristretto/espansivo, debole/energico, poco attivo/molto attivo, piacevole/spiacevole. In un successivo studio, Wallbot (1998) ha trovato che tranne nel caso di configurazioni corporee statiche, tre dimensio-

ni dinamiche possono essere costantemente identificate da chi osserva qualcuno in movimento: quantità di attività, espansività/estensione spaziale, dinamica/energia/potenza del movimento. Gallaher (1992) ha trovato quattro possibili dimensioni di variabilità dipendenti dallo stile proprio di ogni persona: espressività (quantità di “energia” nella comunicazione); animatezza (quantità di “energia” in movimenti non direttamente connessi alla comunicazione); espansività (utilizzo dello spazio, posizione nello spazio dei gomiti); coordinazione (morbidezza, fluidità del movimento). Ball e Breese (2000) fanno riferimento ai risultati di Collier sulla correlazione tra le caratteristiche temporali e spaziali di gesti/postura e personalità/emozione: la frequenza e la velocità dei movimenti sono collegate alla comparsa dell’emozione, così come lo è il livello di “apertura” dei movimenti corporei. La ricerca svolta da Pollick (2004) fa emergere i limiti del decomporre i movimenti in componenti più piccole e assegnare loro dei valori discreti: il modo in cui questi dipendono da fattori spaziali e temporali non è banale e cambia da un tipo di movimento all’altro. L’importanza di considerare la postura come un ulteriore canale di comunicazione è ormai chiara (Mehrabian 1969; Harrigan 1985) ma non integreremo per ora questo ramo della ricerca nel nostro lavoro.

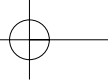
### ***2.3 Attributi per l’espressività***

In base alle osservazioni che abbiamo presentato nei paragrafi precedenti proporremo ora l’insieme di attributi che utilizzeremo nel nostro modello dell’espressività per gli ECA. Questi sei parametri, che descriveremo ora in maniera qualitativa, faranno parte dell’insieme di caratteristiche di ogni agente “individualizzato” e serviranno ad indicare le sue *tendenze* comportamentali generali.

*Overall activation*: rappresenta la quantità di movimento (attività motoria) distribuita sulle varie modalità di comunicazione durante il turno di conversazione dell’agente (influenza per esempio l’uso simultaneo di espressioni facciali e gesti). Servirà in generale a caratterizzare il comportamento dell’agente tra i due estremi: passivo/statico e animato/ dinamico.

*Spatial extent*: ampiezza dei movimenti fisici dell’agente, per esempio le dimensioni dello spazio ricoperto dalle braccia durante il loro tragitto o la quantità di spostamento delle sopracciglia.

*Temporal*: durata dei movimenti, ad esempio veloci/corti oppure lenti/lunghi.



*Fluidity*: morbidezza e continuità dei movimenti, ad esempio quando vengono effettuati due gesti consecutivi (coarticolazione) il movimento potrebbe essere del tutto fluido oppure presentare una spezzatura al termine del primo gesto e prima di incominciare il secondo.

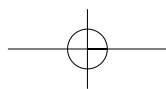
*Power/Energy*: quantità di tensione muscolare nei movimenti, variabile tra debole/rilassato e forte/teso.

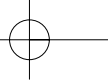
*Repetitivity*: tendenza alla ripetizione ritmica di specifici movimenti su specifiche modalità (ad esempio se consideriamo il momento della testa, un'elevata ripetitività potrebbe tradursi nella tendenza ad annuire molte volte di seguito).

A livello d'implementazione, *overall activation* è un numero reale variabile tra 0 e 1. Ciascuno degli altri attributi invece è un numero reale definito nell'intervallo  $[-1; 1]$ , dove zero corrisponde al comportamento di un agente senza l'aggiunta dei parametri per l'espressività. *Overall activation*, *Fluidity* e *Power* agiscono sull'intero turno di conversazione dell'agente mentre gli altri parametri possono essere ridefiniti in modo da influenzare singoli atti comunicativi. Il concetto di ripetitività è stato aggiunto perché compare in una annotazione di gesti effettuata da Martell *et al.* (2003) su una raccolta di video. Wachsmuth (1999) aveva precedentemente identificato l'importanza del ritmo nei comportamenti non-verbali ma lo aveva utilizzato solo per il riconoscimento dei gesti e non per la sintesi.

