

Mens en Computer

De Menselijke Kant van Digitale media

Emiel Kraemer



Rede, in verkorte vorm uitgesproken bij de openbare aanvaarding van het ambt van hoogleraar in de Digitale Media en Organisatie aan de Universiteit van Tilburg, op 2 november 2007.

© Emiel Krahmer, 2007
Department Communicatie- en Informatiewetenschappen
Faculteit Geesteswetenschappen
Universiteit van Tilburg
Postbus 90153
5000 LE Tilburg

E: E.J.Krahmer@uvt.nl
T: 013 - 4663070
F: 013 - 4663110
U: <http://fdlwww.uvt.nl/~krahmer/>

De foto op de voorkant komt uit een New York Times Magazine feature (17 juni 2007), genaamd *Double Agents*, en toont twee on line gamers en de avatars waarmee zij zich in die games representeren. Ik wil Anita Cremers, Annemarie Krahmer, Fons Maes, Erwin Marsi en Marc Swerts bedanken voor hun commentaar op eerdere versies van deze tekst.

Voor Daan, Bas, en Coen

We don't study elementary particles because they are intrinsically interesting, like people. They are not — if you have seen one electron, you have seen them all.

STEVEN WEINBERG

The computer can't tell you the emotional story. It can give you the exact mathematical design, but what's missing is the eyebrows.

FRANK ZAPPA

Mijnheer de rector magnificus, geachte aanwezigen.

Het kan voor een kersverse hoogleraar behoorlijk ontzuenderend zijn om thuis te vertellen dat hij een nieuwe baan heeft (e.g., Vroomen 2006). Toen onze zoon Bas, destijds 9 jaar, hoorde dat zijn vader “professor” werd, vroeg hij: “Oh nee, word je dan zo iemand in een witte jurk, die in een laboratorium allerlei rare experimenten uitvoert?” Nou Bas, ik kan je geruststellen: Het is, zoals je ziet, een *zwarte* jurk.

Maar genoeg over mijn kleding. Ik wil u ook iets inhoudelijks meegeven vandaag. In het bijzonder wil ik het met u hebben over de ongelofelijke ontwikkelingen op het gebied van de computer, en hoe het juist deze ontwikkelingen zijn die het voor computers mogelijk maakt om zich meer en meer menselijk te gaan gedragen. Daarna wil ik aan de hand van enkele voorbeelden laten zien hoe het onderzoek dat we in Tilburg doen bij kan dragen aan deze ontwikkeling.

De exponentiële groei van digitale media

Ik zou, om te beginnen, eens willen vragen of iedereen die de ontwikkelingen op het gebied van computers en digitale media eenvoudig kan bijbenen nu op zou willen staan. Dat vermoedde ik al: voor de meesten onder ons zijn die ontwikkelingen nauwelijks nog bij te houden. Eindelijk een klein beetje vertrouwd met blogs, wiki's en Second Life, beginnen ze weer over vlogs, tag clouds en Web 2.0. Het gevoel dat de ontwikkelingen niet meer bij te houden zijn is echter eenvoudig te verklaren: ze *zijn* namelijk niet bij te houden; zo snel gaan de veranderingen tegenwoordig.

We leven in een digitaal tijdperk. Computers zijn zo'n integraal onderdeel van ons dagelijks leven, dat we vaak vergeten hoe snel de ontwikkelingen zijn gegaan en nog steeds doorgaan. De eerste computers zoals wij die kennen zijn tijdens en vlak na de tweede wereldoorlog ontstaan. Dit is bijvoorbeeld de ENIAC, ontwikkeld door het Amerikaanse leger om snellere berekeningen voor de artillerie mogelijk te maken (Figuur 1). De ENIAC nam een kleine fabriekshal in beslag en de computer-experts werkten niet *aan* maar *in* de computer. De ENIAC was in staat om een kleine 100 operaties per seconde uit te voeren, een getal dat volledig in het niet valt bij de miljarden operaties die een vele malen goedkopere huis-tuin-en-keuken computer anno nu uit kan voeren.

U heeft misschien wel eens gehoord van Moore's law; de voorspelling van Gordon Moore uit 1965, die grofweg stelt dat de snelheid en capaciteit van computers elke twee jaar zal verdubbelen (e.g., Schaller 1997). Dat is een exponentiële groei, en deze voorspelling blijkt tot nu toe verbluffend goed te kloppen (zie Figuur 2). Mensen hebben de neiging om de snelheid waarmee exponentiële functies groeien enigszins te onderschatten (en maken zich daardoor niet zo druk over bijvoorbeeld overbevolking). In dat verband is het instructief om het verhaal van de uitvinding van het schaakspel nog eens kort te memoriseren (e.g., Sagan 1997). Het schaakspel wordt gespeeld op een bord met 64 velden en is, zoals u misschien wel weet, lang geleden uitgevonden in het oude Perzië. Naar het schijnt was de Shah destijds zo blij met dit nieuwe spel dat hij de Groot-Vizier die het bedacht had wilde belonen. Deze had echter een heel bescheiden verzoek: hij wilde één enkele graankorrel op het eerste veld, en vervolgens steeds een verdubbeld aantal op ieder volgend veld. Dus: 2 graankorrels op het tweede veld, 4 graankorrels op het derde, 8 graankorrels op het vierde, enz. Bedenkt u nu eens hoeveel graankorrels er op het schaakbord

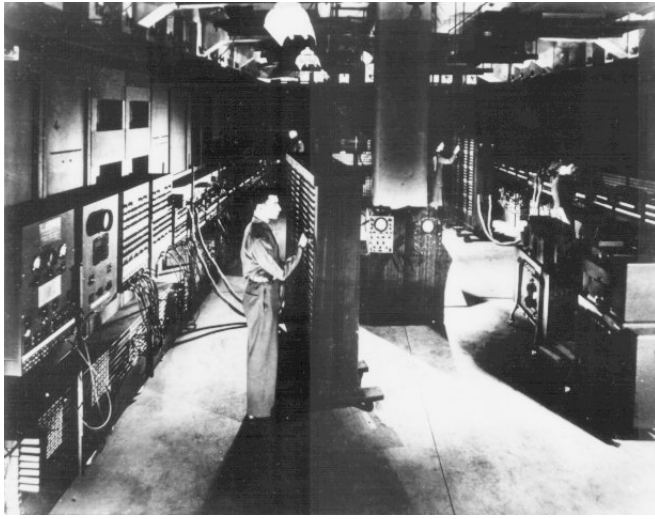


Figure 1: *Twee computers met ruim een halve eeuw ertussen: de ENIAC (links) en de iMac (rechts).*

liggen wanneer we bij het laatste, vierenzestigste veld aanbeland zijn? (Het is niet erg als u er een paar miljard naast zit, hoor.) Het is 18.6 *quintiljoen* (dat is een 18.6 gevolgd door 18 nullen). Uitgedrukt in graankorrels is dat zo'n 150 keer de jaarlijkse wereld-graanproductie van nu. Best veel dus.

Maar het is niet alleen de opslagcapaciteit en snelheid van PCs die exponentieel groeit. Hetzelfde geldt voor de embedded computers die in game consoles, digitale camera's en mp3 spelers zitten. De originele iPod, bijvoorbeeld, die in 2001 op de markt verscheen, kon 1000 nummers bevatten. De "classic" iPod die in september 2007 uitkwam kan al 40.000 nummers bevatten. Als die groei zich doorzet past de hele iTunes store (6 miljoen nummers) op de generatie van 2010 (Pogue 2007).

Veel rekenkracht voor weinig geld is op zichzelf niet voldoende om de digitale revolutie te verklaren. Cruciaal daarvoor was de opkomst in de jaren '90 van toepassingen als e-mail en vooral het world wide web. Het is nauwelijks nog voor te stellen dat het internet zoals we dat nu kennen in de eerste helft van de jaren '90 gewoon niet bestond. Niemand weet precies hoe groot het internet is, maar welke maat je ook neemt (of het nu het aantal servers, het aantal IP adressen, of het aantal weblogs is), steeds opnieuw kom je periodes van explosieve, exponentiële groei tegen. Of zoals één van mijn studentes het ooit kort en bondig formuleerde: "De internetpagina's vallen momenteel als paddestoelen uit de lucht."

