

Beszéd a szavakon túl

Ruttkay Zsófia

PPKE ITK, Budapest
University of Twente, Enschede, The Netherlands,
z.m.ruttkay@cs.utwente.nl

Kivonat: Az emberi kommunikációban a beszélt nyelven túl nagy jelentősége van az azt kísérő nemverbális jeleknek: arckifejezéseknek, gesztusoknak, a tekintetnek, melyeket mindenképpen figyelembe kell venni virtuális emberek kommunikációjának tervezésekor is. A cikkben áttekintést adunk a nemverbális jelek jelentéséről, fajtáiról, morfológiai jellemzőiről. Az arc mimika lehetséges hatásait két kísérlet eredményeivel illusztráljuk. Majd bemutatjuk a GESTYLE szöveg annotáló nyelvet, melynek segítségével virtuális emberek beszédét kísérő nemverbális jeleket lehet több szinten, nemdeterminisztikus módon definiálni.

1. Bevezetés

Az emberi kommunikáció legfontosabb formája a beszélt nyelv. A kimondott, szó szerinti tartalmat modulálja, vagy akár meg is változtatja az, hogy miként hangzik el a mondat: gúnyosan, kételkedve, meglepett hangon. A beszéd jellegzetességei mellett hasonló szerepük van a nemverbális csatornákon közvetített jeleknek: a beszélő arckifejezésének, gesztusainak, tekintetének, sőt testhelyzetének. A nemverbális jelek alapvető velejárói a beszédnek: a beszélő még akkor is gesztikulál és használja mimikáját, ha a hallgató nem láthatja azokat, például, telefonos beszélgetéskor. Továbbá időnként kizárólag nemverbális jelekkel kommunikál az ember. Például a hallgató bólogatással jelzi, hogy követi a beszélő által mondottakat. A nemverbális jelek és a beszéd viszonyát tekintve ma is vitatott, hogy a beszéd és nemverbális jelek egy tőről és időben keletkeznek-e, és együtt, egyenrangú módon közvetítik a kifejezendő tartalmat, vagy a nemverbális kommunikációt alárendelt szerepet tölt be oly módon, hogy a gesztusok segítik a beszélőt a beszéd produkálásában, kiegészítik, modulálják a beszédben közölt tartalmat [8].

A továbbiakban *gesztus* alatt a test egy vagy több részének olyan, többé-kevésbé meghatározott, összehangolt mozgását értjük, amely egy közösségben jelentéssel bír, használati kontextusa körülhatárolt. E definíció egyes aspektusait majd még részletesen szemügyre vesszük, egyelőre néhány példa: felhúzott szemöldökök és tágra nyílt szemek csodálkozást, illetve kérdést fejeznek ki, vagy kiemelik a beszédben elhangzottat. A jobb kéz mutató és középső ujja kb. a váll magasságban, V alakban tartva győzelmet, helyeslést fejez ki.

Minket az emberi gesztusok egy speciális szempontból érdekelnek. Nevezetesen, virtuális embereket készítve, arra vagyunk kíváncsiak, hogy őket az emberi gesztus-repertoár mely elemeivel, és milyen módon ruházzuk fel. Egy *virtuális ember* (VE, angol megfelelői: virtual humans, embodied conversational agents) olyan, megjelenésében az emberre hasonlító számítógépes modell, mely a hétköznapiakban megszokott, természetes, emberi módon képes kommunikálni [1]. Napjainkban az egyre jobb minőségű szintetikus beszéd mellett nagy figyelmet szentelnek a nemverbális modálisítások használatának [2, 17]. Noha kérdésfelvetésünk alapvetően pragmatikusnak tűnik, az emberi nemverbális kommunikáció durván szólva illetően reprodukálása alapvető elméleti kérdéseket is felvet, különböző tudományágak terén:

1. **A nemverbális kommunikáció jelenségei** Mik a gesztusok szerepe a mindennapi kommunikációban? A beszélgető felek, illetve a környezet egyes jellemzői miként befolyásolják a nemverbális jelek használatát?
2. **Gesztusok szerepe ember-virtuális ember párbeszéde esetén** A funkciók és befolyásoló tényezők azonosak-e akkor is, ha az egyik beszélgető fél egy virtuális ember? Fontos-e, hogy egy virtuális ember nemverbálisan is tudjon kommunikálni? Egy-egy alkalmazási szerepkörben milyen gesztusokra kell hogy képes legyen a virtuális ember?
3. **Gesztusok számítógépes modellezése** Ha a fenti kérdéseket tisztáztuk, a számítógépes megvalósítás újabb, immár a gesztus formálás és mozgás részleteinek modellezésre vonatkozó morfológiai problémákat vet fel: hogyan jellemezhető egy gesztus, mik az időbeli változások jellemzői? Mennyire kötöttek ezek a jellemzők, hogyan tehető a gesztusok egyedivé? Mik a beszéd és gesztusok szinkronjának elvei?
4. **Gesztushasználat számítógépes vezérlése** A virtuális embert a fenti elvek alapján akarjuk vezérelni, lehetőleg magas szinten és a gesztusokat automatikusan előállítva. Milyen reprezentációt és számítógépes vezérlési mechanizmust használunk?
5. **Gesztusok számítógépes megjelenítése** Interaktív alkalmazások esetében elengedhetetlen, hogy elfogadható válaszdőn belül előálljon a verbális és nemverbális reakció, és a virtuális ember gesztusai simán és megfelelően időzítve jelenjenek meg a képernyőn.