### 3. Tassonomia delle influenze

La maggior parte degli ECA creati fino ad oggi sono molto generici nel loro comportamento. Diversi studi hanno mostrato l'importanza di tenere in considerazione informazioni più complesse come i fattori culturali (Lee, Nass 1991), la personalità (Isbister, Nass, 2000; de Rosis *et al.* 2003), l'ambiente circostante (McBreen *et al.* 2004) durante la realizzazione di un agente. Tutti questi fattori influiscono nell'interazione che l'utente può avere con l'agente. La capacità di adattare il proprio comportamento alla tipologia di utente che ha davanti può migliorare notevolmente la credibilità dell'agente e la sua capacità di persuasione. Diversi studi hanno mostrato che le modalità in cui le persone comunicano informazioni tra loro, o descrivono eventi, o parlano dei propri sentimenti, dipendono da aspetti culturali (Brislin 1993; Lee, Nass 1998), ma anche dalla per-





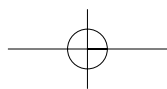
sonalità, il genere (maschile, femminile), l'età (bambino, adolescente, adulto), il ruolo sociale (genitore, insegnante, dottore, ecc.), l'esperienza passata, il livello culturale e l'ambiente circostante. Quello che stiamo cercando di puntualizzare è che, dato un certo obiettivo della conversazione, che è ben chiaro nella mente di chi sta parlando (ad esempio, un paziente che sta spiegando al proprio medico i sintomi di un'influenza), le persone si differenzieranno nel loro modo di esprimersi, sia nella formulazione dei propri pensieri che nell'insieme dei comportamenti verbali e non-verbali messi in atto per comunicarli all'interlocutore (Brown, Nichol-English 1999). Riportando questo ragionamento agli ECA, potremmo dire che tutti i fattori che contribuiscono alla creazione di un agente individualizzato agiscono su differenti livelli. Influenzano la generazione della struttura del dialogo, la costruzione delle frasi, la selezione dei comportamenti non-verbali e della loro espressività (de Rosis *et al.* 2003).

### ***3.1 Influenze intrinseche***

Per prima cosa va considerato il fatto che ogni essere umano ha le proprie abitudini, consce o inconsce, che intervengono nel contenuto dei suoi discorsi e che definiscono le sue attitudini e comportamenti durante la conversazione. Queste abitudini, che chiameremo influenze intrinseche, derivano soprattutto dalla sua personalità, età, sesso, nazionalità, cultura, educazione ed esperienza (de Rosis *et al.* 2003). Alcuni comportamenti non-verbali sono molto dipendenti dalla cultura personale. È il caso dei gesti emblematici, che possono essere tradotti direttamente in parole. Oppure potrebbe essere il caso di alcuni gesti iconici o metaforici, che possono avere radici culturali (l'azione di mangiare non viene rappresentata nello stesso modo se per tradizione si utilizzano delle bacchette invece della forchetta). D'altra parte, non tutti i gesti sono culturalmente dipendenti. Cassell (2000) ha verificato che i tipi di gesti che si fanno durante una conversazione sono molto meno diversi di quanto ci si sarebbe potuto aspettare. La differenza principale è più nella quantità dei gesti prodotti.

### ***3.2 Influenze contestuali***

Molti fattori dipendenti dal contesto possono aumentare o diminuire gli effetti delle influenze intrinseche. Le influenze contestuali possono essere determinate da fattori ambientali come l'illuminazione, il