Een interessante vraag is hoe al deze groeicurves zich in de toekomst zullen voortzetten. Deze vraag leidt vaak tot enigszins hijgerige voorspellingen, zoals bijvoorbeeld die van Ray Kurzweil, die aanneemt dat het moment waarop de rekenkracht van computers die van het menselijk brein voorbijstreeft niet ver meer is. Volgens hem zal het kopen van "één menseli-

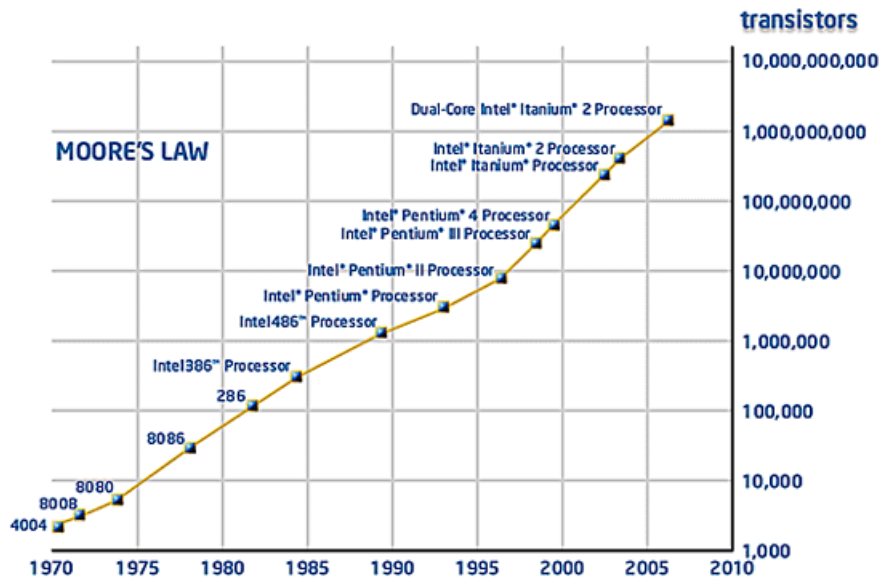


Figure 2: Moore's law: rekenkracht (in termen van transistors) als functie van de tijd. Nota bene: de Y-as geeft een logaritmische schaal, waardoor de exponentiële groei als een min of meer lineaire lijn weergegeven kan worden.

jke breincapaciteit" in 2023 nog \$ 1000 kosten (wat met de huidige stand van de dollar al een koopje genoemd mag worden), in 2037 koop je zo'n zelfde eenheid voor 1 cent. En het internet? Volgens Kurzweil zal het bezoeken van een website in 2030 inhouden dat je je volledig onderdompelt in een virtual reality omgeving (Kurzweil 2005). Of dat allemaal uit gaat komen valt nog te bezien, al heeft Kurzweil wel een naam hoog te houden als digitale voorspeller.

De menselijke kant van digitale media

Het is glashelder dat de ontwikkelingen op het gebied van digitale media de manier waarop mensen communiceren drastisch hebben veranderd. Niet iedereen is blij met deze ontwikkeling; communicatie via het internet (bijvoorbeeld MSNen) zou tot minder diepgravende contacten en zelfs tot sociale isolementen kunnen leiden (Locke 1998, Kraut et al. 1998), hoewel meer recent onderzoek (e.g., Valkenburg & Peter 2007) dat beeld behoorlijk nuanceert. Maar volgens velen gaat er uiteindelijk weinig boven de meest natuurlijke vorm van communicatie tussen mensen: een face-to-face interactie tussen twee of meer personen die elkaar kunnen zien en horen, en die zowel verbaal als non-verbaal communiceren (e.g., Garrod & Pickering 2004). Het is ironisch dat precies de exponentiële toename in computer-rekenkracht (Moore's law) het nu mogelijk maakt om computertoepassingen te ontwikkelen die deze face-to-face interactie proberen te simuleren. Ik heb het dan over computerpersonages die er uitzien als een "echt" mens, en die verbaal en non-verbaal communiceren met de gebruiker

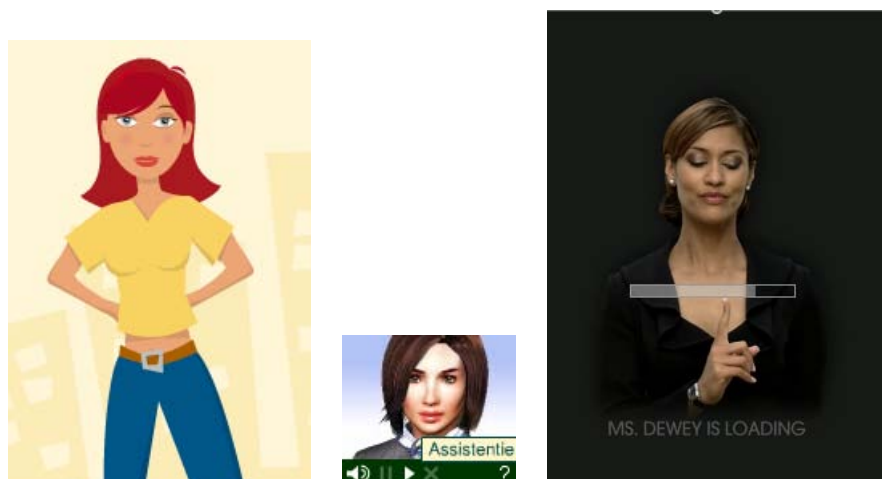


Figure 3: *Drie voorbeelden van virtuele assistentes. Van links-naar-rechts: Steffie, de assistente van de Rabobank, de naamloze assistente op de ABN-AMRO website, en Ms. Dewey, die commentaar geeft op zoekvragen bij LiveSearch.*

Bedrijven en organisaties zijn om uiteenlopende redenen in deze vorm van interactie geïnteresseerd, waarbij de overkoepelende reden vaak het bevorderen van het contact met klanten is. De assistente van de Rabobank bijvoorbeeld, Steffie heet ze, is speciaal in het leven geroepen voor mensen die de nieuwe media moeilijk bij kunnen benen. Voor deze mensen, ik citeer nu, “kunnen alledaagse handelingen als internetten, pinnen, mailen of een treinkaartje kopen bij de automaat behoorlijk frustrerend zijn. Om die frustratie te verminderen legt Steffie met animaties en spraak stap voor stap uit hoe één en ander werkt.” De ABN Amro heeft ook een assistente op haar website, die klanten quasi-leuk toespreekt en onder meer haar eigen knoppen uitlegt (“Het kruisje is mijn minst favoriete optie. Hiermee zet u mij definitief uit.”). LiveSearch van Windows heeft tegenwoordig ook een virtuele assistente, Ms. Dewey, die snedig commentaar geeft op de zoekvragen van gebruikers (Figuur 3).

Naast dit soort assistentes, altijd vrouwelijke personages trouwens, zie je ook in de wereld van computerspelletjes een duidelijk trend naar virtuele werelden waarin gamers het “echte leven” kunnen ontvluchten door het na te spelen op de computer. In de populaire Sims games draait het er bijvoorbeeld vooral om dat je als gamer de keuken regelmatig schoon houdt, en op tijd naar je werk gaat. Anders kun je die felbegeerde promotie natuurlijk op je buik schrijven. Nog een stapje verder gaan spellen als World of Warcraft, Second Life, of Home (de Second Life clone voor de Playstation 3) waarin mensen met elkaar communiceren via digitale computersonages, de zogenaamde “avatars” (Figuur 4).

Inmiddels zien mensen ook het nut in van virtuele personages voor “serious gaming”, een hippe term voor spelenderwijs leren. Bijvoorbeeld leren van een virtuele docent (“pedagogical agents”, Lewis-Johnson et al. 1999), zie Figuur 5. Denk aan al die kinderen (en volwassenen) die momenteel hun brein trainen op de Nintendo DS, aangemoedigd door de enthousiaste



Figure 4: *Virtuele karakters in een spel-omgeving: links voorbeelden van Sims karakters, rechts een screen shot van het avatar-ontwerppaneel in Second Life.*

tips van Professor Kawashima. Ook het Amerikaanse leger heeft het nut van dit soort virtuele leeromgevingen ontdekt. Om officieren-in-opleiding voor te bereiden op lastige situaties hebben ze de zogenaamde Mission Rehearsal Exercise ontwikkeld. Hierbij gaat de trainee een virtuele wereld binnen, om daar —in overleg met verschillende virtuele personages— een crisissituatie op te lossen (Gratch et al. 2002, Rickel et al. 2002). Het nieuwste virtuele personage dat in deze context ontwikkeld werd is Hassan, een Iraakees uiteraard, die de waarheid spreekt of liegt, afhankelijk van de manier waarop hij ondervraagd wordt (Traum et al. 2007).