A fenti kérdések megválaszolásához a társadalomtudomány – pszichológia, nyelvészet, szociális antropológia –, a számítástudomány – mesterséges intelligencia, számítógépes nyelvészet, grafika és animáció, gépi látás –, sőt időnként a művészetek – animációs film, festészet, színjátszás – művelőinek összefogása szükséges. Másrészt a számítógépes modellek, és a minden részletében kontrollálható virtuális ember kísérleti médiumként is szolgál, például pszichológusok, pszichiáterek számára ahhoz, hogy részletekbe menően feltérképezzék az emberi nemverbális kommunikáció jelenségeit.

A cikkünkben a fenti kérdéscsoportokat vesszük sorra, az utolsó, számítógépes grafikait kivéve. A 2. fejezetben először az emberi gesztikuláció jelenségeit tekintjük át. Majd arra keressük a választ, hogy mi a hatása egy virtuális ember az igazi emberéhez többé vagy kevésbé hasonló gesztikulálásának. Saját kutatásunkból idézzük a szemöldökmozgás lényegkiemelő funkciójára és a tekintet mint személyiségjegyre vonatkozó, virtuális emberrel végzett kísérleteink eredményeit. A 3. fejezetben egyetlen arckifejezés, a mosoly esetében mutatjuk be a modellezés megannyi elvi kérdését. A 4. fejezetben az általunk kifejlesztett GESTYLE nyelvre kerül sor, mely lehetővé

teszi, hogy gesztusok morfológiáját és egy virtuális ember által gesztikulációs szokásait deklaráljuk, és ennek alapján az elmondandó szövegbe szűrte, jelentést meghatározó címkéket automatikusan gesztusokra lefordítva, egyedileg gesztikuláló virtuális lényt hozunk létre.

Végül a záró fejezetben a sok nyitott kérdés közül olyanokat sorolunk fel, melyek a számítógépes nyelvészet, illetve a magyar nemverbális kommunikáció szempontjából érdekesek.

2. A nemverbális kommunikáció sajátosságai

2.1 Az emberi gesztusok jellemzői

Az emberi gesztusokat különböző szempontok szerint vizsgálhatjuk [4, 7, 14]. A legfontosabb szempont a *funkció* vagy *jelentés*: mit fejez ki egy gesztus? Egy gesztus *biológia szükségletet teljesíthet* (például a pislogás alapvetően a szemgolyó nedvesen tartását szolgálja), *tagolhatja a beszédet* (például felsorolás jelzése számokat mutató gesztusokkal, vagy ellentét jelzése a fej jobbra illetve balra döntésével vagy a jobb és bal kéz kinyitó mozdulatával), *szabályozhatja a diskurzus menetét* (például ha a szót át kívánjuk adni a partnerünknek, ezt ráemelt tekintettel tesszük, különben csak röviden pillantunk a hallgatóra, hogy időről időre ellenőrizzük, ért-e bennünket). Egy-egy gesztus, illetve a gesztikulálás módja jelezheti a beszélő *érzelmi, szellemi vagy fizikai állapotát*. Például az elégedettség mosolyban nyilvánul meg, míg a szomorúság a lefelé görbülő száj, mint arckifejezés mellett a kéz és testmozgások lassúbb és ritkább voltában is, de ez utóbbiak a fizikai fáradtság jelei is lehetnek. Egy emlék felidézését, illetve a mondandó megformálásának keresését ferdén felfelé emelt tekintet kíséri. Egy gesztus jelezheti a beszédben említett *tárgy vagy esemény bizonyos jegyeit*, mint méret, alak, hely, időbeliség, ismétlődés. Például egy hatalmas farkasról beszélve, a mesélő szeme kitágul, szemöldökét felemeli, és két kezével is jelezheti a farkas méretét. Egyes gesztusok teljes, *absztrakt fogalmakat jelölnek* (lásd a korábban említett győzelem jele). Mások *modulálják az elmondottakat*, annak bizonyosságot-bizonytalanságot, illetve különböző imperatív jelleget adva. Például egy alku során az „ezer forint” puhatolózó, kérő vagy megmásíthatatlan tény jellege nemcsak intonációban, hanem kérdő, kérő vagy közlő arckifejezésben is megnyilvánul.