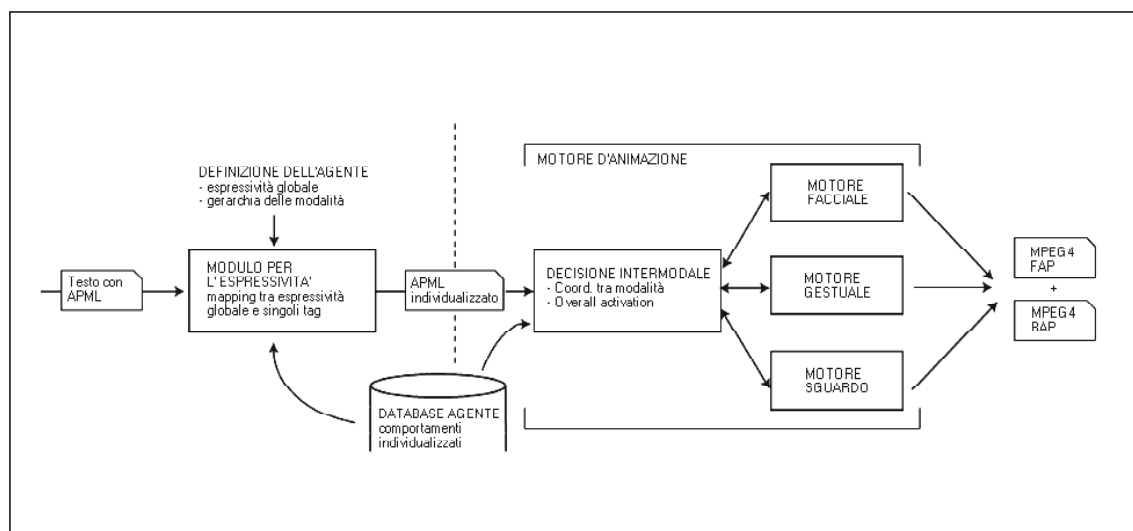


rumore di fondo, la quantità di spazio a disposizione, le caratteristiche proprie del sito in cui si svolge la conversazione. Per esempio in una stanza affollata molti gesti sono praticamente impossibili da effettuare e spesso è necessario alzare il volume della voce. Al contrario in un edificio religioso (chiesa, moschea, ecc.) le persone tendono a parlare sottovoce e muoversi cercando di produrre il minor rumore possibile, per ragioni di tipo sociale. Questo tipo di influenze, al contrario di quelle intrinseche, possono variare durante la conversazione.

### 3.3 Influenze dinamiche

Lo stato mentale di una persona ne influenzerà enormemente il comportamento: ne altererà la prosodia del parlato, l'ampiezza delle espressioni facciali, il ritmo dei movimenti. Per esempio una persona semplicemente non parla nello stesso modo se è molto arrabbiata o no. Anche il tipo di relazione che ha con l'interlocutore contribuirà a cambiarne il comportamento, che sarà diverso a seconda che stia parlando con un amico, uno sconosciuto, un collega di lavoro, un bambino, ecc. (De Carolis *et al.* 2002). Lo stato mentale si evolve durante la conversazione perché lo stato emotivo, gli obiettivi, le convinzioni variano anch'essi, e possiamo quindi affermare che esso è un elemento della conversazione estremamente dinamico.

Fig. 1 – Diagramma del sistema



#### 4. Descrizione del sistema

La Figura 1 mostra l'architettura generale del nostro sistema Greta. Il nostro ECA (vedi foto in Figura 2) riceve in input un testo marcato con dei tag definiti nel linguaggio APML (De Carolis *et al.* 2004) e genera un'animazione composta da parlato sintetizzato, movimento facciale, movimento gestuale, sguardo che può essere visualizzata mediante un apposito *player* da noi realizzato. APML permette di aggiungere delle funzioni comunicative al testo che l'agente deve pronunciare. La sintesi vocale è affidata al sistema Festival (Black *et al.*). Le temporizzazioni calcolate durante la sintesi sono utilizzate per regolare i movimenti prodotti separatamente dai due "motori", quello facciale e quello gestuale. Entrambi generano dei file di animazione compatibili con lo standard MPEG4 FAP/BAP (Taubin 1998), che poi vengono riprodotti dal *player* utilizzando una rappresentazione grafica realistica grazie alla libreria *OpenGL*. Internamente al sistema viene utilizzata una versione arricchita, che noi chiameremo *individualizzata* di APML (Par. 5.3) per rendere possibile la comunicazione tra le due principali componenti del sistema. È proprio nei tag supplementari che vengono aggiunti nella versione individualizzata dell'APML che sono contenute le informazioni sulle varie influenze di cui abbiamo parlato nel paragrafo precedente.

