

## Tilburg University

### Wat gaat er om in een virtueel hoofd?

Krahmer, E.J.; Swerts, M.G.J.

*Published in:*  
Automatiseringgids

*Publication date:*  
2007

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Krahmer, E. J., & Swerts, M. G. J. (2007). Wat gaat er om in een virtueel hoofd? *Automatiseringgids*, 41(48), 18-19.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Automatisering Gids

---

## Wat gaat er om in een virtueel hoofd?

Laatste update: 30-11-2007

door: Emiel Kraemer en Marc Swerts

Virtuele mensen worden op steeds meer plaatsen ingezet. Digitale assistenten helpen gebruikers. Maar voordat zo'n assistent realistisch kan communiceren, moet er heel wat gebeuren. Emiel Kraemer en Marc Swerts onderzoeken de mogelijkheden om virtuele mensen adequaat en flexibel te laten reageren op gebruikers.

Het is glashelder dat de recente ontwikkelingen op het gebied van digitale media de manier waarop mensen communiceren drastisch hebben veranderd. De exponentiële toename in computerrekenkracht maakt het mogelijk om computertoepassingen te ontwikkelen die face-to-face-interactie kunnen simuleren. We hebben het dan over computerpersonages die er uitzien als echte mensen en die verbaal en non-verbaal communiceren met gebruikers. Bedrijven zijn om uiteenlopende redenen in deze vorm van interactie geïnteresseerd. De overkoepelende reden is vaak het bevorderen van het contact met klanten. De ABN AMRO bijvoorbeeld heeft een virtuele assistente op zijn website die klanten quasileuk informeert over de ABN-site en ook haar eigen bedieningsknoppen uitlegt. Naast dit soort assistentes, altijd vrouwelijke personages trouwens, zie je ook in de wereld van computerspelletjes een duidelijke trend naar virtuele werelden waarin gamers het 'echte leven' kunnen ontvluchten door het na te spelen op de computer. Nog een stapje verder gaan spellen als Second Life, waarin mensen met elkaar communiceren via digitale computerpersonages, de zogenaamde avatars.

Inmiddels zien mensen ook het nut in van virtuele personages voor 'serious gaming', een hippe term voor spelenderwijs leren. Denk aan al die kinderen (en volwassenen) die momenteel hun brein trainen op de Nintendo DS, aangemoedigd door de enthousiaste 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.

In de meeste toepassingen die we hier hebben genoemd, wordt het computerpersonage bestuurd door een echt mens 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' genoemd.

Alle bouwstenen die nodig zijn om een echt virtueel karakter te maken, zijn in rudimentaire vorm aanwezig (zie kader Bouwen). Hoewel nog niet alle basistechnologie goed genoeg is, werken onderzoekers sinds de tweede helft van de jaren negentig serieus aan de ontwikkeling van dit soort virtuele mensen. En dit roept meteen allerlei basale vragen op, waarbij we ons hier gemakshalve beperken tot – letterlijk en figuurlijk – de hoofdzaken. Wat moet je bijvoorbeeld zien aan een virtueel hoofd en hoe kun je dat implementeren?

Lippen die automatisch meebewegen met de spraak (zie kader Visuele spraak) is een sine qua non voor virtueel pratende hoofden, maar als het personage verder niet beweegt, ziet dat er buitengewoon onnatuurlijk uit. Vandaar dat allerlei technische trucs worden uitgehaald om de suggestie van leven te wekken. Zo'n intens sturende 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. 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 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 zijn dit essentiële voorwaarden voor een succesvolle toepassing van virtuele personages. We weten inmiddels dat gebruikers een virtueel mens alleen

waarden als het een toegevoegde waarde in een applicatie heeft, en als het op een functionele, maar toch realistische manier communiceert met gebruikers, zowel verbaal als non-verbaal.

Maar dan beginnen de problemen pas echt. Menselijke sprekers gebruiken bijvoorbeeld de wenkbrauwen om belangrijke woorden visueel te benadrukken, zo bleek uit het pionierswerk van Paul Ekman. Op zich is het eenvoudig om een computerpersonage zo te programmeren dat het de wenkbrauwen zo nu en dan omhoog of omlaag beweegt bij het uitspreken van belangrijke woorden, maar automatisch vaststellen wat de belangrijke woorden zijn, is een stuk ingewikkelder. Hoeveel van die belangrijke woorden zitten er in een zin? Als het 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. De enige oplossing is om nog maar eens goed naar echte mensen te kijken, en de kans te berekenen dat een spreker een bepaald woord (bijvoorbeeld afhankelijk van de specifieke positie van het woord in de zin) uitspreekt en van een wenkbrauwbeweging vergezeld doet gaan.