A *beszéd és gesztus viszonyát tekintve*, bizonyos gesztusok *beszédtől függetlenül*, önmagukban is használhatók. A legtöbb gesztus azonban beszéddel együtt használatos. Ezek vagy redundáns módon *megerősítik, vagy kiegészítik* a beszédben foglalt információt. Speciálisan, a rámutatásnak a beszédben nem szereplő referenciák megadásában, illetve többértelmű referenciák feloldásában lehet szerepe. Például a „Nyisd ki az ablakot!” felszólításban a megszólítottat és a kinyitandó ablakot is rámutatás jelezheti a beszélő, ha több személy és ablak is szóba jön. A rámutatás nemcsak kézzel, hanem tekintettel, sőt fejbiccentéssel is történhet.

Egy másik osztályozási szempont az, hogy egy gesztus *milyen körben ismert*, illetve milyen helyzetben használatos. Vannak csak egy-egy etnikai, foglalkozási illetve társadalmi körön belül használatos gesztusok. Görögországban a tagadást fejezi ki a nyugat-európai igenlő bólintás. Egy másik példa az üdvözlés gesztusai: a partnerek

neme, kora, társadalmi helyzete, ismeretsége és nem utolsósorban etnikai hovatartozása határozza meg, hogy például a tekintettel történő biccentés és (adott számú és sorrendű) arcon csókolás közti lehetőségek közül melyiket használják.

A gesztusok *morfológiájukat tekintve unimodálisak vagy multimodálisak, egyszereiek vagyisméltlódók, statikusak illetve dinamikusak*. Statikus gesztus például a V jel mutatása, míg dinamikus a hangsúlyozó leütő kézmozdulat. Egy gesztust az arc egyes jegyeinek, illetve a kéznek a koordinált mozgása jellemez. Egy-egy gesztus többféle változatban is használatos, mely változatokat lényegében azonosként érzékelünk és értelmezzük. Pl. a V jel mutatása esetén a kézfej helye nem pontosan meghatározott, a kézforma a karakterisztikus jellemző, melyet látható magasságban kell felmutatni. Viszont a hangsúlyozásra a fejbiccentés és kézleütés két, modalitásában és morfológiájában különböző gesztus. Másrészt egyetlen gesztus többféle jelentést is kifejezhet, lásd a már említett felemelt szemöldök. Tehát a gesztusok és jelentések között *több-többértelmű a megfeleltetés*.

Ezzel eljutottunk a két utolsó, a gesztikuláló személyéhez kapcsolódó szempontig. Az, hogy valaki a morfológiailag és mozgáskarakteristikájában különböző alternatívák közül milyen gesztusokat használ, *személyiségétől is függ*. Egy nyitott, magabiztos személy nagyobb gesztikulációs teret tölt ki maga körül, több és nagyobb intenzitású gesztust használ, mint egy zárkózott ember. Továbbá a gesztikulálásban, beleértve az arc mimikáját is, mindig van több-kevesebb *egyéni jellegzetesség*.

2.2 Hogyan értelmezzük egy virtuális ember gesztusait?

Szükséges-e virtuális embereket a valódiak körében megismert gazdagságban és részletességben gesztusokkal felruházunk? Nem elegendő-e csak a szigorúan funkcionális, referenciákat feloldó vagy információt hordozó gesztusokra szorítkoznunk, és azokat egyetlen prototípus formájában használni? Az utóbbi választásra az adhatna alapot, hogy a számítógép előtt egy virtuális emberrel beszélgető felhasználó tudatában van annak, hogy nem egy másik emberrel, hanem egy többé vagy kevésbé élet-hűnek tűnő számítógépes szimulációval cserél eszmét, akitől nem is vár el gazdag és élethű gesztikulálást. Ám hogy ez nem így van, újabb és újabb kísérletek erősítik meg.

Egyrészt az ember igen érzékeny a mindennapi életben nagy változatosságban és tömegben tapasztalt gesztusokat illetőleg. Mechanikus, nem az emberi mozgásdinamikát követő, és mindannyiszor pontosan azonos formában ismétlődő mozgások unalmassá és zavaróvá válnak, és a virtuális lény elveszti az élő lény illúzióját. De nem elég csupán változatos és a mozgást tekintve emberszerű gesztikulációról gondoskodni. C. Nass és kollegái egy sereg vizsgálatot végeztek, melyek azt bizonyították, hogy a virtuális lény kommunikációját meghatározó paraméterek közül egyet megváltoztatva, más lett a virtuális lény által keltett szubjektív benyomás, és ami még fontosabb, objektív hatás [9, 10, 13]. Például, gesztusokra szorítkozva, attól függően, hogy a virtuális lény milyen (mozdulatlan) testhelyzetben jelent meg, a felhasználók többé vagy kevésbé adtak hitelt a tanácsainak. A preferencia attól függött, hogy a felhasználó maga extrovertált vagy introvertált személyiségű-e. Egy másik példa azt mutatta, hogy pusztán a szemöldököt kissé összevonva, alig észrevehetően szigorúbb kifejezést adva egy virtuális lénynek, a kísérleti alanyok jobb hatásfokkal végzik el az általa kirótt feladatot, viszont kevésbé élvezik a feladatvégzést [20]. Tehát a gesztusok részleteit is értelmezi az ember, és azok alapján (is), jóllehet öntudatlanul, reagál.