Fig. 2 – Greta





## 5. Modulo per l'espressività

Come abbiamo spiegato nel paragrafo precedente il nostro sistema riceve in input un file di testo che è stato precedentemente marcato con dei tag APML (Poggi *et al.* 2000). Come mostrato in Figura 1, il modulo per l'espressività si occupa dell'*individualizzazione* dell'APML, cioè dell'introduzione delle informazioni circa le influenze intrinseche, contestuali e dinamiche (Par.3) all'interno dei tag APML stessi per permettere poi al motore d'animazione (Par.6) di generare un'animazione facciale e gestuale coerente con tali influenze. Il modulo utilizza una base dati contenente un lungo elenco di coppie (*significato, segnale*). L'input del sistema (nel prossimo paragrafo descriveremo una per una tutte le sue componenti) è costituito da:

1. il testo che l'agente deve pronunciare, annotato con dei tag APML; per ogni tag è presente un attributo chiamato *intensityCA* che rappresenta l'intensità (nel senso di enfasi, importanza ai fini della comunicazione in atto) associata al tag;
2. la definizione dell'agente, composta da:
  - a. *fattori di predisposizione*;
  - b. *gerarchia di modalità*;
  - c. *overall activation* (v. Par.6.1);
  - d. *attributi per l'espressività* (Par. 2.3).

Per prima cosa il *modulo per l'espressività* seleziona un insieme di possibili comportamenti non-verbali: cercando nella basi dati seleziona alcuni segnali in base ai parametri di definizione dell'agente. Infine calcola i valori degli attributi per l'espressività per ogni tag APML presente in input e produce il file di output che chiameremo *AMPL individualizzato* (Par. 5.3).

### 5.1 Intensità dell'atto comunicativo (*intensityCA*)

Il parametro *intensityCA* specifica l'intensità dell'atto comunicativo (tag APML) a cui è associato. Può essere utilizzato per esprimere che un certo tag deve ricevere un'attenzione maggiore rispetto agli altri. Per esempio l'agente può aggiungere dell'enfasi nel parlato o mostrare maggiormente il proprio stato emotivo attraverso le espressioni facciali.

## 5.2 Definizione dell'agente

I parametri di definizione dell'agente servono a caratterizzarne il comportamento e contengono l'informazione ricavata dalle influenze intrinseche che abbiamo visto nel paragrafo 3.1 (al momento le influenze contestuali e dinamiche non vengono considerate).

*Fattori di predisposizione.* Ad ogni modalità viene assegnato un numero che rappresenta quanto l'agente è *espressivo* per quella modalità. Ad esempio esistono persone che sono molto espressive come espressioni facciali ma utilizzano poco il resto del corpo.

*Gerarchia di modalità.* Un agente potrebbe essere predisposto per utilizzare maggiormente alcune modalità rispetto ad altre. Per esempio potrebbe scegliere di usare prevalentemente le espressioni facciali e, nei casi in cui volesse aggiungere ulteriore enfasi, sovrapporre anche i gesti o aumentare il tono della voce. Per modellare questo tipo di comportamento abbiamo definito una gerarchia delle modalità e durante l'istanziamento dei tag in segnali il sistema tende ad utilizzare i comportamenti che meglio corrispondono alla gerarchia.

