

Planiranje i sinteza animacija utjelovljenih razgovornih agenata u stvarnom vremenu pomoću BML jezika

Aleksandra Čereković

Zavod za telekomunikacije i informatiku, Fakultet elektrotehnike i računarstva, Zagreb
(aleksandra.cerekovic)@fer.hr

Sažetak— Krajem 90-tih godina prošlog stoljeća Justine Cassel je uvela izraz „utjelovljeni razgovorni agenti“, odnosno ECA agenti (engl. *Embodied Conversational Agents*). Izraz je nastao kao definicija novog trenda u svijetu 3D računalne grafike i multimodalnih sučelja. Riječ „ECA“ jednoznačno definira virtualnog, kompjuterski generiranog čovjeka koji je sposoban za prirodnu komunikaciju sa ljudima putem govora i gesti. Ovaj članak prati razvoj ECA sustava i opisuje ih, naglašavajući problematiku generiranja prirodnih animacija koje prate govor ECA. Proces planiranja i sinteze animacija je opisan polazeći od *Dialogue Management* sustava koji kombiniraju ulazne korisničke informacije i na temelju stanja konverzacije određuju ponašanje na izlazu. Ponašanje je uvjetovano i modelom ličnosti i emocionalnim stanjem agenta koji su izraženi u njegovim verbalnim karakteristikama i ekspresivnosti u izvodenju animacija. Sinteza ponašanja se vrši pomoću opisnih jezika kojima se specificira njihovo ponašanje.

Ključne riječi— utjelovljeni razgovorni agenti, ponašanje agenata, sinteza animacija, gestikulacija, ECA

I. UVOD

Povijest ljudske težnje da komunikaciju na razini računalo – čovjek svede na prirodni govor seže još u 1966. godinu kada je njemački znanstvenik Joseph Weizenbaum predstavio ELIZU [1]. ELIZA je bila prvi tzv. chatterbot koji je putem tekst konzole komunicirao sa čovjekom. ELIZA je predstavljala Rogerian terapeuta i sa setom od 200 fraza koje su u većini slučajeva bile protupitanja na postavljena pitanja odgovarala na tekst koji bi korisnik unio u konzolu. Tako npr. odgovor na izjavu “Boli me glava” bi mogao biti “Zašto kažeš da te boli glava?”. Iako uskoro su nakon ELIZE nastali PARRY, TINYMUD i Hex, ELIZA ostaje zabilježena kao prekretnica u računalnoj znanosti kao prvi programerski pothvat koji je interakciju između čovjeka i računala pokušao svesti na interakciju između dva čovjeka.

Četrdeset godina poslije ELIZE tehnologija obrade govora, detekcija ljudskih pokreta te multimodalna sučelja su zamijenila chatterbotove u komunikaciji ljudi sa računalom. Pokušaj znanstvenika da poboljšaju mapiranje tekstualnih uzoraka u govoru je proizvelo nekoliko sustava za prirodnu komunikaciju [2][3][4]. Virtualni ljudi (engl. *Virtual Characters*), grafički prikazi ljudi na računalu, su doživjeli svoj pomak iz računalnih igara i filmova na aplikacije široke primjene (telekonferencije, trgovina, edukacija) što je rezultiralo nastankom novih konverzacijskih sučelja i potaklo istraživanja na tome području. Krajem 90-tih godina prošlog stoljeća Justine Cassel je uvela izraz „utjelovljeni razgovorni agenti“,

odnosno ECA agenti (engl. *Embodied Conversational Agents*) [5]. ECA agenti su računalno generirana čovjekolika sučelja koja imaju ljudske osobine u svakodnevnoj konverzaciji, a to su:

- Prepoznavanje i razumijevanje verbalnih i neverbalnih podražaja
- Generiranje govora i neverbalnog ponašanja (gestikulacija)
- Procesiranje konverzacijskih funkcija (pravo prednosti, odgovaranje na upite, mehanizmi oporavka)
- Davanje signala koji označuju stanje konverzacije, kao i davanje doprinosa novim elementima konverzacije (smjerovima konverzacije)

Da bi se ispunili navedeni zahtjevi, sustavi bazirani na ECA agentima trebaju imati komponente koji će detektirati govor i pokrete, odnosno zahtjeve korisnika, obraditi ih sukladno stanju konverzacije te generirati odgovor putem slike i zvuka, odnosno grafičkog prikaza čovjeka koji govori i gestikulira sa korisnikom. Navedena tvrdnja se jednostavno može prikazati slikom 1. Kao ulaz sustav objedinjuje verbalni i neverbalni kanal korisnika i detektira volju korisnika iz skupljenih informacija. U središnjem dijelu ECA sustava, koji predstavlja „inteligenciju“ ECA agenta se ulazna informacija koristi kao pobuda personifikacijskog modela agenta (emocije, osobnost, kultura), a koje ovisi o stanju konverzacije. Kao rezultat se dobije ponašanje agenta koje se generira na izlazu. Sve ovo objedinjuje širok spektar istraživačkih znanosti, od obrade signala zvuka, detekcije pokreta i obrade slike, preko umjetne inteligencije i personalizacije agenata (modeliranje emocija i osobnosti), do generiranja govora i animacija pokreta tijela.



Slika 1. Općeniti model sustava sa ECA agentima

Nimalo jednostavni zadaci koji se dizajnerima sofisticiranih ECA sustava predaju na zahjev u proteklom desetljeću su potakli razna istraživanja na tome području. To je rezultiralo velikim brojem sustava sa ECA agentima koji se međusobno razlikuju, kako u primjeni tako i u

dizajnu. Aspekti primjene se kreću preko interaktivnih igara, edukacijskih aplikacija, internet informacijskih servisa, web-site vodiča, agenata za e-trgovinu itd. Gandalf, jedan od prvih „oživljenih“ virtualnih likova je kreacija Kristinna Thorissona [10]. Gandalf je 2D crtani lik, znalac o sunčevom sustavu koji razgovara s ljudima govorom i gestikulacijom te odgovara na njihove upite ili ih čak može odvesti na upoznavanje sa sunčevim sustavom. Gandalf je baziran na računalnom modelu psihološkog dijaloga koji povezuje percepcijsku analizu multimodalnih događaja i generiranje multimodalnih akcija. Nedugo nakon Gandalfa je nastala Rea (engl. *Real Estate Agent*) [11], 3D generirani virtualni prodavač kuća koji omogućava socijalnu interakciju, odnosno posjeduje elemente konverzacijskih funkcija, odgovara na različite upite, a može i dati svoj doprinos tijeku konverzacije. Nakon pionira Gandalfa i Rea agenta je uslijedio cijeli niz znanstvenih projekata koji je rezultirao raznolikim kreacijama interaktivnih virtualnih ljudi. Max[12], Greta[14], Steve[13] su primjeri nekih od njih.