Egy virtuális lény nemverbális kommunikációját ezért igen körültekintően kell megtervezni. A nemverbális gesztusoknak összhangban kell lenniük továbbá az egyéb modalitások (kinézet, nyelvhasználat, hanghordozás) által keltett benyomásnak, hogy az összbenyomás egy konzisztens személyiség legyen.

2.3 A szemöldök szerepe kiemelésben

Ebben a fejezetben egy saját kutatási eredményt ismertetünk annak illusztrálására, hogy egy virtuális ember nemverbális kommunikációjának egyetlen részlete is miféle hatást vált ki. Azt vizsgáltuk, hogy egy beszélő fej esetében [16] a megemelt szemöldök jelzi-e a dialógus kontextusában az információ fontosságát, újdonságát (prominence) [5]. A kísérlet keretében egy beszélő fej hollandul a „blauw vierkant” (kék négyzet) jelzős szerkezetet ismételte különböző változatokban: az első, a második, illetve mindkét szót hangsúlyozva, és mindhárom esetben további 4 variációban aszerint hogy kíséri-e megemelt szemöldök az első, a második, mindkét vagy semelyik szó kiejtését. A nézőnek két kérdésre kellett válaszolnia két-két variáció megtekintése után: melyiket tartja természetesebbnek? Melyikben kap hangsúlyt a jelző, illetve a jelzett szó?

A kísérlet két tanulsággal szolgált. Egyrészt a vokális hangsúllyal egybeeső megemelt szemöldökös változatokat tekintették a nézők a legtermészetesebbnek. Ha mindkét szó hangsúlyozott volt, akkor csak az elsőt kísérő szemöldökmozgást kedvelték.

Továbbá a megemelt szemöldök igenis megerősítette a vokálisan is kiemelt szó prominenciáját, és „leárnyékolta” a megelőző vagy követő szón lévő vokális hangsúly kiemelő szerepét.

A kísérlet részletei megtalálhatók a cikkekben. Itt csak azt emeljük ki, hogy nem ismert hasonló adat igazi beszélőket tekintve, aminek egyik oka, hogy igen nehéz egy embernek úgy beszélnie, hogy vokálisan és vizuálisan más szót hangsúlyozzon. A mesterséges beszélővel kapott eredményeket tekintherjük az emberi beszédre is jellemzőnek.

2.4 Extrovertált virtuális ember tervezése

A mindennapi életben tapasztalt jelenség, hogy egy nyitott, extrovertált személyiségű embernek élénkebb a beszéde és az arc mimikája, mint egy zárkózott, introvertált személyiségé. Virtuális emberek esetében is fontos, hogy a lény konzisztens személyiségjegyeket mutasson, ezek megléte – aszerint, hogy a virtuális lény személyiség illel-e a felhasználóhoz – kedvezően vagy éppen negatívan befolyásolja a virtuális lény megítélését, és közvetett módon, az általa képviselt szolgáltatás igénybevételét vagy kirótt feladat elvégzésének sikerességét [9].

Az előzőt követő kísérletünkben arra kerestünk választ, hogy a beszédet milyen arc mimikai jellegzetességekkel kell kísérni, hogy a virtuális lény extrovertált vagy introvertált benyomást keltsen [5]. Három paraméter változtattunk egy beszélő fejen: a szintetizált beszéd, a tekintet és a szemöldökmozgás mindegyikét elkészítettük extrovertált és introvertált formában. A szemöldök introvertált esetben nem mozgott, extrovertált esetben kétszer megemelkedett két vokálisan hangsúlyozott szón. Az introvertált tekintet azt jelentette, hogy a fej 300 ms alatt lefele-balra mozgatta a sze-

mét, és csak a beszéd vége felé emelte fel a tekintetét ismét (ld. 1. ábra). Extrovertált esetben a karakter mindvégig a hallgató szemébe nézett, pislogásokkal megszakítva. A 8 lehetséges változatban jelentéssel nem bíró modern verssorokat mondatunk a.beszélő fejjel. A vokális jellemzők és a kísérlet részletei megtalálhatók a cikkben.



Fig. 1. Introvertált és extrovertált beszélő fej.