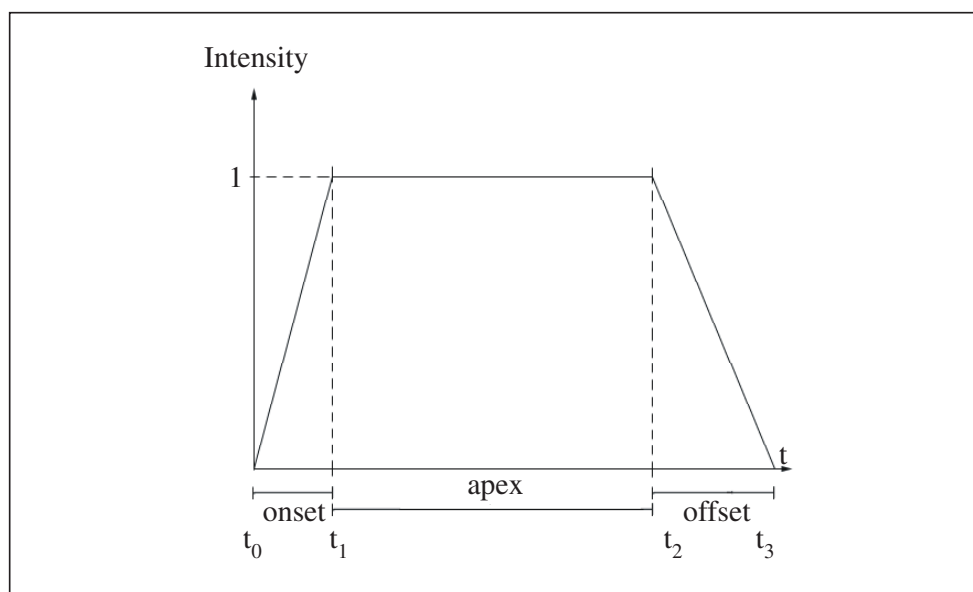
## 5.3 Generazione dell'APML Individualizzato

L'output del modulo per l'espressività è un testo marcato con tag APML ai quali sono stati aggiunte delle liste di possibili segnali da associare a ciascun tag più l'elenco degli attributi per l'espressività. All'inizio, per ogni possibile segnale associato a ogni tag viene letto dal database dell'agente il *range* (*min*, *max*) di valori in cui possono variare gli attributi per l'espressività (detti *domainValues*), più un valore di riferimento. L'attributo *intensityCA* associato al tag viene comparato a questi valori e vengono calcolati gli attributi per l'espressività da scrivere in output nella versione "individualizzata" del tag APML. Infine il modulo sceglie quali segnali vanno effettivamente espressi in base alla gerarchia delle modalità e in modo da non provocare conflitti (Pogg, Pelachaud 2002). Nel caso in cui *intensityCA* non sia compreso nell'intervallo dei *domainValues*, il processo di selezione prosegue identificando il segnale per il quale l'intervallo dei *domainValues* è la migliore approssimazione di *intensityCA*. Al contrario, se *intensityCA* è contenuto in più di un intervallo, il segnale viene scelto in base alla distanza tra *intensityCA* e il valore di riferimento.

## 6. Motore d'animazione

Passiamo ora alla descrizione del secondo modulo del nostro sistema, che nella Fig.1 è nella metà sinistra dello schema. In input il motore riceve il testo ottenuto dall'elaborazione effettuata dal *modulo per l'espressività*. Come possiamo allora guidare la generazione dei comportamenti non-verbali (facciali e gestuali) per comunicare il contenuto espressivo richiesto? Nei prossimi paragrafi suggeriremo un modello per il trattamento dell'espressività nella fase di sintesi dell'animazione. Nell'elenco degli attributi per l'espressività introdotti nel paragrafo 2.3, solo *Overall activation* agisce uniformemente a livello globale e verrà ora preso in esame per primo. Poi, per ciascuna delle due modalità considerate – faccia e gesti – riproporremo l'elenco di tutti gli altri parametri e vedremo quale può essere la loro influenza sulla generazione dell'animazione finale.

Fig. 3 – Svolgimento temporale di un'espressione facciale



### 6.1 Overall activation

Questo elemento del sistema (Fig.1) svolge una funzione di filtro tra le varie modalità. Ogni tag APML porta con sé un attributo che ne definisce l'intensità, cioè l'importanza comunicativa (Par. 5.1). I tag per i quali questo valore non supera una certa soglia non vengono presi in considerazione e quindi nessun comportamento non-verbale viene generato. Un filtro simile è stato presentato e implementato da Cassell *et al.* (2001).