In de meeste toepassingen die ik hierboven genoemd heb wordt het computer-personage bestuurd door een echt mens (de Second Life avatars, bijvoorbeeld) of verloopt de interactie via een vooraf bepaald script, zodat van een echte, natuurlijke interactie nog geen sprake is. In beide gevallen is het virtueel karakter niet *zelf* in staat om te communiceren met de gebruiker. Virtuele karakters die dat wel kunnen —die in staat zijn om de verbale en non-verbale communicatie van “echte” mensen te begrijpen, en daar weer op een adequate wijze verbaal en non-verbaal op kunnen reageren— worden om voor de hand liggende redenen ook wel “virtuele mensen” (Gratch et al. 2002), “life-like characters” (Prendinger & Ishizuka 2004) of “embodied conversational agents” (Cassell et al. 2000) genoemd.

Hoe bouw ik een virtueel personage?

Vanuit een technisch oogpunt kun je dit soort “virtuele mensen” zien als een combinatie van “real time computer graphics”, kunstmatige intelligentie en taal- en spraaktechnologie. Aan de ene kant is technologie nodig om automatisch te kunnen verstaan en te begrijpen wat de gebruiker zegt (spraakherkenning en taalanalyse) en ook om gebaren en gezichtsuitdrukkingen van de gebruiker te kunnen herkennen en interpreteren. Aan de andere kant moet een “virtueel mens” zelf ook weer gesproken taal kunnen produceren, met bijpassende non-verbale signalen, en er verder een beetje fatsoenlijk uitzien. Alle bouwstenen die nodig zijn om een echt virtueel karakter te maken zijn in rudimentaire vorm aanwezig (e.g., Lester et al. 2004). Hoewel nog niet alle basistechnologie goed genoeg is, werken onderzoekers sinds de tweede helft van de



Figure 5: *Toepassingen van virtuele karakters in uiteenlopende educatieve toepassingen: links Dr. Kawashima uit de Nintendo DS Brain Training, rechts een screen shot uit de eerste Mission Rehearsal Exercise.*

jaren negentig serieus aan de ontwikkeling van dit soort virtuele mensen. En dit roept meteen allerlei basale vragen op, waarbij ik me hier gemakshalve beperk tot —letterlijk en figuurlijk— de hoofdzaken.

Wat moet de mond bijvoorbeeld doen wanneer het personage praat? Meebewegen uiteraard, maar hoe? Een taal als het Nederlands kent ongeveer 40 elementaire klanken (of fonemen), als /a/, /o/ en /ü/. Dat zijn er meer dan er letters in het alfabet zitten, maar sommige letters kunnen we op meerdere manieren uitspreken. De visuele tegenhangers van fonemen worden wel visemen genoemd; “elementaire mondposities”. Daarvan kent het Nederlands er ongeveer een dozijn. Er zijn dus minder visemen dan fonemen, en dat komt omdat sommige hoorbare klankverschillen niet of nauwelijks te zien zijn; de /o/ en de /ü/ klinken anders, maar zien er hetzelfde uit. Visuele spraak kan nu gemaakt worden door met een bepaalde frequentie (bijvoorbeeld 25 keer per seconde) te bepalen welk foneem er op dat moment uitgesproken wordt, en het bijbehorende viseme te laten zien (Figuur 6). Dat ziet er nog niet helemaal natuurlijk uit, omdat de visuele realisatie van klanken context-afhankelijk is (spreekt U maar eens zachtjes “stoel” en “steen” uit, en voel dat bij de /s/ uw lippen reeds rekening houden met het feit dat ze straks een /oe/ of een /ee/ gaan produceren). Dit is gelukkig eenvoudig op te lossen door de associatie van klanken met lipposities afhankelijk te maken van de omringende klanken.

Goed, nu hebben we een virtueel personage dat zijn of haar lippen mee kan bewegen met de spraak. Dat is mooi, maar als het personage verder niet beweegt, ziet dat er buitengewoon onnatuurlijk uit. Vandaar dat allerlei trucs worden uitgehaald om de suggestie van leven te wekken. Hele kleine, willekeurige bewegingen kunnen daarbij al heel effectief zijn, zoals bleek uit het onderzoek van Kenneth Perlin (hoogleraar bij het Media Research Lab in New York, die voor dit onderzoek een Oscar heeft gekregen). Maar voor de levendigheid van een virtueel personage is dit niet genoeg. Zo’n intens starende blik bijvoorbeeld, is —net als in het echte leven— erg onprettig om naar te kijken. De meeste mensen knipperen zo gemiddeld eens per 4.8 seconde met de oogleden, om op die manier de oogbol prettig vochtig en schoon te houden, als een soort ruitenwissers voor de ogen (e.g., Karson, 1983). De meeste virtuele karakters sluiten de ogen daarom ook grofweg eens in de 5 secondes, hoewel daar natuurlijk geen enkele fysiologische noodzakelijkheid voor is. Het wachten is nog op een virtueel personage dat realistisch



Figure 6: *Twee visemen. Het linker-visem correspondeert met de /a/ klank (als in 'slaap'), het rechtervisem zowel met de /o/ ('loop') als met de /ii/ ('fuut').*

ademhaalt, maar met de potentiële levendigheid van virtuele karakters zit het dus wel goed.

De vraag is natuurlijk hoe relevant dit is voor de communicatie. Grofweg kun je stellen dat veel onderzoek tot nu toe gericht was op het creëren van een realistische buitenkant voor virtuele mensen, maar dat er weinig gedaan is aan de binnenkant (wat gaat er om in dat virtuele hoofd?) en aan hoe de buitenkant zichtbaar kan maken wat er aan de binnenkant gebeurt. Toch lijken dit essentiële voorwaarden voor een succesvolle toepassing van virtuele personages. We weten inmiddels dat een virtueel karakter alleen gewaardeerd wordt door gebruikers als het op een functionele, maar ook realistische manier communiceert met gebruikers, zowel verbaal als non-verbaal (e.g., Dehn & van Mulken 2000).

Hoe dat moet is precies wat Marc Swerts en ik in ons gezamenlijk werk proberen te onderzoeken. Het eerste dat we op dit gebied hebben geprobeerd was te bepalen hoe en vooral wanneer een virtueel karakter de wenkbrauwen zou moeten bewegen. In het pionierswerk van Ekman (Ekman 1979) werd al gesteld dat wenkbrauwbewegingen een communicatieve functie hebben, en bijvoorbeeld door een spreker gebruikt worden om belangrijke woorden te benadrukken. Maar zodra je dit soort ideeën in een virtueel personage wilt realiseren stuit je op een probleem. Wat zijn de belangrijke woorden? En hoeveel belangrijke woorden zitten er in een zin? Als een virtueel karakter bij te weinig woorden de wenkbrauwen beweegt is dat weinig expressief en erg onrealistisch. Als het bij te veel woorden gebeurt komt het karakter buitengewoon neurotisch over. En dat is ook weer niet goed (Krahmer & Swerts 2004). De enige oplossing bleek te zijn om nog maar eens goed naar "echte" mensen te kijken, en Marc zal daar straks een aantal voorbeelden van laten zien. Wenkbrauwbewegingen hebben niet alleen een communicatieve functie, ze zijn ook een belangrijk onderdeel van emotionele gezichtsuitdrukkingen (omhoog bij verbazing, omlaag bij woede), en daar wil ik het nu graag in meer detail over hebben.