A teszt személyeknek a beszélő fejet két szempontból kellett minősíteniük: mennyire természetes, és milyen személyiségű. Az utóbbi kérdés meglepő eredménnyel szolgált. A beszélő fej már akkor is extrovertált benyomást keltett, ha a 3 jegyből kettő volt extrovertált. A megítélést csak kissé javította, ha mindhárom modalitásban extrovertált jeleket produkált a beszélő fej.

Ez az eredmény egyrészt felveti a kérdést, hogy mi ennek a jelenségnek az emberi kommunikációban és percepcióban rejlő, mélyebb oka. Másrészt elvi alapot ad arra, hogy tudatosan válasszuk ki azt a két modalitást, mellyel sikeresen kelthetjük az extrovertált személyiség benyomását. Például jó minőségű, paraméterezhető beszéd-szintetizátor hiányában (ami gyakori eset) elég az arc mimikára nagyobb gondot fordítani, hogy kompenzálódjon a szintetizált beszéd semleges vagy akár introvertált jellege.

3. Gesztusok modellezése

Ha egy virtuális lény arc vagy kéz gesztusainak megtervezéséhez kezd az ember, három alapvető problémával találja szembe magát:

1. Milyen valójában a reprodukálendő gesztus? Mik a morfológiai és dinamikai karakterisztikái? Milyen apró eltérések lehetségesek, sőt, szükségesek, hogy életszerű és nem mechanikus, robot hatást keltsenek?
2. Egy gesztust különböző helyzetekben, például más és más beszédet kísérve vagy más-más célpontra mutatva akarunk használni. Ez a morfológiai és dinamikai jellemzők paraméterezésével érhető el. Mik az alkalmas paraméterek, és ezek lehetséges értékei?
3. A gesztusokat szeretnénk alkalmazni különböző virtuális lényekre, ahelyett hogy minden esetben újból létrehozzuk a gesztus repertoárt.

A 3. probléma a szabványos virtuális lény modellek és animációk témájához vezet, mellyel itt nem foglalkozunk. Az 1. és 2. problémát alább egyetlen gesztus, a mosoly példáján tárgyaljuk.

3.1 A mosoly dinamikája és fajtái

Ahhoz, hogy modellezni tudjuk a mosolyt, igazi emberi mosolyokat kell részleteiben megfigyelni. Erre, durván, két út van: a tudományos kísérletezés és leírás, illetve a művészi absztrahálás. Mi korábbi munkánk során mindkettőt alkalmaztuk, amikor mosolyt terveztünk [3]. A tudományos módszer bevett gyakorlata szerint igazi emberek arcán előhívott mosolyt tapogattunk le, az arcon megjelölt, a mosoly során mozgó pontok koordinátáinak időfüggvénye formájában.

Amint az a 2. ábrán látható, az egyes pontok (miként az azok mozgását előidéző megfelelő arcizmok) koordináltan, többé-kevésbé szimmetrikusan mozognak. Elkülöníthető a mosoly *megjelenésének*, *megtartásának* és *eltűnésének* 3 fázisa. Noha mi is és mások is tapasztalták a szájcücskők mozgásának „kettős hegycsúcs” alakú függvénygörbéjét, mind a mai napig bevett gyakorlat, hogy a mosolyt (és a többi arkifejezést is) trapéz alakú aktivációs függvényekkel jellemzik, az arkifejezés egy ideig mozdulatlanul van jelen az arcon. Nem tisztázott, hogy az ábrán is látható görbék mik az általános jegyei, mik az egyéni jellegzetességek, és mi a mintavételi vagy a kísérlet egyéb körülményeiből adódó zaj?

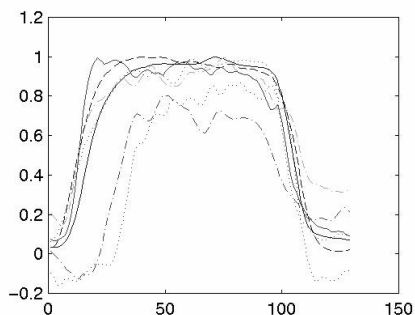


Fig. 2. Egy emberi mosoly során az arcon megjelölt pontok mozgása.

Az egyszerű, trapéz alakú aktivációs függvényt feltételezve, máris számtalan kérdést kell tisztáznunk ahhoz, hogy ilyen függvényekkel illúziókeltő mosolyt fakasszunk egy virtuális arcon. Milyen rövid, illetve hosszú lehet egy mosoly? Ha időben skálázunk egy mosolyt, hogyan változik a 3 szakasz időtartama? Mi a helyzet az intenzitást illetően, milyen az átsuhanó mosoly és széles mosoly szájmozgás függvénye? Szimmetrikusnak kell-e lennie az arc mozgásának?