## 6.2 Motore facciale

Un'espressione facciale viene identificata non solo da una configurazione dei muscoli della faccia. Il modo in cui un'espressione facciale si evolve nel tempo ne determina infatti l'espressività (Par. 2). Nella Fig. 3 possiamo osservare il tipico svolgimento temporale di un'espressione (Ekman 1979): l'asse delle ordinate rappresenta l'intensità dell'espressione, dove con il valore 0 indichiamo l'assenza totale di tensione nei muscoli facciali mentre il valore 1 significa che la configurazione facciale che rappresenta l'espressione è stata raggiunta; il tempo necessario per passare da *intensità=0* a *intensità=1* è chiamato *onset* (nella figura corrisponde all'intervallo  $[t_0; t_1]$ ); il tempo durante il quale l'espressione viene mantenuta è detto *apex* (intervallo  $[t_1; t_2]$ ); il tempo necessario a tornare nella posizione di riposo è chiamato *offset* (intervallo  $[t_2; t_3]$ ). Come detto, il linguaggio APMML utilizzato per descrivere il testo ricevuto in input dal nostro sistema permette di specificare il significato di ogni atto comunicativo dell'agente (Pelachaud *et al.* 2002); successivamente il motore di animazione facciale effettua una ricerca in una libreria di espressioni per ciascuna delle quali vengono descritte le caratteristiche temporali dei muscoli facciali coinvolti. Una volta istanziate, le espressioni facciali vanno opportunamente modificate in modo da aggiungere le caratteristiche di espressività secondo quanto esposto nel Par. 2. Vedremo ora come il nostro set di parametri per l'espressività (Par. 2.3) viene utilizzato per alterare lo svolgimento temporale e spaziale delle espressioni facciali.

*Spatial extent*: questo parametro determina la quantità di contrazione dei muscoli facciali; a livello di implementazione si traduce in un fattore moltiplicativo associato ai parametri d'animazione delle sopracciglia e della bocca che ne provoca un maggiore o minore spostamento durante l'animazione.

*Temporal*: in generale, un valore pari a 0 per questo parametro significa che l'espressione si deve svolgere con una temporizzazione che segua l'allineamento tra tag APMML e testo; un valore positivo invece provoca un allungamento dei tempi di esecuzione dell'espressione, vale a dire che, usando la notazione della Figura 3, il tempo  $t_0$  viene anticipato e il tempo  $t_3$  ritardato; un valore negativo avrà l'effetto contrario, cioè accorcerà l'espressione.

*Fluidity*: questo parametro influenza la velocità di contrazione e decontrazione dei muscoli facciali e il tipo di interpolazione impiegata

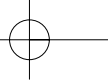
nel passaggio da un'espressione alla successiva (coarticolazione); valori positivi producono movimenti facciali più rallentati, cioè allungano gli intervalli  $[t0; t1]$  e  $[t2; t3]$ ; valori negativi al contrario danno movimenti facciali più rapidi, facendo diminuire la lunghezza degli intervalli  $[t0; t1]$  e  $[t2; t3]$ .

*Power*: questo parametro agisce sull'intensità della contrazione dei muscoli facciali: lo stesso tipo di influenza del parametro *Spatial extent*. Ma a differenza di quello, che agisce solo sui tag APML nei quali compare, *power* ha effetto su tutti i tag contenuti nel turno di conversazione dell'agente. Oltre a questo, il tipo di influenza sul movimento labiale è più complesso. Il nostro modello di movimento labiale infatti prevede la possibilità di controllare il livello di tensione labiale presente durante il parlato. Questa tensione può comparire durante la manifestazione di espressioni come la paura o la rabbia. Oppure mentre vengono pronunciate delle consonanti bilabiali come la *b*, o quando l'apertura labiale aumenta (in larghezza) e le labbra si stirano diventando più sottili. Valori positivi di *power* provocheranno un aumento della tensione labiale, valori negativi ne provocheranno la diminuzione.