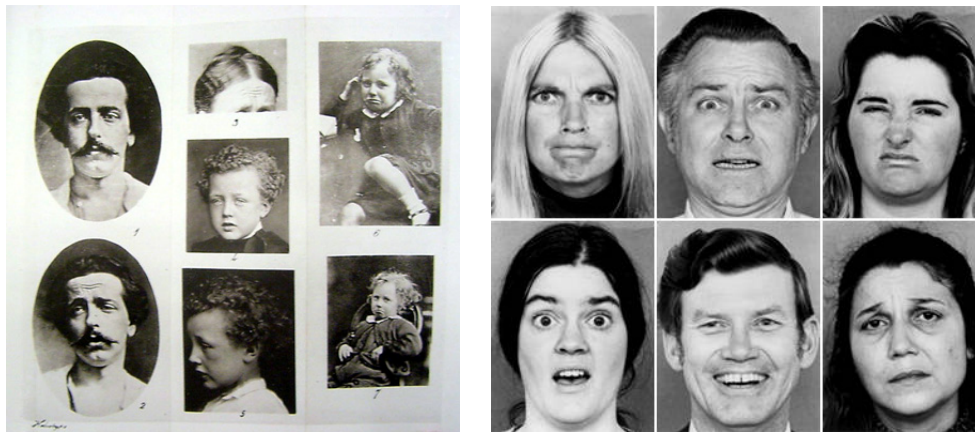


Figure 7: Links: een bladzijde uit Darwin (1872). Rechts: de zes basis-expressies uit het werk van Ekman (1972): met de klok mee: woede, angst, walging, verdriet, blijdschap, verbazing. Merk op dat het bij alle foto's van volwassen om geposeerde expressies gaat.

De emoties van echte en virtuele mensen

Over het algemeen wordt aangenomen dat het belangrijk is dat een virtueel karakter in staat is om de emoties van de menselijke gebruiker te herkennen, en daar op een passende manier kan reageren (e.g., Brace et al. 2005, McQuiggan & Lester 2007, Prendinger & Ishizuka 2005). Denk bijvoorbeeld aan de virtuele docenten, waarvoor het belangrijk is om te kunnen zien of de leerling verveeld raakt (omdat de opdrachten te makkelijk zijn) of gefrustreerd (omdat de opdrachten juist te moeilijk zijn). Daarnaast kan het van belang zijn om te kunnen laten zien wanneer de virtuele docent tevreden is, of juist teleurgesteld is in de leerling. Om deze beide aspecten in een virtueel personage te implementeren, is het dus zowel van belang om automatisch emotionele gezichtsuitdrukkingen van gebruikers te kunnen interpreteren, als om realistische emotionele expressies te kunnen tonen op het virtuele gezicht.

Over emotionele gezichtsuitdrukkingen is al veel bekend. Het moderne onderzoek naar gezichtsuitdrukkingen bij emoties heeft een lange traditie, en gaat tenminste terug op het werk van Darwin (1872), die de emotionele expressies van mensen met die van dieren vergeleek (Figuur 7). Onderzoek in deze traditie richt zich vaak op de zogenaamde basisemoties: woede, angst, walging, verdriet, blijdschap en verbazing (e.g., Ekman 1972, Adolphs 2002). Experimenten hebben keer op keer laten zien dat mensen goed in staat zijn om gezichtsuitdrukkingen van deze basisemoties bij anderen te herkennen, ongeacht de cultuur, leeftijd of achtergrond van de observatoren (e.g., Ekman 1972, Elfenbein en Ambady 2002, Russell 1994, Schmidt & Cohn 2002). Het zal toch zeker geen probleem zijn om dat in een virtueel personage in te bouwen, hoor ik u denken? Iemand die lacht is blij, toch? De werkelijkheid is een stuk complexer...

Als je vanuit het perspectief van virtuele mensen naar de uitgebreide literatuur over emotionele gezichtsuitdrukkingen kijkt vallen een paar dingen op. Ten eerste wordt het onderzoek bijna altijd uitgevoerd op basis van foto's (e.g., Adolphs 2002). In onze dagelijkse interacties



Figure 8: *Representatieve fragmenten van ongeacteerde (boven) en geacteerde (onder) expressies van emoties, met links de positieve en recht de negatieve versies.*

vormen we natuurlijk geen oordeel over iemands emotionele toestand op basis van een statische momentopname, maar op basis van veranderende en bewegende signalen. Daarnaast zou het wel eens zo kunnen zijn dat voor de interactie met virtuele personages de sociale emoties zoals frustratie of onzekerheid belangrijker zijn dan basisemoties als angst en walging (e.g., Barkhuysen et al. 2005, Kraemer & Swerts 2005, Swerts & Kraemer 2005). Tenslotte blijkt het eerdere onderzoek bijna altijd te gaan over geacteerde visuele expressies. Al dan niet professionele acteurs wordt gevraagd om een bepaalde emotie uit te drukken, en vervolgens worden de meest duidelijke gezichtsuitdrukkingen geselecteerd, en weer als de stereotiepe expressies gebruikt in verdere experimenten. Een methode met een dubieuze ecologische validiteit (Russell et al. 2003), om het maar eens voorzichtig uit te drukken.

We weten dus veel van de visuele expressie van dit soort statische, stereotiepe en geacteerde basisemoties, maar hoe representatief zijn die? Kunnen we die gebruiken voor de ontwikkeling van virtuele personages? We weten feitelijk niet wat het verband is tussen acteren —doen alsof je een bepaalde emotie voelt— en die emotie daadwerkelijk voelen. Niet dat daar geen ideeën over zijn. William James dacht bijvoorbeeld dat het net doen alsof je een bepaalde emotie

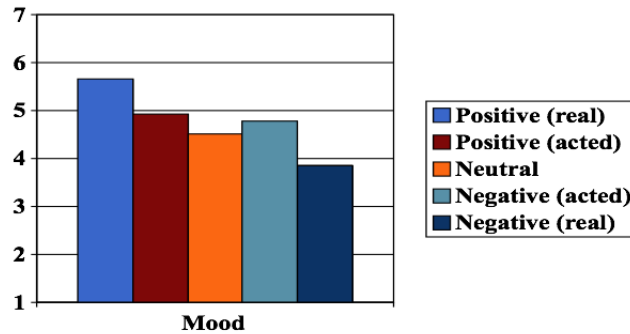


Figure 9: Zelf-gerapporteerde emotionele toestand als functie van de 5 condities.

voelt er toe leidt dat je die emotie inderdaad gaat voelen; als je de hele dag zit te mokken en te zuchten, en op alles met een sombere stem antwoordt voel je je aan het einde van de dag vermoedelijk behoorlijk gedeprimeerd (James 1884, ook Stepper & Strack 1993). Omgekeerd is er ook evidentie dat mensen die continu moeten glimlachen voor hun werk (stewardessen, hamburgerverkopers) desondanks niet vrolijk worden (Kotchemidova 2005). Daarnaast is het vooral onduidelijk in welke mate geacteerde emoties er hetzelfde uitzien als echte emoties (al zal de gemiddelde GTST kijker zich daar weinig illusies over maken).

Voor we een virtueel personage van een emotionele intelligentie kunnen voorzien, is het dus zaak om eerst nog maar eens goed naar echte mensen te kijken. In dat verband hebben we recent een serie experimenten uitgevoerd, waarin we echte en geacteerde emoties systematisch met elkaar vergelijken (Krahmer & Swerts 2007b, Wiling et al. 2006). We hebben hierbij gebruik gemaakt van de Velten techniek, een methode om emoties op te wekken onder experimentele condities (Velten 1968). Deze techniek heeft haar oorsprong in de cognitieve gedragstherapie, en is gebaseerd op het voorlezen van zinnen met een sterke emotionele lading, en daarmee zeer geschikt voor het bestuderen van audiovisuele expressies van emoties. De oorspronkelijke Velten techniek omvat drie condities, één met neutrale zinnen (*99.1% of Alaska is owned by the federal government*), één met negatieve, deprimerende zinnen (*I want to go to sleep and never wake up*) en één met positieve, vrolijke zinnen (*God I feel great!*). Van deze test hebben we een aangepaste Nederlandse versie gemaakt, die we hebben gebruikt in onze experimenten. Proefpersonen werd gevraagd om plaats te nemen voor een computermonitor waarop de zinnen één voor één verschenen. Ze lazen deze zinnen eerst in zichzelf en vervolgens hardop voor, terwijl een camera hun spraak en gezichtsuitdrukkingen registreerde. Aan de drie condities uit de originele Velten methode hebben we twee “acteer”-condities toegevoegd. In één daarvan wordt proefpersonen gevraagd om de negatieve zinnen uit te spreken, en daarbij te acteren dat ze in een positieve stemming verkeren. In de andere conditie werd proefpersonen juist gevraagd om de positieve zinnen uit te spreken en daarbij te acteren dat ze in een negatieve stemming waren. Figuur 8 laat enkele representatieve plaatjes uit de verschillende condities zien.