Mindezek megválaszolásában a fő nehézség, hogy nincs elég, természetes helyzetben felvett mosolyról adatunk. Márpedig a spontán helyzet alap követelmény (lenne), mivel ismert, hogy a parancsra kiváltott mosoly különbözik az igazi érzelmet tükrözőtől. Egy másik probléma, hogy mi is az az érzelem, ami a mosolyt kiváltja? Az elégedettség, viszontlátás öröme, csodálat, káröröm, gúny mind mosollyal járnak – melyik ezek közül az az érzelem, amely a mosoly? Megfelel-e ez az általunk megcél-

zott esetben, amikor is a virtuális lény például eladó, tanár vagy pszichológus szerepét kell hogy betöltse?

Számunkra igen tanulságos volt, amikor az arckifejezések letapogatásának nehézségeivel szembesülve, egy grafikusművészt kértünk meg, hogy fejből tervezzen arckifejezéseket. Ő a fentiek szerint minősített mosolyok sokaságát tervezte, melyek mind a szubjektív szemlélő, mind a számítógépes felismerő rendszer számára karakterisztikusabbak és részletgazdagabbak voltak, mint a kísérleti körülmények között előhívott igazi mosolyok.

3.2 Parametrizált gesztus repertoár

Egy virtuális lény gesztus repertoárját szisztematikusan, és egy-egy alkalmazási területnek megfelelően építhetjük fel. Gesztusokat *gesztus primitívek párhuzamos és szekvenciális komponálásával* határozhatunk meg. Pl. az ijedt arckifejezés három primitív, a szemöldökemelés, szem és szájkerekítés. Egy ilyen arckifejezést numerikus vagy kvalitatív módon paraméterezhetünk (pl. intenzitás és 1 között, kissé ijedt vagy rettentően ijedt). A gesztusok időzítését rendszerint beszédhez igazítjuk. Itt merül fel, hogy milyen hosszú is legyen-lehet egy arckifejezés: egy-egy szótag, szó, vagy mondat időtartamú? Láttuk, hogy egy-egy gesztus, különösen kéz gesztusok, többféle változatban is kivitelezhető. A mozgás amplitúdója, simasága, a dinamikai jellemzői (lásd indulatos gesztikulálás) mind paraméterekkel vezéreltek kell hogy legyenek, mivel a gesztus ilyen részletei például érzelmi vagy egyéni jelleget tükröznek. Az általunk kifejlesztett, alább bemutatásra kerülő nemverbális kommunikációt deklaráló GESTYLE nyelvben párhuzamos és szekvenciális kompozícióval definiálhatók összetett arc és kéz gesztusok. A gesztus egyes fázisainak hossza, a mozgás milyenség, pontossága, arckifejezések szimmetriája mind paraméterekkel szabályozhatók [11].

4. Multimodális kommunikáció GESTYLE-ban

4.1 A vezérlés szintjei: gesztustól a stílusig

A GESTYLE nyelv egy olyan szöveg annotáló nyelv, mellyel az előbb felsorolt követelményeknek megfelelően, több szinten és több modalitást tekintve lehet előírni a beszédet kísérő (vagy helyettesítő) nemverbális jeleket [11].

1. **Gesztus repertoár definiálása** Elemi unimodális gesztusokból multimodális, összetett gesztusokat lehet definiálni. A virtuális lényt megjelenítő animáló motornak végül pontosan időzített elemi gesztusokat kell majd megjelenítenie.
2. **Gesztus könyvtár megadás** Egy-egy gesztus (egy vagy több) jelentést fejezhet ki. A jelentés-gesztus több-többértelmű, és a választások tekintetében nemdeterminisztikus hozzárendelés egy-egy könyvtárba összegyűjtve szerepel. Egy ilyen könyvtár valamilyen foglalkozási, etnikai vagy egyéb csoport körében honos gesztushasználatot rögzíti.

3. **Stílus** A virtuális lény stílusát a használatos gesztus könyvtárak, valamint a gesztikuláció specifikus jegyeinek megadásával definiáljuk. Pl. egy introvertált tanár, akire aszimmetrikus szemöldökmozgás jellemző, a tanári foglalkozás és introvertált személyiség gesztus könyvtárait fogja használni úgy, hogy az ottani, szemöldök mozgást magukba foglaló gesztusok m módosulnak.
4. **Szöveg annotálás címkékkel** A (beszédszintetizátor által majd elmondandó) szöveget magas szinten, *jelentés címkékkel* tűzdelhetjük meg, melyek értelmezése a használatra kerülő gesztus könyvtárak alapján képeződik le egy-egy gesztusra, mely időzítése a szintetizált beszédhez igazítva, automatikusan történik. Emellett *gesztus címkékkel* gesztusok közvetlenül, a beszédhez igazítva vagy abszolút időzítéssel is előírhatók.
5. **Dinamikus hatások** Az érzelmi és egyéb dinamikus címkékkel egy VL gesztikulációja dinamikusan is változtatható, a stílusban definiáltakat időlegesen felülírva. Például szomorú érzelmi állapotban, a gesztusok száma és intenzitása csökken, az arckifejezésnek, szemmozgásnak lesz nagyobb szerepe a kéz gesztusokkal szemben.