Razvoj ECA sustava je pratio istraživanja o utjecaju ECA na ponašanje korisnika. Znanstveni rezultati su pokazali da prisutnost ECA agenata za vrijeme interakcije između čovjeka i računala može imati velik utjecaj na korisnika, kao npr. povećanje uključenosti korisnika te reduciranje vremena za koje korisnik može obaviti neki posao [6]. Ovi rezultati su prilično razumljivi s obzirom da ECA omogućuju prirodnu interakciju sa računalom s obzirom da već znaju na koji način treba komunicirati sa ljudima. Rickenberg i Reeves [7] su dokazali da uključenost virtualnih likova u web stranice može povećati pouzdanje korisnika, ali također i njegovu anksioznost (što može imati i pozitivne i negativne aspekte). De Angeli et al. [8] podržavaju upotrebu personificiranih likova što objašnjavaju socijalnom stimulacijom koju virtualni ljudi mogu izazvati pri komunikaciji čovjeka i računala. Scotty et al. [9] pokazuju kako interakcija sa animiranim i statičnim virtualnim ljudima može pozitivno utjecati na studente pri učenju. Elliot et al. [10] su napravili studiju o pedagoškim agentima na koje utječu emocionalna stanja studenata i koji su sposobni razlučivati o različitim aspektima pri rješavanju problema.

Sa razvojem računalne grafike današnji likovi koji se mogu modelirati su vrlo uvjerljivi 3D likovi koji se kreću i govore. Ipak, većini tih likova nedostaje individualnosti pri interakciji sa ljudima. Individualnost i uvjerljivost ECA se pokušava postići integracijom raznih aspekata poput emocija i osobnosti što zahtjeva opsežno istraživanje o tome kako ljudi reagiraju na specifične događaje u specifičnim okruženjima. Osobnost se sve češće pripisuje ECA agentu bez obzira na njegovu funkciju: ekstrovertni tipovi i prijateljski nastrojene ličnosti se najviše proučavaju kao modeli ličnosti za ECA agente. Andre et al. [15] kombiniraju specifična ponašanja ovisno o osobnosti i prezentacijskim vještinama. U ljudskom ponašanju kretanje ruku koje prate govor se smatraju kao sastavni dio načina komunikacije individue [46]. Model ličnosti i emocionalno stanje ECA pri interakciji s ljudima odlučuje o ponašanju ECA te o razini ekspresivnosti koju oni izražavaju pokretima. Pokreti se odnose na skupinu neverbalnih ponašanja, a uključuju pokrete lica i tijela koje kontroliraju parametri visoke razine, odnosno izrazi lica i geste. Pokreti se prezentiraju animacijama koje bi trebale biti neovisne o virtualnom liku. Primjer je

reprezentacija animacije tijela i lica koju definira MPEG-4 standard [16].

Istraživanje u ovome članku se fokusira na planiranje ponašanja te modeliranje ekspresivnosti ECA agenata pri generiranju animacija. U narednim poglavljima će biti dan pregled sustava koji djeluju na razini planiranja te metoda koje kontroliraju pokrete lica, a koje povećavaju razinu realizma ekspresivnosti pri generiranju.

II. PRINCIP ANIMACIJE VIRTUALNIH LJUDI

Jednostavno rečeno animacija znači „oživljavanje“ objekata i ljudi. Na osnovnom nivou animaciju možemo definirati kao vremensku funkciju uzastopnih okvira animacije, odnosno slika koje predstavljaju stanja objekta u definiranom vremenu. Osnovni princip generiranja animacija grafičkih prikaza ljudi je uporaba skeletona unutar mreže poligona koja predstavlja ovojnici, odnosno 3D prikaz virtualnog lika. Sa stajališta dizajnera postoje 2 osnovne tehnike koje se koriste u izradi animacija virtualnih ljudi:

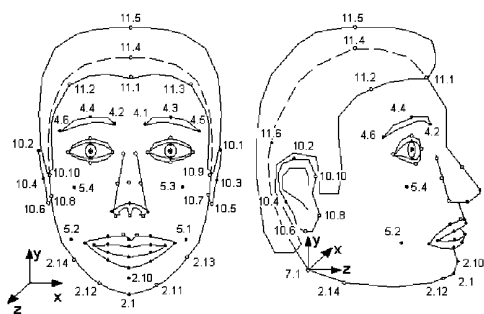
- Ručni pristup: animacija se vrši putem ključnih položaja (engl. *keyframe*) koje ručno kreira animator upravljanjem kostura. Danas u svijetu postoje mnogi alati koji omogućuju ovaj pristup (3ds Max, Maya). Ipak, ukoliko animator ne potroši puno vremena razina realizma ručno modeliranih animacija je mala.
- Snimljene animacije: animacija se snima uporabom sustava za praćenje pokreta (engl. *Motion Capture*). Većina njih koriste kamere i posebne markere koji prate pokrete lica i tijela, npr Vicon BLADE [28], ili reflektirajuće trake ili detektore pokreta. Danas se pojavljuju i nova rješenja kod kojih se eliminira upotreba dodatne opreme te se koristi samo kamera (Organic Motion [29]). Korištenjem ovakvog pristupa se šteti na vremenu, a rezultat su realistične animacije u različitim formatima (.BVH, .ASF) koje se mogu učitati u većini današnjih sustava za modeliranje i animaciju (Maya, 3ds Max). Osnovni nedostatak ovog pristupa je problem tranzicije između različitih animacija te generiranje varijacija pokreta.

Zbog anatomskih dijelova ljudskog tijela te načinom upravljanja različitim pokretima, animacija virtualnih ljudi se često se dijeli na animaciju tijela i animaciju lica.

A. Animacija lica

Razvoj metoda za animaciju lica je započeo još u '70-tim godinama prošlog stoljeća [41], no najveći napredak na tom području je postignut krajem '80-tih godina. Tehnike koje danas postoje su se razvijale paralelno sa načinima prikazavanja ljudskog lica, odnosno definicijom modela.