Na afloop moesten alle proefpersonen een korte vragenlijst invullen, waarin ze aangaven hoe ze zich op dat moment voelden en de resultaten daarvan zijn te zien in Figuur 9. Het blijkt dat mensen die niet acteren en de positieve zinnen voorlezen zich het vrolijkst voelen,

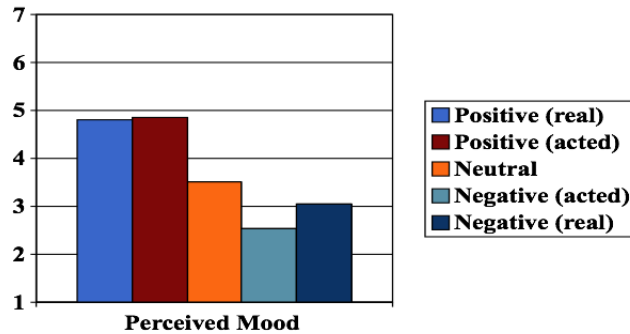


Figure 10: *Waargenomen emotionele toestand als functie van de 5 condities.*

terwijl mensen die de negatieve zinnen voorlezen zich na afloop het minst vrolijk voelden. De mensen die neutrale zinnen lazen voelden zich inderdaad het meest neutraal. Interessant is dat de acteurs zich gemiddeld genomen ook neutraal voelden, ongeacht of ze een positieve of negatieve stemming acteerden. Met andere woorden: de proefpersonen die acteren voelden de geacteerde emotie niet, in tegenspraak met de suggestie van William James dus (vgl. Konijn 2000).

De volgende interessante vraag is hoe die emotionele expressies overkomen bij anderen. Daarvoor hebben we een zogenaamd perceptie-experiment opgezet waarin we van alle proefpersonen een stukje film hebben geselecteerd die we door andere proefpersonen lieten beoordelen. Uit die filmpjes hebben we de spraak weggefilterd, zodat de proefpersonen zich alleen op de gezichtsuitdrukkingen konden concentreren. Proefpersonen kregen hierbij de vraag: hoe positief of negatief denk je dat de spreker in dit filmpje zich voelt. Figuur 10 geeft de gemiddelde scores voor de sprekers in elk van de vijf condities. Wat opvalt in de resultaten is dat proefpersonen die acteren dat ze in een positieve stemming verkeren positiever worden ingeschat dan mensen die echt in een positieve stemming zijn. En bovendien: proefpersonen die doen alsof ze in een negatieve stemming zijn worden behoorlijk negatiever ingeschat dan mensen die zich echt negatief voelen.

De conclusie lijkt duidelijk: geacteerde emoties worden niet gevoeld door de persoon die acteert, maar ze komen sterker over dan echte, doorvoelde emoties. Als we dit onderzoek ergens presenteren zeggen mensen (en dan met name reviewers) vaak dat deze resultaten mogelijk verklaard kunnen worden uit het feit dat we geen echte acteurs gebruikt hebben. Professionele acteurs zouden dat natuurlijk veel beter en veel realistischer doen. Om dat te testen hebben we het onderzoek onlangs opnieuw gedaan, maar nu met professionele acteurs, allemaal met meerdere jaren praktijkervaring. Dat ziet er dan zo uit (Figuur 11). Vervolgens hebben we de net beschreven perceptieproef herhaald, maar nu met opnames van de professionele acteurs erbij. Hieruit bleek dat de gezichtsexpressies van de professionele acteurs nog extremer worden ingeschat dan die van onze niet-professionele proefpersonen. Het ligt dus niet aan de acteervaardigheden van onze proefpersonen. Inmiddels hebben we dit onderzoek ook voor andere culturen (met name Zuid-Aziatisch) en met kinderen herhaald, en steeds opnieuw krijgen we hetzelfde resultaat: geacteerde emoties worden niet gevoeld, maar komen extremer over dan natuurlijke emoties.

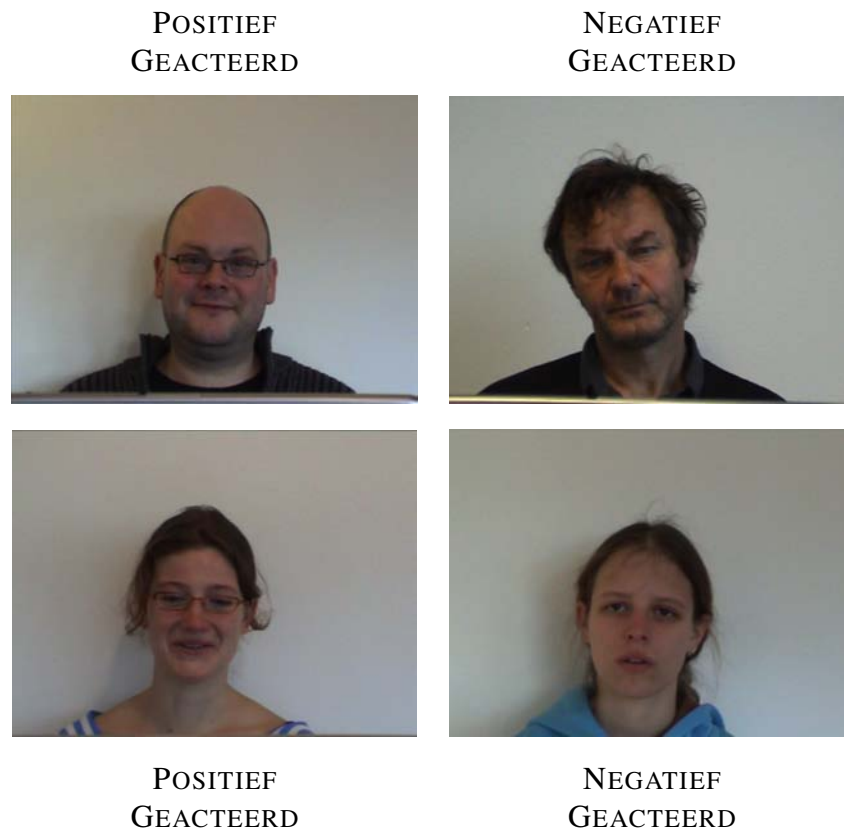


Figure 11: *Representatieve fragmenten van mannelijke (boven) en vrouwelijk (onder) ervaren acteurs, met links de positieve en rechts de negatieve versies.*

Wat betekent dit nu voor de ontwikkeling van virtuele karakters? Het betekent in ieder geval dat eerdere inzichten over het tonen en herkennen van emotionele gezichtsuitdrukkingen niet zonder meer gebruikt kunnen worden; kennis over geacteerde emoties zal waarschijnlijk niet erg helpen bij het detecteren van de emotie van de gebruiker (want die zal minder expressief zijn dan de acteurs), en kan verkeerde associaties op wekken bij het tonen van emoties door een virtueel karakter (want hoewel de stereotiepe expressies misschien wel duidelijk zullen zijn, komen ze waarschijnlijk karikaturaal over).

Virtuele mensen met inhoud

Even recapituleren: we zijn begonnen met virtuele personage die er “levend” uitzagen (met bewegend lippen, knipperende ogen, etc.) en hebben het toen over personages gehad die met emotionele expressies om kunnen gaan. Nu is het natuurlijk leuk om te kunnen communiceren met een virtueel personage dat emoties kan tonen en jouw emoties kan herkennen, maar het zou nog leuker zijn als het personage ook iets te zeggen had. De communicatieve vaardigheden van de huidige generatie virtuele personages zijn echter zeer minimaal. In de toekomst ligt het voor



Figure 12: *Een eenvoudige voorbeeld-scène.*

de hand dat virtuele mensen op een meer flexibele en gepersonaliseerde manier zullen communiceren met individuele gebruikers, en daarvoor zullen de huidige vastgebakken methodes niet langer volstaan.

Op dat moment ontstaat behoefte aan technieken die in staat zijn om informatie op een flexibele manier in gesproken taal uit te drukken, ondersteund door bijpassende gebaren. Aan dit soort technieken, ook wel automatische taalgeneratie genoemd, wordt door computerlinguïsten al een jaar of 20 gewerkt. Taalgeneratie wordt bijvoorbeeld gebruikt om automatisch weerberichten te produceren op basis van numerieke data (over de te verwachten windsterkte en -richting), of om medische informatie op een patiënt-vriendelijke manier te presenteren (Reiter & Dale 2000).