4.2 Egy példa

Az előadás során látható és hallható lesz egy virtuális lény, melyet az alább annotált szövegrészlettel vezéreltünk:

```

1 <StyledText>
2 <StyleDeclaration>
3 <OrderedElements>
4 <Style aspect="PERSONALITY" dict="Extravert" />
5 </OrderedElements>
6 </StyleDeclaration>
7 <TextBody>
8 <Meaning name = "Happy">
9 <Meaning name = "Greet"> Hello </Meaning>
10 I am a gesturing avatar.
11 </Meaning>
12 <Meaning name = "Sad">
13 Sorry for
14 <Meaning name = "EmphasizeMild"> not </Meaning>
15 being properly dressed ...
16 </Meaning>
17 Do you allow
18 <Gesture name="PointAtSelf" intensity="INTENSE"> me </Gesture>
19 in?
20 </TextBody>
21 </StyledText>

```

A fenti példában a VE személyisége az egyetlen, amely meghatározza a gesztikulációját. Csupán a 4. sorban a PERSONALITY címke megváltoztatásával introvertált személyiséggé, a VE ugyanezt a szöveg más, introvertált gesztikulációval fogja előadni (ld. 3. ábra). Megjegyezzük, hogy a GESTYLE fordító többször egymás után esetleg más-más gesztusokat rendel a jelentés címkékhez, a könyvtárban szereplő valószínűségi megfeleltetés alapján. Így pl. verbálisan adott üdvözlés többszöri ismét-

léskor többé vagy kevésbé más gesztusokkal történik, hasonlóan az emberi szokásokhoz.

Láthatjuk azt is, hogy a szöveg során a VE kedve megváltozik (8, 12 sorok). Ennek eredményeként pl. a 9 sorban az üdvözlést nagyobb amplitúdójú integető gesztus fejezi ki. A jelentés címkék mellett a 18 sorban egy intenzitás paraméterrel megadott, a szöveghez szinkronizált gesztus szerepel.

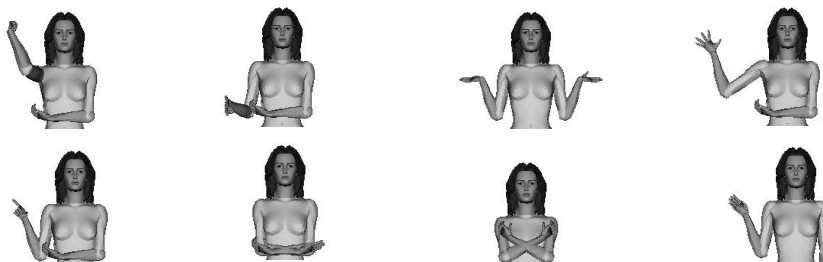


Fig. 3. Gesztusok extrovertált és introvertált személyiségű VE esetében.

6. Nyitott kérdések

Végül a sok lehetséges nyitott kérdés közül olyanokat sorolunk fel, amelyek számítógépes nyelvészek számára érdekesek lehetnek.

A fentiekben azt vázoltuk, hogy miként lehet címkézett szövegből kiindulva előállítani a szöveget kísérő gesztusokat. Kérdés, hogy egy szöveg szintaktikus és szemantikus elemzésével generálhatók-e a jelentés címkék automatikusan? Milyen módon, nyelvtannal írhatók le a címkékkel megtűzdelt magyar mondatok?

A nemverbális stílus generálását össze kéne kapcsolni a nyelvben és diskurzusban megjelenő stílussal is [19]. Legvérmesebb álmunk egy olyan, stílusában paraméterezhető VL, mely Queneau „Stilusgyakorlatok”-jának mintájára egyetlen történetet különböző nyelvi, narratív és nemverbális stílusban képes előadni [12].

Szinte semmit nem tudunk a nemverbális jelek és a magyar beszéd szinkronizálásáról. Vannak-e, és mik a magyar nyelvből adódó sajátosságok? Az idézethez hasonló kísérlet megismétlését tervezzük magyar résztvevőkkel – érdekes lenne közel azonos kísérleti környezetben összehasonlítani pl. holland és magyar alanyokkal kapott eredményeket.