3D modeli glave se danas najčešće aproksimiraju mrežom poligona, mišića ili parametara. Tehnike animacije koje se baziraju na mišićima lica rezultiraju jako realističnim animacijama, ali su teške za implementaciju [43][44]. Image-morphing animacijska tehnika je interpolacija točaka lica između sekvenci slika ili videa, a najčešće se koristi pri animaciji govora [42]. Tehnika se temelji na označavanju mreže poligona ili parametara na 2D slikama te pridruživanjem odgovarajućim modelima. Parametrizacija je najzastupljenija metoda modeliranja lica. Parametrizaciju su prvi uveli Ekman i Freisen [45]



Slika 2. Parametri za animaciju lica prema MPEG-4 FBA specifikaciji

definicijom FACS (engl. Facial Action Coding System) sustava. FACS definira regije na licu koje omogućuju izražavanje gesti lica.

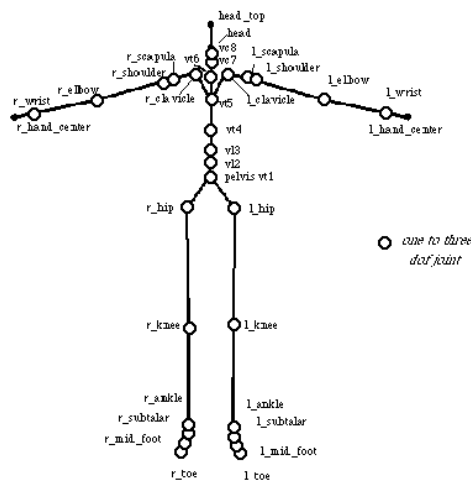
Zasada jedini postojeći standard za animaciju lica se temelji na prikazivanju lica parametrima. MPEG-4 FBA (engl. *MPEG-4 Face and Body Animation*) specifikacija je dio MPEG-4 standarda koja pruža učinkovit prikaz oblika animacije tijela i lica [30]. Široko je prihvaćen u istraživačkim krugovima, a kroz nove proizvode i usluge prodire i u komercijalne sustave. MPEG-4 FBA specifikacija definira dva seta parametara: prvi set definira geometriju i deformaciju FBA modela tijela (oblik, tekstura), a drugi definira animaciju lica i tijela. MPEG-4 FAP (engl. *Facial Animation Parameters*) definira 68 parametara koji pokreću pojedine dijelove lica, a od kojih se 2 parametra koriste za definiciju izraza lica visokog nivoa (vizemi, izrazi lica).

B. Animacija tijela

Animacija tijela se danas uglavnom svodi na pokretanje kostura, odnosno upravljanje stupnjevima slobode kostura. Ovaj pristup omogućuje izradu animacija koje su neovisne o modelu kojim se upravlja. Danas postoje dva standarda za animaciju tijela, MPEG-4 i H-Anim.

MPEG-4 FBA specifikacija, dio standarda MPEG-4 za animaciju tijela, koristi 186 BAP parametara (engl. *Body Animation Parameters*). BAP parametri su kutovi rotacije spojnih točaka koje spajaju različite dijelove tijela. Zglobovi tijela uglavnom imaju 2 stupnja slobode gibanja (engl. *Degrees of freedom, DOF*) i to rotacije (2 stupnja slobode, 2 osi). Translacija definirana samo za cijeli kostur. Spojne točke tijela koje definira MPEG-4 FBA specifikacija su: nožni prst, članak, koljeno, bedro, kralježnica (3 dijela), rame, ključna kost, ručni zglob i prsti ruku. Rotacijski kutovi su pozitivni ukoliko se kreću u smjeru kazaljke na satu. Specifikacijom se mogu opisati i složeni pokreti ruku koje su također uključene u hijerarhiju tijela. Ukupno postoji 29 stupnjeva slobode za svaku ruku.

H-Anim specifikacija za animaciju virtualnih ljudi je dio VRML standarda [31]. MPEG-4 FBA je preuzeo način definiranja kostura iz H-Anima i to definiciju točaka na kosturu i stupnjeve slobode DOF. Osnovna razlika u između definicije tijela u MPEG-4 i H-Animu je u znatno većoj kompresiji struje podataka virtualnog lika u MPEG-4 standardu [32]. Osim toga, H-Anim virtualni lik može



Slika 3. Kostur za izradu animacija tijela prema MPEG-4 FBA specifikaciji

postojati samo u VRML svijetu, dok se MPEG-4 sintetički kodira.

Kako se vidi iz prikaza kostura, pokretanje kostura je prilično složen zadatak. Postoji velik broj stupnjeva slobode, a teško je postići da animacije izgledaju prirodno. Pri izradi animacija danas se najviše pristupa ključnim okvirima (engl. *key frames*). Temelj ovoga pristupa je definiranje niza sekvenci ključnih okvira između kojih se radi interpolacija (engl. *interpolation*). Animator sekvence animacije izrađuje sam (ručna tehnika), a slijedi pravilo da između susjednih sekvenci ne smije biti drastičnih pomaka kako bi se izbjegli neprirodni položaji (npr. zavrtnanje ruke). Linearna interpolacija, najjednostavniji oblik interpolacije, često ne daje prirodne rezultate pri izradi animacija. Različite brzine pokreta npr. geste rukom koje su u prvom dijelu ubrzane, a u drugom usporene nije moguće postići linearnom interpolacijom. Zbog toga postoje i druge metode, poput Hermite interpolacije, SLERP [33], SQUAD [34].

Drugi krajni način izrade animacija, spomenut u uvodnom dijelu ovog poglavlja su postupci slijeđenja putem trackera i snimanje animacijskih sekvenci. Na ovaj način se dobijaju jako lijepe i realistične animacije, ali metoda je skupa i kako raspolaže sa velikom opremom, nepraktična. Na internetu postoje baze snimljenih animacija koje se mogu kupiti ili besplatno downloadati [35][36], ali to ne rješava problem fleksibilnosti i upravljanja – ukoliko želimo promijeniti neki pokret, u najčešćem slučaju ćemo morati snimiti novu sekvencu. Promjena brzine izvođenja snimljenih animacija također ne znači dobar rezultat – promjenjena dinamika ne mora izgledati prirodno.