Een onderdeel van taalgeneratie dat momenteel sterk in de belangstelling staat is het verwijzen naar objecten of personen (e.g., Belz & Gatt 2007, Dale & Reiter 1995, van Deemter 2002, Kraemer and Theune 2002, Siddhartan & Copestake 2004). Neem bijvoorbeeld de volgende eenvoudige scène met drie personen (Figuur 12), en stel je voor dat je iemand op één van die personen wilt attenderen. Dat kan op veel manieren, maar de meeste sprekers hebben er geen enkele moeite mee om snel te besluiten welke informatie ze daarvoor nodig hebben, en hoe ze die informatie in taal kunnen uitdrukken met zogenaamde verwijzende expressies, zoals “de man met de stropdas”, “het meisje”, of “de jongere man”. De kunst is om een verwijzende expressie te vinden die precies past op een bepaald doelobject, en niet op één van de andere objecten in de scène (de zogenaamde afleiders). Dit is primair een keuze-probleem: er zijn veel verschillende manieren waarop je bijvoorbeeld naar een persoon kunt refereren: in plaats van “de man met de stropdas” hadden we ook kunnen verwijzen met, bijvoorbeeld, “de man met de bril”, “de wat oudere man”, “de man links”, “de man in pak”, of zelfs (onder Apple-gebruikers) “de PC”.

Stel nu dat een virtueel personage aan het communiceren is met een gebruiker over onze voorbeeld-scène (een eenvoudige en misschien niet zo interessante scène, maar daar kom ik dadelijk op terug). Laten we aannemen dat ons virtueel personage weet dat er drie personen

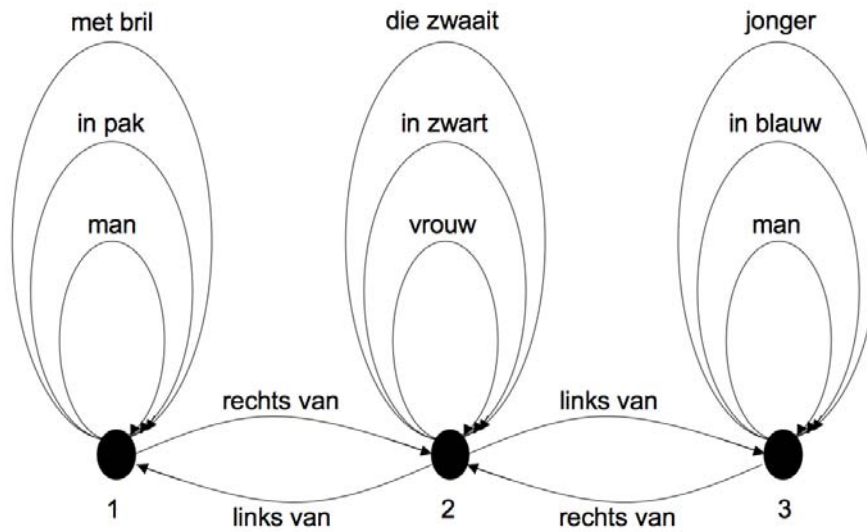


Figure 13: *Representatie van de informatie in onze voorbeeld-scène met behulp van een graaf-structuur (niet alle mogelijke lijnen zijn getekend om de figuur overzichtelijk te houden).*

in de scène voorkomen, en weet wat alle eigenschappen van die personen zijn. Hoe kan ons virtuele personage dan bepalen wat een natuurlijke manier is om naar één van die personen te refereren?

Samen met twee studenten, Sebastiaan van Erk en André Verleg, heb ik een methode ontwikkeld die dit automatisch doet (Krahmer et al. 2003). Deze methode is gebaseerd op *graf*en (e.g., Wilson 1996), en een graaf is hier een wiskundige structuur die bestaat uit een verzameling punten (of knopen), die met elkaar verbonden kunnen zijn door lijnen (of zijden). Het basis-idee is heel simpel: elk object in een scène correspondeert met een punt, en elke eigenschap van een object correspondeert met een lijn. Onze voorbeeld-scène kun je als volgt in een graaf weergeven (zie Figuur 13): er zijn 3 punten, eentje voor elke figuur in onze scène. Daarnaast ziet u twee soorten lijnen. Allereerst lijnen die beginnen en eindigen in dezelfde punt (lussen); die drukken eigenschappen uit van de personen (bijvoorbeeld dat persoon 1 een bril draagt). En daarnaast lijnen tussen verschillende punten, die relaties tussen personen uitdrukken (bijvoorbeeld dat de enige vrouw, persoon 2, zowel links als rechts naast zich iemand anders heeft staan). Ons virtueel personage hoeft nu alleen maar uit te rekenen welk deel van de graaf de persoon waarnaar hij wil verwijzen uniek karakteriseert.

Wat hebben we nu gewonnen door informatie over onze scène in de vorm van een graaf uit te drukken? Wel, het blijkt dat het vinden van zo'n unieke graaf overeenkomt met een wiskundig probleem (het vinden van zogenaamde subgraaf-isomorfismen) waar vele efficiënte oplossingen voor beschikbaar zijn. We lossen het ene probleem (het vinden van verwijzende expressies) dus op door het te vertalen naar een probleem in grafen-theorie dat al opgelost is!

Er is nog één extra ingrediënt nodig: wie goed heeft opgelet, heeft begrepen dat er in het algemeen meerdere unieke oplossingen gevonden zullen worden. Eén manier om dit keuzeprobleem op te lossen is door te kiezen voor de kleinste graaf die ons doel dient. Op die manier produceert het virtueel personage altijd een minimale, unieke beschrijving voor een object (mits die bestaat), wat in overeenstemming is met de regel van Grice (1975) dat sprekers precies voldoende informatie geven. Op deze manier is het mogelijk om automatisch allerlei soorten verwijzende expressies te produceren (van Deemter & Krahmer 2007), zelfs verwijzende expressies die deels uit taal en deels uit gebaren bestaan (van der Sluis & Krahmer 2007).

Er is eigenlijk maar één probleem, en dat is gerelateerd aan de eenvoud van de voorbeeldscène. De echte wereld ziet er namelijk een stuk complexer uit. Kijk maar eens om u heen. Overall zitten personen, en u zou uw buurman of -vrouw op elk van de hier aanwezigen kunnen attenderen middels een verwijzende expressie (“heb je dat leuke jongetje met de krullen en de bril vooraan gezien?”). In termen van grafen hebben we het dan toch gauw over een paar honderd punten, en een veelvoud daarvan aan lijnen. Het aantal mogelijke combinaties van eigenschappen dat bekeken moet worden groeit exponentieel als functie van het aantal punten en lijnen in de graaf-representatie. Dat dit echt een probleem is weet u inmiddels; mensen hebben weliswaar de neiging om de kracht van dergelijke exponentiële groeifuncties te onderschatten, maar u niet meer. In ieder geval kan het automatisch vinden van een minimale verwijzende expressie binnen zo’n groot, realistisch domein erg lang duren.

Een virtueel personage kan ook in dit verband dus weer veel van u leren. Mensen hebben doorgaans geen enkele moeite met het snel produceren van verwijzende expressies, terwijl uit de computer-analyses blijkt dat het proces in theorie heel complex is. Hoe doet u dat nu zo snel? Onderzoekers realiseren zich sinds een paar jaar dat de manier waarop mensen verwijzende expressies produceren anders is dan de manier waarop dit tot nu toe in computersystemen wordt gedaan. Een opvallende eigenschap is dat mensen vaak meer informatie geven dan strikt genomen noodzakelijk is. Ze zeggen dingen als: “het leuke jongetje met de krullen en de bril vooraan”, terwijl er toch maar één leuk jongetje met krullen hier is . . . Mensen produceren dus geen minimale beschrijvingen, maar voegen informatie toe die strikt genomen overbodig is (Maes et al. 2004, Engelhardt et al. 2007, van der Sluis & Krahmer 2007). Nu is menselijke spraakproductie zo’n efficiënt medium dat het uitspreken van een paar extra woorden nauwelijks tijdsverlies oplevert, maar waarom en hoe mensen overspecificeren is onduidelijk. Het zou kunnen dat mensen overspecificeren omdat het hun eigen zoekproces vereenvoudigt (e.g., Engelhardt et al. 2006); het zoeken naar minimale verwijzingen vraagt immers veel rekenkracht. Maar het zou ook kunnen dat mensen overspecificeren om de hoorder te helpen bij het interpreteren van de verwijzing (e.g., Paraboni et al. 2007). Hoe dit alles computationeel verwerkt kan worden is nog een open vraag, maar het is duidelijk dat kijken naar echte mensen een aantal van de problemen op kan lossen.