Nem térünk ki arra, hogy az egyéniség, érzelmi állapot a beszédben is kell hogy tükröződjön. Vannak kezdeti kutatások arra nézve, hogy miként lehet, a gesztusokhoz hasonló módon, a szintetikus beszédet is például szomorúra vagy izgatottra hangolni. A GESTYLE nyelv elvben képes ilyen változtatásokra is [15, 18]. A Profivox TTS rendszerrel lehetne a magyar nyelvre hasonló környezetet kialakítani. A munka dandárját az érzelmes beszéd analitikus leírása és modellezése jelentené.

Végül, egy igen érdekes, hasznos és szakmailag is figyelemreméltó alkalmazás lenne különböző nyelvi, nemverbális és diskurzus stratégiákat megtapasztalni, kipróbálni programozható virtuális lényekkel. Például sokan nem is veszik észre, hogy milyen zavaró hatású lehet, ha túl kevés vagy túl sok szemkontaktust tartanak beszélgető partnerükkel. Egy virtuális lényvel beszélgetve ezt gyakorolni, tanulni lehetne.

A legutolsó kérdés elvezet a VL alkalmazásának legkritikusabb problémájához: a mai virtuális lények elég emberszerűen tudnak előadni, percepciók képességük, beszédértésük viszont igen korlátozott.

Köszönetnyilvánítás

A cikk a Magyar Felsőoktatási Minisztérium Szent-Györgyi Albert Ösztöndíjának támogatásával készült.

Irodalomjegyzék

1. Cassell, J., Sullivan, J., Prevost, S., Churchill, E. “*Embodied Conversational Agents*”, MIT Press, Cambridge, MA. 2000.
2. Chi, D., Costa M., Zhao L., Badler N. “The EMOTE Model for Effort and Shape”, *Proc. of Siggraph*, 2000. pp. 173-182.
3. Hendrix, J., Ruttkay, Zs. Exploring the space of emotional faces of subjects without acting experience, *CWI Report INS-R0013*, Amsterdam, 2000
4. Kendon, A. “Human Gesture”, In: Ingold, T. and Gibson K. (eds.) *Tools, Language and Intelligence*, Cambridge University Press, 1993.
5. Krahmer, E., S. van Buuren, Ruttkay, Zs., W. Wesselink: Audio-visual Personality Cues for Embodied Agents: An experimental evaluation, Proc. of the AAMAS03 Ws on “Embodied Conversational Characters as Individuals”, 2003, Melbourne, Australia
6. Krahmer, E., Ruttkay, Zs., Swerts, M., Wesselink, W. Audiovisual Cues to Prominence. In: Proceedings International Conference Spoken Language Processing, Denver, CO, 2002, pp. 1933-1936.
7. McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press. 1991.
8. McNeill, David (Ed.) *Language and gesture*. Cambridge: Cambridge University Press, 2000.
9. Nass C., Isbister K., Lee E-J. “Truth is Beauty: Researching Embodied Conversational Agents,” In: [1], pp. 374-402.
10. Nass, C.I., Steuer, J., Tauber, E. “Computers are Social Actors”, *Proc. of CHI'94*, Boston, MA, 1994.
11. Noot, H. Ruttkay, Zs. Variations in Gesturing and Speech by GESTYLE, International Journal of Human-Computer Studies, Special Issue on ‘Subtle Expressivity for Characters and Robots’, to appear in 2005.
12. Queneau R., Wright B. (Translator): *Exercises in Style*, New Directions 1981.
13. Reeves, B., Nass, C. *The Media Equation – How People Treat Computers, Television and New Media Like Real People and Places*, CUP, 1996.
14. Ruttkay, Zs., Pelachaud, C., Poggi, I., Noot, H. Exercises of Style for Virtual Humans, In: L. Canamero, R. Aylett (Eds.), *Animating Expressive Characters for Social Interactions*, Advances in Consciousness Research Series, John Benjamins Publishing Company, to appear.

15. Ruttkay, Zs., V. van Moppes, Noot, H. The jovial, the reserved and the robot, Proc. of the AAMAS03 Ws on “Embodied Conversational Characters as Individuals”, 15th July, 2003, Melbourne, Australia
16. Ruttkay, Zs., Noot, H. Animated CharToon Faces, *Proceedings of NPAR 2000 - First International Symposium on Non Photorealistic Animation and Rendering*, pp 91-100, June 2000.
17. Ruttkay, Zs., Pelachaud, C. (Eds.), *From Brows to Trust: Evaluating Embodied Conversational Agents*, Kluwer, 2004.
18. Van Moppes, V. Improving the Quality of Synthesized Speech Through Mark-up of Input Text with Emotions, Master Thesis, VU, Amsterdam, 2002.
19. Walker, M., Cahn, J., Whittaker, S. “Improvising Linguistic Style: Social and Affective Bases for Agent Personality”, *Proc. of Autonomous Agents Conf.* 1997.
20. Walker, J., Sproull, L., Subramani, R. “Using a Human Face in an Interface”, *Proc. of CHI'94*, pp. 85-91.