Inverzna kinematika, preuzeta iz robotike je metoda kojom se iz konačnog položaja tijela mogu odrediti promjene kuta zglobova koji vode do toga položaja. Pristup je naivan ukoliko se žele dobiti prirodne sekvence – postoji velik broj rješenja, a samo neka izgledaju prirodno. Sustavi sa inverznom kinematikom razvijeni u industriji koriste IK samo na udovima (Alias Maya) ili mjerenjem sličnosti sa pozom koju zovemo referentnom pozom [37][38]. U tome slučaju animator mora zadati velik broj ograničenja. Na internetu su dostupna rješenja istraživačkih grupa koja nisu u komercijalne primjene,

poput IKAN biblioteke [39] razvijene u programskom jeziku C++. IKAN poput Aliasa koristi IK samo na rukama i nogama te u nekim slučajevima daje animacije koje izgledaju mehanički. S druge strane postoje i rješenja koja se baziraju na skupu poza dobijenih iz videa stvarnih ljudi koja generiraju prirodnije animacije[40].

III. GENERIRANJE PONAŠANJA UTJELOVLJENIH RAZGOVORNIH AGENATA

Planiranje i sinteza ponašanja ECA agenata zahtjeva daleko više od izrade animacija koje se generiraju na izlazu sustava. Kako je rečeno u uvodnom dijelu članka, ponašanje agenata ovisi o velikom broju aspekata: ponašanju korisnika, modela ličnosti koji utječe na ekspresivnost agenata, emocionalnog stanja, kulture iz koje dolaze, do onoga što utjelovljuju (prezenteri vijesti, psihoterapeuti, trgovci).

U ovom odjeljku će se razmotriti aspekti generiranja ponašanja ECA od središnjeg dijela modela ECA sustava (slika 1) do ekspresivnosti u generiranju pokreta tijela koji su odgovori na zahtjeve korisnika.

A. Dialogue Management sustavi

Dialogue Management sustavi (engl. *Dialogue Manager Systems*) su kompleksne tehnologije koje uključuju obradu govorne informacije, umjetnu inteligenciju (engl. *Artificial Intelligence*), prezentiranje znanja i planiranje.

ELIZA, prvi chatterbot program sa tekstualnom konzolom nije imao sofisticiran dijaloški model. Temelj komunikacije sa ELIZOM je bilo mapiranje dijela ulaznih podataka (odnosno tekstualnog niza) u izlaznu rečenicu, što je bilo vrlo jednostavno implementirati. Alice[17], svojevrsna ekstenzija ELIZE za reprezentaciju dijaloškog modela koristi jezik umjetne inteligencije AIML (engl. *Artificial Markup Language*) [18]. Alice je tri puta dobio nagradu „Loebner Prize Contest for Most Human Computer“ no današnji zahtjevi korisnika nadilaze zahtjeve koje pokriva AIML. AIML je baziran na XML jeziku te kao ELIZA koristi mapiranje teksta u odgovore. Nedostatak ovoga pristupa je posloženje modela dijaloga ukoliko bi se u ECA sustav uvele emocije i osobnost te kontrola toka konverzacije (koju su neki znanstvenici pokušali napraviti pomoću automata stanja [19]). Kao rezultat je uslijedila izgradnja objektno orijentiranih distribuiranih arhitektura za DM kao što su Communicator[20], RavenClaw [21][22], Queen's Communicator [23], SesAME [24] i Jaspis [25]. U svim ovim pristupima dijalog je implementiran koristeći objektno-orijentirani pristup nasljeđivanja.

U domeni virtualnih agenata su predstavljeni i drugi DM sustavi poput BDI-ja, koji je originalno izrađen kao alternativna pravom planiranju koja je mogla balansirati između relativnog i dogovornog ponašanja [26][27]. BDI model se može aproksimirati crnom kutijom koja za kontrolu toka konverzacije koristi statistički model nastao iz korpusa pravih ljudskih konverzacija. Osnovni nedostatak ovog modela je nedostatak emocija.

B. Model ličnosti i emocije ECA agenata

Kako bi se postigla uvjerljivost u simulaciji ponašanja ECA, odnosno kako bi se agenti svojim ponašanjem približili čovjeku, znanstvenici pribjegavaju raznim pristupima u psihologiji. Istraživanjima se nastoji povezati

različiti model ličnosti, emocionalna stanja i raspoloženja sa ulaznim informacijama u dizajniranju centralnog dijela ECA sustava (slika 1). Ukratko, traži se najoptimalniji način koji osobnost i emocije mogle utjecati na ulazne informacije i na ponašanje ECA.

Osobnost ili model ličnosti su karakteristike ponašanja koje jednu osobu razlikuju od druge. Emocije su trenutno stanje svijesti, a raspoloženje je rezultat kumulativnog djelovanja emocija koje traje dulje od emocija. Koristeći navedene definicije moguće je napraviti vezu između modela ličnosti i emocija na razini aktivacija (model ličnosti) -> evaluacija (emocije). Ovaj pristup se često koristi u izgradnji ECA sustava.

Najpopularniji model ličnosti u psihologiji je pet-faktorski model ličnosti koji se pokazao snažnijim i replikabilnijim od ostalih taksonomija [66]. 5-faktorski model razlikuje 5 crta koje definiraju ličnost: ekstraverzija, ugodnost, savjesnost, emocionalna stabilnost i otvorenost za nova iskustva. Eysenck smješta različite crte ličnosti u 2- dimenzionalni koordinatni sustav u skladu s njihovom pozicijom uzduž glavnih dimenzija ličnosti: introverzija-ekstraverzija i neuroticizam. Ekman [68] definira 6 univerzalnih izraza lica koji su rezultat emocija, a iz kojih se kombinacijom mogu dobiti drugi izrazi lica. To su izraz sreće, tuge, ljutnje, straha i gađenja. Njegova definicija je vrlo popularna i koristi se za istraživanje u procesu izrade animacija lica. OCC model je nastao iz kognitivne teorije emocija [69], a koja definira emocije u ovisnosti o procjeni pojedinca da li su neki događaji dobri ili loši za njega. OCC model razlikuje 22 emocije i to 4 koje definira Ekman kao izraze lica (sreća, tuga, strah, ljutnja).