Conclusie

Ik begon dit verhaal met de onstuimige, exponentiële groei van digitale media in de afgelopen jaren. In dat verband is het niet verwonderlijk dat steeds meer organisaties geïnteresseerd zijn in het ontwikkelen van virtuele personages; computertoepassingen waarin menselijke commu-

nicatie in al haar verbale en non-verbale rijkdom gesimuleerd wordt. Ik heb vandaag geprobeerd een idee te geven van wat er allemaal komt kijken bij het bouwen van dergelijke virtuele mensen en hoe het onderzoek dat we in Tilburg doen hier aan bij kan dragen. We zijn begonnen met de meest basale externe kenmerken; lippen die meebewegen met de spraak, ogen die de gebruiker niet al te onplezierig aanstaren, wenkbrauwen die op gezette tijden omhoog of omlaag gaan. Allemaal bedoeld om de suggestie van leven te wekken, maar op zichzelf nog weinig communicatief. Dat is anders voor de twee studies die ik vervolgens in wat meer detail heb besproken: het tonen en herkennen van emotionele gezichtsuitdrukkingen, en het automatisch produceren van taal.

Deze twee studies, hoe verschillend ook, hebben meer met elkaar gemeen dan u op het eerste gezicht misschien zou denken. Het gaat in beide studies om een spannende mix van communicatie- en informatiewetenschap met psychologie en informatica. Zowel emoties als spraakproductie zijn al eerder onderwerp van studie geweest, maar het perspectief van het virtuele personage werpt een heel ander licht op dit eerdere onderzoek, en roept allerlei nieuwe vragen op. En het belangrijkste punt: om die nieuwe vragen te beantwoorden is het essentieel om opnieuw te kijken naar hoe echte mensen met elkaar communiceren.

Het bouwen van computerpersonages die net zo communiceren als echte mensen: het is en blijft een ambitieus streven (maar ja, ik moet ook iets te doen hebben tot aan mijn pensioen). En wat nu als binnenkort blijkt dat de huidige trend naar meer menselijke computers zich niet doorzet? Mocht dat inderdaad zo zijn, is er in ieder geval vanuit wetenschappelijk oogpunt niets verloren. Het bouwen van virtuele mensen is misschien wel de beste manier om echte mensen beter te leren begrijpen. En ik kan me geen onderwerp van studie voorstellen dat inherent interessanter is dan dat.

Dankwoord

Mensen denken wel eens dat onderzoek doen een eenzame aangelegenheid is, maar niets is minder waar. Een niet onaanzienlijk deel van mijn dagen vul ik met gesprekken met studenten, promovendi en andere collega-onderzoekers. Samen brainstormen over onderzoeksplannen kan enorm inspirerend zijn en is één van de leukste kanten van dit werk. Het liefst ging ik nu alle studenten en collega's met wie ik in de afgelopen jaren dergelijke gesprekken gevoerd heb met naam en toenaam bedanken, maar ik zie dat Marc al onrustig begint te bewegen. Laat ik me dus beperken tot de mensen die voor mij de hoofdrollen hebben gespeeld in de weg tot hier en nu.

Die weg begon feitelijk 12 jaar geleden, bijna op de dag precies, toen ik ook hier stond, om het proefschrift te verdedigen dat ik onder begeleiding van Harry Bunt en Reinhard Muskens had geschreven. Van de aandacht voor helderheid, zowel in denken als in schrijven, die ik van Reinhard heb geleerd, heb ik nog steeds veel plezier, ook al pas ik die tegenwoordig op een wat andere manier toe.

Na mijn promotie kon ik dankzij Jan Landsbergen aan de slag bij wat toen nog het Instituut voor Perceptie Onderzoek (IPO) heette. Terugkijkend blijken de jaren op het IPO voor mij heel belangrijk geweest te zijn; daar kwam ik aanraking met taalgeneratie, en met een meer empirische, meer toegepaste manier van met taal en met communicatie omgaan. Ik ben Jan uiteraard heel dankbaar voor de kans die hij mij op het IPO gegeven heeft. Dat we nog steeds

een leuk contact hebben met Jan, en met Maureen, vind ik heel erg bijzonder. Op het IPO zijn ook een aantal plezierige samenwerkingsverbanden begonnen die tot op de dag van vandaag doorlopen, waarvan ik met name Kees van Deemter (nu in Aberdeen), Mariët Theune (nu in Twente) en Marc Swerts (nu in Tilburg) zou willen noemen. Erwin Marsi (nu even fysiek in Tilburg) hoort ook in dit rijtje thuis, al is onze samenwerking strikt genomen van na het IPO.

Mijn meest intensieve samenwerking is die met Marc geweest, eerst op het gebied van prosodie, en later op het bredere vlak van non-verbale communicatie. Marc: ik heb erg veel van je geleerd, met name van je scherpe intuïties en goede ideeën op experimenteel vlak. Ik denk met heel veel plezier terug aan de experimenten die we samen gedaan hebben, niet zelden ontstaan tijdens informeel werkoverleg in de Slegte. Dat we hier vandaag allebei staan zie ik als een mooie afronding van de eerste 10 jaar samenwerken. Maar we zijn nog lang niet klaar, ons beste werk gaat nog komen, en ik verheug me dan ook erg op de komende jaren.

Sinds 2001 ben ik weer terug in Tilburg, en enigszins tot mijn eigen verbazing met meer plezier dan ooit. Als één iemand daarvoor verantwoordelijk is, dan is dat Fons Maes. Het is vooral dankzij hem dat Bedrijfscommunicatie en Digitale Media een multidisciplinaire en wetenschappelijk verantwoorde opleiding is geworden, waaraan ook nog een mooi onderzoeksprogramma gekoppeld is dat zich richt op communicatie en cognitie. De meeste betrokkenen bij dit onderwijs en onderzoek zitten op de 4e verdieping van het Dante gebouw, en dat is, zoals iedereen weet, de hoogst haalbare sfeer binnen onze faculteit (in ieder geval tot de volgende interne verhuizing). Ik prijs mezelf gelukkig dat ik daar bij mag horen.

Het is nog maar de vraag of ik hier zou staan zonder al die mensen die ik hiervoor genoemd heb. *Zeker* is dat ik hier niet zou staan zonder het vertrouwen van het Faculteitsbestuur en van het Bestuur van de Universiteit, en ik ben uiteraard heel blij met dit vertrouwen.

Er zijn nog twee mensens zonder wie ik hier zeker niet zou staan: mijn ouders. Ik wil hen bedanken voor hun steun en de belangstelling die ze altijd gehad hebben in alles wat ik doe. Ik ben heel blij dat jullie er vandaag bij zijn.

Tot slot: ik heb werkelijk geen idee waar ik nu zou staan zonder Annemarie, en zonder onze kinderen. Vermoedelijk helemaal nergens, en een heel stuk minder gelukkig bovendien. Ik draag deze rede dan ook op aan Daan, Bas en Coen, zodat ze later nog eens na kunnen lezen wat hun vader toch voor rare experimenten uitvoerde. Als hij tenminste niet in een zwarte jurk rondliep. **Ik heb gezegd.**

References

- [1] Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinions in Neurobiology*, 12,169–177.
- [2] Barkhuysen, P., E. Kraemer and M. Swerts (2005), Problem Detection in Human-Machine Interactions based on Facial Expressions of Users, *Speech Communication*, 45(3), 343-359
- [3] Anja Belz and Albert Gatt (2007) The Attribute Selection for GRE Challenge: Overview and Evaluation Results, in *Proceedings of UCNLG+MT: Language Generation and Machine Translation*, pp. 75-83