*Repetitivity*: questo parametro esprime quanto spesso un'espressione facciale viene ripetuta; un valore più alto rappresenta un'alta ripetitività, che provoca ad esempio un maggior numero di "head nod" durante il turno di conversazione in cui l'agente è l'ascoltatore oppure l'aumento del numero di volte in cui l'agente solleva le sopracciglia mentre parla.

### 6.3 Motore gestuale

Basandosi sull'analisi delle funzioni comunicative contenute nel testo marcato con i tag APML, il motore gestuale sceglie i corrispondenti gesti da far eseguire all'agente Greta. Subito dopo i gesti che sono stati scelti vengono tradotti in animazione tenendo in considerazione l'insieme degli attributi per l'espressività visti nel Par. 2.3 e i vincoli di sincronia col parlato. Questa strategia di funzionamento dovrà essere applicata su più livelli – da quello di pianificazione (decidere se effettuare o no la ricerca di un certo gesto per un dato testo in input) a quello di istanziamento (decidere se ripetere o no un certo gesto, modificare il profilo dinamico di esecuzione del gesto). Come abbiamo già fatto per le espressioni facciali vedremo ora come l'insieme di attributi per l'espressività che viene aggiunto



nell'APML individualizzato influenza la generazione dell'animazione gestuale.

*Spatial extent*: questo parametro è stato modellato in modo da espandere (per valori positivi) o contrarre (per valori negativi) lo spazio di fronte all'agente in cui si muovono le mani durante l'esecuzione dei gesti. Questo spazio è ricoperto da un insieme di settori secondo il diagramma di McNeill (1992).

*Temporal*: assumendo che per motivi di sincronia la fine della fase di *stroke* del gesto deve coincidere con la sillaba più enfatizzata del parlato (McNeill 1992; Kendon 2004), possiamo calcolare la temporizzazione delle altre fasi del gesto (precedenti e successive allo *stroke*) utilizzando la legge invariante descritta da Gibet *et al.* (2004).

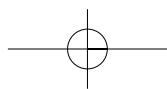
*Fluidity*: il concetto di fluidità per i gesti indica quanto il movimento delle braccia è continuo tra l'esecuzione di un gesto e il successivo (coarticolazione). Per l'animazione delle braccia del nostro agente utilizziamo delle curve matematiche definite sull'intera animazione, quindi per modificare la loro fluidità possiamo agire sui loro parametri di definizione.

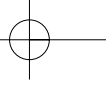
*Power*: per visualizzare la quantità di energia e tensione del movimento facciamo riferimento ancora alle proprietà dinamiche dei gesti. Possiamo aspettarci che dei movimenti con un alto livello di energia manifesteranno grosse accelerazioni e decelerazioni e, come nel caso del parametro *Fluidity*, potremo modellare questi comportamenti agendo sui parametri delle curve di animazione.

*Repetitivity*: questo attributo controlla la tendenza a ripetere più volte di seguito lo stesso gesto. La configurazione della mano viene mantenuta durante la prima esecuzione del gesto e trasformata in un *beat* nelle successive ripetizioni, che avvengono seguendo il ritmo degli accenti del parlato.

## 7. Sviluppi futuri

In questo articolo abbiamo presentato un sistema per ECA che tiene in considerazione diversi tipi di influenze nella pianificazione del comportamento dell'agente. Per superare le difficoltà della modellizzazione di concetti complessi come le influenze culturali, della personalità e così via, abbiamo preferito realizzare un sistema che partendo da un insieme prestabilito di parametri decida come modificare il comportamento dell'agente, separatamente per la





modalità delle espressioni facciali e dei gesti. Attualmente siamo ancora nella fase di sperimentazione e i valori assegnati agli attributi per l'espressività sono del tutto empirici. Il nostro obiettivo è quello di ottenere delle strategie per la combinazione dei segnali su diverse modalità (usando ridondanza, complementarità, sostituzione, ecc.). Inoltre sarà necessaria una fase di valutazione per verificare se i cambiamenti dovuti all'espressività vengono realmente percepiti nel modo adeguato.

#### Ringraziamenti

Vorremmo ringraziare in modo particolare Myriam Lamolle e Vincent Maya per il loro contributo al nostro progetto.