Kshirsagar i Magnenat-Thalmann [67] za ECA sustave razmatraju problem pronalaženja osobnosti agenata kroz: logičku razinu („mozak“) agenta, emocije, ekspresivnost u izvođenju gesti i fizičku pojavu agenta. Prema radu Vel'asqueza [70] predlažu slojeviti pristup modeliranju emocija (osobnost, raspoloženje, emocije) te implementiraju 5-faktorski model koristeći Bayesian mreže. U DM sustavima za obogaćivanje ekspresivnosti izgovoreni rečenica koriste OCC model. OCC model u sintezi izraza lica povezuju sa Ekmanovom klasifikacijom. André et al.[71] koriste kognitivnu strukturu emocija i 5-faktorski model ličnosti na tri različita projekta: virtualni recepcionar, virtualni teatar i trgovački centar. Generator neverbalnog ponašanja kojeg predlažu Egges i Magnenat-Thalmann [72] se temelji na aktivacijsko-evaluacijskom modelu, odnosno 2-dimenzionalnom modelu kojim se predstavlja ovisnost emocija o stanju ličnosti. Svakom intervalu emocija sa modela je pridružen skup gesti koje su snimljene slijeđenjem pravih ljudi. Nakon odabira intervala, iz skupa gesti se nasumce odabire gesta koja odgovara tom intervalu.

C. Ekspresivnost ECA agenata

Ekspresivnost ECA agenata predstavlja način na koji se generiraju ponašanja agenta koja prate govor, a odnose se na sintezu gesta lica, tijela, pogled agenta i položaje tijela. Ekspresivnost se istražuje proučavanjem i anotacijom videa ponašanja ljudi u stvarnom svijetu, ali i likova iz crtanih filmova [47] te dovodenjem ponašanja u direktnu vezu sa modelom ličnosti agenata i emocionalnim stanjem.

U procesu komunikacije ljudi svjesno ili nesvjesno rade različite pokrete. Današnja istraživanja na području ECA

se pojedinačno fokusiraju na dijelove tijela koja sudjeluju u komunikaciji: na geste lica/glave, na geste ruku, pogled, položaje tijela. Razlog ove podjele je složenost izrade prirodnih animacija i granularnost koja se dobija ovim pristupom- kombinacija više gesti može imati različito značenje.

Postoji više načina podjela gesti ruku. Prema kognitivnoj konstrukciji Poggi [48] razlikuje 2 vrste gesti ruku; kodirane geste i kreativne geste. Kodirane geste su pohranjene u dugotrajnom pamćenju (kao veza: gesta - značenje) poput *emblems* gesti (engl.; geste koje se mogu literano prevesti, npr. „OK“ palac podignut prema gore) i *battons* (engl.; geste koje naglašavaju govor, npr. mali pokreti rukom). Kreativne geste mogu biti *iconic* (engl.: opisivati neki objekt) ili *deictic* geste (apstraktne geste, javljaju se ako je taj objekt prisutan npr. pokazivanje).

Geste lica Zorić et al.[49] dijele na:

- pokrete glavom (mogu izražavati slaganje, neslaganje ili naglašavati govor)
- pokrete usnama (naglašavaju govor, izražavaju emocionalno stanje, biološke potrebe)
- pokrete obrvama (najčešće naglašavaju govor)
- pokrete očnim kapcima (biološki signali, naglašavaju govor)
- pokrete očima
- pokrete na čelu (neslaganje)

Oči su najizraženiji komunikacijski kanal i u njima se odražavaju različite emocije. Koriste se u naglašavanju govora te u kontroli toka razgovora (npr. kada želimo nekome reći da smo završili sa razgovorom te da je „na njega red“). Postoje različite studije i sustavi za ECA agente koje se bave problemom komunikacije očima [50][51][52].

Promjene položaja tijela se koriste u komunikaciji poput gesti, a ovisno o kulturi i socijalnim odnosima. Osobe koje su u bližim odnosima će stajati na manjoj udaljenosti sa tijelom okrenutim jedno prema drugom. Položaji tijela se istražuju i u tzv. stanjima mirovanja [53] (engl. *Idle Motions*). Stanja mirovanja su prirodna stanja ECA koji miruju i čekaju da im korisnik pristupi (agenti nisu „zamznuti“ nego proizvode pokrete male amplitude, npr. promjene stajne noge, češkanje, stavljanje ruke u džep – kao osoba koja miruje).

Neverbalni pokreti u komunikaciji se u postojećim ECA sustavima sintetiziraju na različite načine.

Sinteza ponašanja u REA agentu [11] (MIT Media grupa) se vrši pomoću BEAT-a. BEAT toolkit, [54] koristi lingvističku i kontekstualnu analizu teksta za sintezu ponašanja. Princip je da korisnik utipka tekst kojeg agent izgovara, a kao rezultat se dobijaju animacije koje su sinkronizirane s govorom. Max agent [12] koristi DM sustav koji opisuje scenarij ponašanja. Animacije se generiraju automatski, sljedeći XML specifikaciju ponašanja agenta. GRETA [14] automatski generira geste koristeći ključne položaje točke lica, dok za geste ruku koristi biblioteku IKAN [39].

Većina animacija u opisanim sustavima se generira proceduralno, odnosno programskom kontrolom parametara na tijelu (definiranih npr. MPEG-4 standardom) koja prati fizikalne zahtjeve gibanja. Na ovaj način se postiže visoka kontrola pokreta, ali rezultat su najčešće animacije koje nemaju prirodnu dinamiku