- [4] Brave, S., C. Nass and K. Hutchinson (2005), Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent, *International Journal of Human-Computer Studies*, 62, 161–178.
- [5] Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (2000). *Embodied Conversational Agents*, Cambridge, MA: The MIT Press.
- [6] Darwin, C. (1872). *The expression of the emotions in man and animals*. 3rd edition, Oxford University Press [1998].
- [7] Dale, R. and E. Reiter (1995), Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions, *Cognitive Science* 19, 233-263
- [8] Van Deemter, K. (2002), Generating Referring Expressions: Boolean Extensions of the Incremental Algorithm, *Computational Linguistics* 28 (1), 37-52.
- [9] van Deemter, K. and E. Krahmer (2007), Graphs and Booleans, in: *Computing Meaning* (vol. 3), H. Bunt and R. Muskens (eds.), Studies in Linguistics and Philosophy, Kluwer Academic Publishers, 397-422.
- [10] Dehn, D and S. van Mulken (2000), The impact of animated interface agents: a review of empirical research, *International Journal of Human-Computer Studies*, 33, 1-22.
- [11] Ekman, P. (1972). *Emotion in the human face*. Pergamon Press.
- [12] Ekman, P. (1979). About brows: Emotional and conversational signals, in: *Human ethology: Claims and limits of a new discipline*, M. von Cranach, K. Foppa, W. Lepenies, D. Ploog (eds.), Cambridge: Cambridge University Press, pp. 169-202.
- [13] Engelhardt, P., K. Bailey and F. Ferreira (2006), Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54, 554–573.
- [14] Elfenbein, H. & Ambady, N. (2002), On the universality and cultural specificity of emotion recognition: a meta-analysis, *Psychological Bulletin*, 128, 203-235.
- [15] Garey, M. and D. Johnson (1979), *Computers and Intractability*, New York: Freeman.
- [16] Garrod, S. & M. Pickering (2004), Why is conversation so easy?, *Trends in Cognitive Sciences*, 8(1): 8-11.
- [17] Gratch, J., J. Rickel, E. André, N. Badler, J. Cassell and E. Petajan (2002), Creating Interactive Virtual Humans: Some Assembly Required, *IEEE Intelligent Systems*, 54–63
- [18] Grice, H.P. (1975), Logic and Conversation, in: *Syntax and Semantics* (vol. 3), Cole, P. and J. Morgan (eds.), New York: Academic Press (pp. 41-58).
- [19] James, W. (1884), What Is an Emotion? *Mind*, 9, 190–204.

- [20] Karson, C. (1983), Spontaneous eye-blink rates and dopaminergic systems, *Brain*, 106, 643–653.
- [21] Konijn, E. (2000), *Acting Emotions: shaping emotions on stage.*, Amsterdam University Press.
- [22] Kotchemidova, C. (2005), From good cheer to “Drive-By Smiling: a social history of cheerfulness, *Journal of Social History*, 39(1)
- [23] Krahmer, E. and M. Theune (2002), Efficient context-sensitive generation of referring expressions, in: *Information Sharing: Givenness and Newness in Language Processing*, K. van Deemter and R.Kibble (eds.), CSLI Publications, Stanford, 223-264
- [24] Krahmer, E., S. van Erk, & A. Verleg (2003), Graph-based Generation of Referring Expressions, *Computational Linguistics*, 29(1), 53-72.
- [25] Krahmer, E., van Dorst, J., & Ummelen, N. (2004). Mood, persuasion and information presentation, *Information Design Journal + Document Design*, 12, 40–52.
- [26] Krahmer, E. and M. Swerts (2004), More about brows: a cross-linguistic analysis-by-synthesis study, in: *From Brows to Trust: Evaluating Embodied Conversational Agents*, C. Pelachaud and Zs. Ruttkay (eds.), Kluwer Academic Publishers, 191-216.
- [27] Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, 48, 29–53.
- [28] Krahmer, E. & M. Swerts (2007a), The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception, *Journal of Memory and Language*, 396-414.
- [29] Krahmer, E. & M. Swerts (2007b), Perception of Congruent and Incongruent Emotional Speech, manuscript in voorbereiding.
- [30] Kraut, R., M. Patterson, V. Lundmark, S. Kiesler, T. Mukopadhyay & W. Scherlis (1998), Internet paradox: A social technology that reduces social involvement and psychological well being? *American Psychologist*, 53, 1017–1031.
- [31] Kurzweil, R. (2005), *The singularity is near: when humans transcend biology*, Viking Penguin.
- [32] Lester, J., K. Branting & Mott (2004), Conversational Agents, in *Practical Handbook of Internet Computing*, M. Singh, (Ed.), Chapman Hall & CRC Press, Baton Rouge.
- [33] Lewis Johnson, W., J. Rickel and J. Lester (2000), Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments, *International Journal of Artificial Intelligence in Education*, 11, 47–78.
- [34] Locke, J. (1998), *The de-voicing of society: Why we don't talk to each other anymore*, New York: Simon & Schuster.

- [35] Maes, A., A. Arts and L. Noordman (2004), Reference management in instructive discourse, *Discourse Processes*, 37, 117–144.
- [36] McQuiggan, S. and J. Lester (2007). Modeling and Evaluating Empathy in Embodied Companion Agents, *International Journal of Human-Computer Studies*, 65(4),
- [37] Paraboni, I., K. van Deemter and J. Masthoff (2007), Generating Referring Expressions: Making Referents Easy to Identify, *Computational Linguistics*, 33, 229-254.
- [38] Pogue, D. (2007), Making over the iPod family (again), *New York Times*, September 13, 2007.
- [39] Prendinger, H. and M. Ishizuka (2004), *Life-like characters: tools, affective functions and applications*. Springer: Berlin.
- [40] Prendinger, H. and M. Ishizuka (2005), The emphatic companion: a character-based interface that addresses user’s affective states, *Applied Artificial Intelligence*, 19, 267–285.
- [41] Reiter, E. and R. Dale (2000), *Building Natural-Language Generation Systems*. Cambridge University Press.
- [42] Rickel, J., S. Marsella, J. Gratch, R. Hill, D. Traum & B. Swartout (2002), Towards a new generation of virtual humans for interactive experiences, *IEEE Intelligent Systems*, 32–38.
- [43] Russell, J. (1994), Is there universal recognition of emotion from facial expressions? A review of cross-cultural studies. *Psychological Bulletin*, 115, 102-140.
- [44] Russell, J., J. Bachorowski, & J. Fernandez-Dols (2003), Facial and Vocal Expressions of Emotion, *Annual Review of Psychology*, 54, 329-349.
- [45] Sagan, C. (1997), *Billions & Billions*, Random House, New York.
- [46] Schaller, R. (1997), Moore’s law: past, present and future, *IEEE Spectrum*, 34, 52-59.
- [47] Schmidt, K., & Cohn (2002). Human facial expressions as adaptations: Evolutionary questions in facial expressions, *Yearbook of Physical Anthropology*, 44, 3–24.
- [48] Siddharthan, A. and A. Copestake (2004), Generating Referring Expressions in Open Domains, in *Proceedings of the 42th Meeting of the Association for Computational Linguistics Annual Conference (ACL 2004)*, Barcelona, Spain.
- [49] van der Sluis, I and E. Kraemer (2007), Generating Multimodal References, *Discourse Processes*, 44(3), 1–30.
- [50] Stepper, S. and F. Strack (1993). Proprioceptive Determinants of Emotional and Non-Emotional Feelings, *Journal of Personality and Social Psychology*, 64, 210–22.
- [51] Swerts, M. & E. Kraemer (2005), Audiovisual prosody and feeling of knowing, *Journal of Memory and Language*, 53(1), 81–94.

- [52] Swerts, M. & E. Kraemer (2007), Facial expressions and prosodic prominence: Comparing modalities and facial areas, *Journal of Phonetics*, in press.
- [53] Traum, D., A. Rogue, A. Leuski, P. Georgiou, J. Gerten, B. Martinovski, S. Narayanan, S. Robinson & A. Vaswanti (2007), Hassan: A virtual human for tactical questioning, in: *Proceedings of the 8th SIGdial workshop on Discourse and Dialogue*, S. Keizer, B. Bunt & T. Paek (eds), Antwerp, Belgium, pp. 75–78.
- [54] Valkenburg, P. M., & Peter, J. (2007), Preadolescents and adolescents online communication and their closeness to friends, *Developmental Psychology*, 43, 267- 277.
- [55] Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research & Therapy*, 6, 473–482.
- [56] Vroomen, J. (2006), *Horen met de ogen, zien met de oren*, inaugurele rede, 22 september 2006, Universiteit van Tilburg.
- [57] Weinberg, S. (1998), The revolution that didn't happen, *The New York Review*, 8 oktober, pp. 48-52.
- [58] Wilson, R. (1996), *Introduction to Graph Theory*. Longman, Harlow, England, fourth edition.
- [59] Wilting, J., E. Kraemer and M. Swerts (2006), Real vs. acted emotional speech, in: *Proceedings of the International Conference on Spoken Language Processing (Interspeech 2006)*, Pittsburgh PA, USA
- [60] Zappa, F. (1989). *The Real Frank Zappa Book*, New York: Poseidon Press.