Slika 4. Rezultat sinteze gesti u ECA sustavima: REA; GRETA; MAX

kretanja. Problem izrade realističnih animacija se svodi na problem pridruživanja gesti na visokoj razini parametrima niske razine, npr. kada čovjek želi uzeti neku stvar, onda će posegnuti cijelim tijelom, a ne samo rukom. Istraživanja u ovoj domeni se baziraju na traženje sredine između sekvenci koji se dobijaju slijeđenjem pokreta pravoga čovjeka (engl. *Motion Capture Data*) i proceduralnih animacija. Stone et al.[55] su razvili kolekciju alata koji pomažu u anotaciji i ekstrakciji parametara skupljenih iz podataka dobijenih slijeđenjem. Dobijeni parametri pomažu u sintezi novih animacija u ECA sustavu. Sinteza animacija radi samo sa snimljenim govorom što onemogućuje korištenje alata u sustavima koji koriste sintezu govora iz teksta (engl. *Text-To-Speech*). EMOTE Model [56] istraživačke grupe HMS predlaže model koja se temelji na Laban analizi i *Effort* (engl.; trud, napor) i *Shape* (engl.; oblik) parametrima kojima se kontrolira visoki nivo animacija. Kshirsagar et al. [57] koriste principijelnu komponentu ili PCA (engl. *Principal Component Analysis*) analizu kojom se FBA parametrima MPEG-4 standarda pridružuju svojstveni vektori lica. Ovaj način se koristi za sintezu lica pri animaciji govorom, a uključuje vizeme i izraze lica. Princip PCA analize su naslijedili Edges i Molet [64] u sintezi promjena položaja tijela. Iz animacija dobijenih slijeđenjem stvarnih pokreta oni generiraju PC vektore te računaju euklidske udaljenosti kako bi izvršili prijelaz između ključnih okvira animacija koji su pohranjeni u bazi pokreta. Michael Kipp et al. [65] predlažu sustav za planiranje i sintezu gesti ruku koji se temelji na anotaciji videa stvarnih osoba. Iz videa se stvaraju sekvence okvira iz kojih se kasnije vrši sinteza putem jedinica gesti (engl. *G-Units*) i ulaznog teksta. Njihov princip može kreirati geste ovisno o osobi i daje prilično uvjerljive rezultate, no generira samo pokrete ruku.

D. Jezici za opis ponašanja ECA agenata

Razvoj sustava sa ECA agentima je pratio razvoj različitih jezika za opis njihovog ponašanja. Max system koristi MURML opisni jezik (engl. *Multimodal Utterance Representation Markup Language*) [58] sa vremenskim anotacijama koje omogućuju sinkronizaciju pokreta i govora. APML (engl. *the Affective Presentation Markup Language*) [59] je nastao na osnovu konverzacijskih funkcija gesti koje definira Poggi [48]. Za ECA sustave ona predlaže leksikon sa parovima: značenje – signal. Ovakva taksonomija omogućuje podjelu između nivoa planiranja ponašanja agenta i nivoa realizacije. RRL opisni jezik (engl. *Rich Representation Language*) [60] je vezan za DM sustav te se može koristiti za opisivanje ponašanja većeg broja agenata u sceni. VHML (engl. *Virtual Human Markup Language*) [63] je dizajniran kako bi se koristio u DM sustavu, sintezi animacija lica i tijela, a koristi XML/XSL.

Većina spomenutih jezika je bazirana na XML-u i ima međusobno velike sličnosti. Kao posljedica u istraživačkoj zajednici se javila potreba za osmišljavanjem univerzalnog jezika opisa ponašanja virtualnih ljudi. Zbog toga je 2005. godine od strane znanstvenika sa različitih područja (stručnjaka za računalne znanosti, psihologa i lingvisti) pokrenuta BML inicijativa. Grupa je i danas aktivna, a cilj njihove suradnje je stvoriti standard koji bi se koristio kako u istraživačkoj zajednici, tako i u industriji. Kao rezultat djelovanja je nastao SAIBA Framework (engl. *Situation, Agent, Intension, Behavior, Animation*, slika 5.). SAIBA framework objedinjava multimodalno generiranje ponašanja virtualnih agenata i model realizacije ponašanja i podijeljen je u tri faze (planiranje namjere, planiranje ponašanja i realizacija ponašanja). Framework koristi BML jezik (engl. *Behavior Markup Language*) za opis multimodalnog ponašanja agenta [61][62] i FML jezik (engl. *Functional Markup Language*) koji opisuje razinu planiranja ponašanja. Iako još nije doživio svoju konačnu verziju, BML jezik se sve više koristi u različitim istraživačkim projektima. Za razliku od njega, FML jezik još nema svoju potpunu strukturu.



Slika 5. SAIBA Framework

BML opisni jezik je dio baziran na XML jeziku i koristi 8 osnovnih BML elementa nazvanih ili prema anatomskim dijelovima tijela (glava, lice...) ili prema funkcijama u neverbalnom ponašanju (kretanja, pozicija, gesta). Elementi omogućuju opis ponašanja na visokoj razini. Sinkronizacija elemenata je omogućena putem 7 sinkronizacijskih točaka definiranih za svaki element. Taksonomija točaka dijeli gestu na 6 različitih faza: start, ready, stroke-start, stroke, stroke-end, relax, end. BML omogućuje korištenje i drugih reprezentacijskih jezika (MURML, SML...) te ponašanja/animacija koje nisu definirani BML elementima. Na ovaj način je omogućena sloboda kreativnosti pri implementaciji BML realizatora što BML jezik razlikuje od drugih opisnih jezika. Naime, izvođenje tzv. *Core BML-a* (BML elementi) mora biti sastavni dio svakog BML realizatora. Izvođenje ostalih animacija, odnosno ponašanja agenta koja se ne mogu opisati putem BML-a je unutar realizatora moguće omogućiti proširenjem tagova BML skripti sa *namespaceov*-ima i opisnim razinama (engl. *Description Level*). Stoga, velika je vjerojatnost da dva BML realizatora razvijena u različitim istraživačkim grupama imaju različite mogućnosti izvođenja.

IV. ZAKLJUČAK

U ovome članku je predstavljen općeniti model sustava sa utjelovljenim razgovornim agentima - ECA agentima (engl. *Embodied Conversational Agents*). Model prema funkcionalnosti definira tri dijela ECA sustava: ulaz, centralni dio i izlaz. Na ulaznom dijelu se prikupljaju informacije korisnika, dok centralni dio čini „mozak“ agenta koji odlučuje kakav će biti odgovor agenta na zahtjev korisnika. Odgovor agenta objedinjuje skup gesti koje se u sinkronizaciji sa govorom generiraju na izlazu sustava.

Članak se fokusira na izradu animacija za ECA agente, odnosno na izlazni dio sustava. U prvome dijelu je dan pregled općenitih načina izrade animacija te je predstavljen standard za sintezu animacija virtualnih ljudi. Radi boljeg razumijevanja složenosti i problematike izrade skupa animacija za ECA sustave prezentirano je područje o kojem ovisi pristup izrade, a kojeg čine *Data Management* sustavi, modeli emocija i ličnosti u centralnom dijelu sustava te opisni jezici koji specificiraju ponašanje agenta na izlazu.

Unatoč velikim doprinosima znanstvenika na ovome području, ECA agentima još uvijek nedostaje individualnosti u izražavanju pokreta tijela. Istraživanja na razini sinteze prirodnih animacija koja su opisana o članku bi se mogla iskoristiti u izgradnji generatora animacija koji bi uključio nadolazeći BML jezik (engl. *Behavior Markup Language*) za opis ponašanja agenata. Izrada animacija za BML realizator bi uključio traženje efikasnog načina kojim bi se specificirala jednostavna izrada prirodnih animacija tijela i gesti lica za „univerzalnog“ ECA agenta. Također, vrlo je važno napomenuti i utjecaj modela ličnosti i emocija na pokrete koji se generiraju na izlazu. Sve skupa navedeno otvara i mogućnost za izradnju frameworka za planiranje i sintezu animacija utjelovljenih razgovornih agenata uz pomoć BML jezika.

REFERENCE

- [1] J. Weizenbaum, "ELIZA - A computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, pp. 36-45, 1966.
- [2] Alice, (2005), Artificial Linguistic Internet Computer Entity, [Online]. Available: <http://www.alicebot.org>
- [3] B. Alfonsi, "'Sassy' Chatbot Wins with Wit," in *IEEE Intelligent Systems*, 2006, pp. 6-7.
- [4] J. Hutchens and M. D. Alder, "Introducing MegaHAL," presented at The Human-Computer Communication Workshop, 1998.
- [5] Cassel, "Embodied Conversational Interfaces Agents", *Communications of the ACM*, vol. 43., No. 4, p.p. 70 – 78, 2000
- [6] Blom, J. & Monk, A. "One-to-one e-commerce: who's the one?" *Proceedings of CHI 2001: Conference on Human Factors in Computing Systems*. Seattle, Washington, 31 March - 5 April 2001.
- [7] Rickenberg, R. and Reeves, B. The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces. *CHI 2000 The Hague, Amsterdam, The Netherlands*. 1-6 APRIL 2000.
- [8] De Angeli, A.; Lynch, P. and Johnson, G. Personifying the e-market: A framework for social agents. *Interact 2001 - Eighth IFIP TC.13 Conference on Human-Computer Interaction*, Tokyo, Japan, July 9-13, 2001.
- [9] Scotty D. Craig, Barry Gholson, and David M. Driscoll. *Animated Pedagogical Agents in Multimedia Educational Environments*. *Journal of Educational Psychology*, 2002, Vol. 94, No. 2, 428-434.
- [10] Thorisson K., PhD thesis: "Communicative Humanoids A Computational Model of Psychosocial Dialogue Skills", 1997
- [11] Cassel et. al, "Embodiment in Conversational Interfaces: Rea", *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, p.p.: 520 - 527, 1999
- [12] Kopp S., B. Jung, N. Lessmann, I. Wachsmuth: "Max--A Multimodal Assistant in Virtual Reality Construction". *KI 4/03: 11-17.*, 2003
- [13] Gratch J, Marsella S. Tears and Fears: modeling emotions and emotional behaviors in synthetic agents. In: *Proceedings of the Fifth International Conference on Autonomous Agents*, Montreal, Canada, June 2001.

- [14] Pelachaud C. et al.: "Embodied contextual agent in information delivering application", Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2, p.p. 758-765, 2002.
- [15] Andre, E., Rist, T., Van Mulken, S., Klesen, M., and Baldes, S. (2000) The automated design of believable dialogues for animated presentation teams. In: Cassell, J., Prevost, S., Churchill, E. (Eds.) Embodied Conversational Agents, pp. 220-255, MIT Press, Cambridge.
- [16] MPEG-4. <http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm>
- [17] Alice webpage. <http://www.alicebot.org>.
- [18] <http://en.wikipedia.org/wiki/AIML>
- [19] G. E. Churcher, E. S. Atwell, and C. Souter. Dialogue management systems: a survey and overview. Technical Report 97.6, School of Computer Studies, University of Leeds, February 1997.
- [20] <http://communicator.sourceforge.net/index.shtml>
- [21] Rudnicki, A. Thayer, E., Constantinides, P., Tchou, C., Shern, R., Lenzo, K., Xu, W., Oh, A. Creating Natural Dialogs in the Carnegie Mellon University Communicator System. In Proc. of Eurospeech 1999: 1531-1534, 1999.
- [22] Bohus, D., Rudnicki, A. RavenClaw: Dialog Management Using Hierarchical Task Decomposition and an Expectation Agenda. In Proc. of the Eurospeech 2003: 597-600, 2003.
- [23] O'Neill, I., Hanna, P., Liu, X., McTear, M. The Queen's Communicator: An Object-Oriented Dialogue Manager. In Proc. of the Eurospeech 2003: 593-596, 2003.
- [24] Pakucs, B. Towards Dynamic Multi-Domain Dialogue Processing. In Proc. of Eurospeech 2003: 741-744, 2003.
- [25] Turunen, M., Hakulinen, J. Jaspis, J. - A Framework for Multilingual Adaptive Speech Applications. In Proc. of 6th International Conference of Spoken Language Processing (ICSLP 2000): 719-722, 2000.
- [26] Bratman, M.E., D.J. Israel, and M.E. Pollack, "Plans and Resource-Bounded Practical Reasoning", in Computational Intelligence, 4, 1988
- [27] Ardissono, L and G. Boella An Agent Architecture for NL dialog modeling in "Artificial Intelligence: Methodology, Systems and Applications", Springer (LNAI 2256) 1998
- [28] <http://www.mocap.lt/motion-capture/software/vicon-blade.html>
- [29] <http://www.organicmotion.com/>
- [30] Igor S. Pandzic, "Virtualna okruženja", Element, Zagreb, 2004.
- [31] <http://www.w3.org/MarkUp/VRML/>
- [32] Preda, Preteux, "Virtual Character Within MPEG-4 Animation Framework eXtension", IEEE Transactions on circuits and systems for video technology, vol 14, no 7, p.p. 975 – 988, 2004
- [33] K. Shoemake. Animating rotation with quaternion curves. In Proceedings SIGGRAPH 1985, pages 245–254. ACM Press, 1985.
- [34] W. Boehm. On cubics: a survey. Computer Graphics and Image Processing, 19:201—226, 1982.
- [35] <http://www.truebones.com/>
- [36] <http://www.mocapdata.com/>
- [37] Yamane, Nakamura, "Natural motion animation through constraining and de-constraining at will", IEEE Transactions on Visualization and Computer Graphics 9, 3, p.p. 352 – 360, 2003
- [38] Zhao, Badler, "Gesticulation Behaviors for Virtual Humans", In Pacific Graphics, p.p. 161 – 168, 1998
- [39] <http://cg.cis.upenn.edu/hms/software/ikan/ikan.html>
- [40] Grochow et al., "Style-Based Inverse Kinematics", ACM Transactions on Graphics, SIGGRAPH, 2004
- [41] F. I. Parke. Computer generated animation of faces. In Proceedings ACM Annual Conference, pages 451–457. ACM Press, 1972.
- [42] Ezzat, Poggio, "MikeTalk: A Talking Facial Display Based on Morphing Visemes", in Proceedings of the Computer Animation Conference, Philadelphia, Pennsylvania, June 1998
- [43] S. M. Platt and N. Badler. Animating facial expression. Computer Graphics, 15(3):245–252, 1981.
- [44] K. Waters. A muscle model for animating three-dimensional facial expression. Computer Graphics, 21:17–24, 1987.
- [45] P. Ekman and W. V. Friesen. Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto, CA, USA, 1978.
- [46] Kendon, A. (1980) Gesticulation and speech: two aspects of the utterance. In: Key, M. R. (Ed.) The relationship of Verbal Nonverbal Communication, pp. 207-228, Mouton Publishers.
- [47] Chafai, Pelachaud et al., "Gesture expressivity modulations in an ECA application ", in Proceedings of IVA '06 conference, Marina del Ray, California, 2006
- [48] Poggi: "Mind, Hands, Face and Body", Weidler Buchverlag, Berlin 2007, ISBN 978-3-89693-263-1
- [49] Goranka Zoric, Karlo Smid, Igor S. Pandzic: "Facial Gestures: Taxonomy and Application of Non-Verbal, Non-Emotional Facial Displays for Embodied Conversational Agents" in Nishida T.: "Conversational Informatics", ISBN-13: 978-0-470-02699-1 - John Wiley & Sons, November 2007
- [50] Lee et al.: "Eyes Alive", ACM Transactions on Graphics 21 (3), July 2002, ACM:NY, pages 637-644
- [51] Peter, Pelachaud et al.: "A model of attention and interest using gaze behavior", IVA'2005. Greece.
- [52] Torres, Cassel et al. "Modeling Gaze Behavior as a Function of Discourse Structure.", First International Workshop on HumanComputer Conversation, July, 1997
- [53] Egges, Visser, Magnenat-Thalmann; "Example-Based Idle Motions in a Real-Time Application", CAPTECH Workshop 2004. Zermatt, Switzerland.
- [54] Cassel et al. "BEAT: the Behavior Expression Animation Toolkit", Siggraph, 2001
- [55] M. Stone, D. DeCarlo, I. Oh, C. Rodriguez, A. Stere, A. Lees, and C. Bregler. "Speaking with hands: Creating animated conversational characters from recordings of human performance." In Proceedings SIGGRAPH 2004, pages 506–513. ACM Press, 2004.
- [56] Chi et al. : "EMOTE Model for Effort and Shape", in Proceedings of the 27th annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '00, pages 173-182.
- [57] Kshirsagar et al. "Principal Components of Expressive Speech Animation". In Computer Graphics International 2001, pages 38-44. IEEE Computer Society, February 2001.
- [58] Kopp S., Wachsmuth I. (2004): Synthesizing Multimodal Utterances for Conversational Agents. Computer Animation and Virtual Worlds, 15(1):
- [59] DeCarolis B., Pelachaud C., Poggi I, and Steedman M. (2004). APMML, a mark-up language for believable behavior generation. In H. Prendinger and M. Ishizuka, editors, Life-like Characters. Tools, Affective Functions and Applications, 65–85. Springer.
- [60] Piwek P., Krenn B., Schröder M., Grice M., Baumann S., Pirker H. (2002). RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA. In Proceedings of the Workshop Embodied conversational agents - let's specify and evaluate *them!*, held in conjunction with AAMAS-02, July 16 2002, Bologna, Italy.
- [61] Kopp et al.:" Towards a Common Framework for Multimodal Generation: The Behavior Markup Language", in Proceedings of the IVA '06 conference, Marina del Ray, USA, 2006
- [62] Vilhjalmsón et al.:" The Behavior Markup Language: Recent Developments and Challenges", in Proceedings of the IVA '07 conference, Paris 2007
- [63] <http://www.vhml.org/>
- [64] Edges, Moles: "Personalized Real-time Idle Motion Synthesis", In Pacific Graphics 2004, pages 121 – 130, 2004.
- [65] Kipp M, Neff M., Kipp K., Albrecht I.: "Towards Natural Gesture Synthesis: Evaluating gesture units in a data driven approach to gesture synthesis", in Proceedings of the IVA '07 conference Paris 2007
- [66] Goldberg, L. R. (1993). The structure of phenotypic personality traits. American Psychologist, 48, 26-34.
- [67] Sumedha Kshirsagar, Nadia Magnenat-Thalmann: "A Multilayer Personality Model", ACM International Conference Proceeding

- Series, Vol. 24, Proceedings of the 2nd international symposium on Smart graphics, Hawthorne, New York, Pages: 107 – 115, 2002
- [68] Ekman, P., *Emotion in the Human Face*, Cambridge University Press, New York, 1982.
- [69] Ortony, A., Clore, G., and Collins, A., *Cognitive Structure of Emotions*, Cambridge University Press, 1988.
- [70] Juan D. Velasquez. Modeling emotions and other motivations in synthetic agents. In *Proceedings AAAI-97*, pages 10--15. AAAI, AAAI Press and The MIT Press, 1997
- [71] André, E., Klesen, M., Gebhard, P., Allen, S., and Rist, T., Integrating models of personality and emotions into lifelike characters, In A. Paiva and C. Martinho, editors, *Proceedings of the workshop on Affect in Interactions - Towards a new Generation of Interfaces in conjunction with the 3rd i3 Annual Conference*, Siena, Italy, October 1999, 136-149.
- [72] Edges, Magnenat-Thalmann. "Emotional Communicative Body Animation for Multiple Characters", *V-Crowds'05*, Lausanne, Switzerland. Pages 31-40. November 2005.