Uiteraard doen sprekers meer dan enkel de belangrijke woorden visueel onderstrepen. We kunnen bijvoorbeeld ook aan menselijke sprekers zien of ze zeker of juist onzeker zijn over wat ze zeggen, over hoe lang ze nog aan het woord willen blijven, en of ze vinden dat de communicatie vlotjes verloopt of juist moeizaam. Ook over dit soort aspecten van de communicatie zouden dus automatisch beslissingen genomen moeten worden; zowel om vast te stellen of het virtuele personage wil signaleren dat het niet helemaal zeker is over de informatie die het geeft, als over de manier waarop het deze onzekerheid kan signaleren (bijvoorbeeld door het automatisch synthetiseren van spraakmarkers zoals 'uh' of het visueel tonen van een weifelende gezichtsuitdrukking).

Het bouwen van computerpersonages die net zo communiceren als echte mensen is en blijft een ambitieus streven. Allereerst gaat het om een spannende mix van informatica met psychologie en met communicatie- en informatiewetenschap. Bovendien roept het perspectief van een virtueel mens allerlei nieuwe vragen op over de manier waarop echte mensen nu eigenlijk met elkaar communiceren. Het bouwen van virtuele mensen is misschien wel de beste manier om echte mensen beter te leren begrijpen.

*Emiel Kraemer en Marc Swerts zijn respectievelijk hoogleraar Digitale Media en Organisatie en hoogleraar Tekst- en Informatiewetenschap bij de Faculteit Geesteswetenschappen van de Universiteit van Tilburg. Dit artikel is gebaseerd op hun oraties (2 november 2007). Voor meer informatie: zie <http://foap.uvt.nl/>.*

#### Visuele spraak

Een taal als het Nederlands kent ongeveer veertig elementaire klanken (of fonemen), als /aa/, /oo/ en /uu/. 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 /oo/ en de /uu/, bijvoorbeeld, klinken anders, maar zien er hetzelfde uit. Visuele spraak kan nu automatisch geproduceerd 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, met eventueel een lineaire interpolatie ertussen. Dat ziet er nog niet helemaal natuurlijk uit, omdat de visuele realisatie van klanken contextafhankelijk is. Wie zachtjes 'stoel' en 'steen' uitspreekt, voelt dat bij de /s/ de lippen al rekening houden met het feit dat ze straks een /oe/ of een /ee/ gaan produceren. Dit is eenvoudig op te lossen door de associatie van klanken met lipposities afhankelijk te maken van de omringende klanken.

#### Bouwen van een virtueel mens

Vanuit een technisch oogpunt kun je het bouwen van 'virtuele mensen' zien als een combinatie van 'real-time computer graphics', kunstmatige intelligentie en taal- en spraaktechnologie. Aan de ene kant (invoer) is technologie nodig om automatisch te kunnen verstaan en te begrijpen wat de gebruiker zegt (spraakherkenning en taalanalyse/natural language processing) en ook om gebaren en gezichtsuitdrukkingen van de gebruiker te kunnen herkennen en interpreteren ('face tracking' en 'gesture tracking'). Aan de andere kant moet een virtueel mens zelf ook weer gesproken taal kunnen produceren (via 'natural language generation' en spraaksynthese), met

bijpassende non-verbale signalen, en er verder een beetje fatsoenlijk uitzien (3D real-time computer graphics). Het verbeteren en integreren van de 'enabling technologies' is al een behoorlijk technische klus. Maar er zijn ook nieuwe technische uitdagingen: • Een echt virtueel mens moet op een adequate en flexibele manier op bijvoorbeeld een vraag van de gebruiker kunnen reageren, en moet dus ook zelf een antwoord kunnen formuleren. Dit vereist niet alleen automatische taalgeneratie, ook moet automatisch bepaald worden hoe het antwoord audiovisueel uitgesproken zal worden (Welke woorden moeten bijvoorbeeld benadrukt worden en hoe dat te doen?). • Het virtuele personage moet automatisch kunnen beslissen wanneer en waarom het een bepaald visueel signaal laat zien. Het moet dus zelf kunnen vaststellen wat bijvoorbeeld een goed moment is om de wenkbrauwen op te richten of om een onzekere gezichtsuitdrukking op het virtuele gezicht te tonen.

*Voor reacties en nieuwe bijdragen van deskundigen: Henk Ester (h.ester@sdu.nl, (070) 378 03 9 )*

Verschenen in: Automatisering Gids, 48, 2007

---

© Sdu Uitgevers, Den Haag

<p><b>Automatisering Gids</b> biedt u: meer dan 2000 whitepapers, business cases en reference cases. Ga naar <a href="http://automatiseringgids.nl/whitepapers">automatiseringgids.nl/whitepapers</a> en bekijk het complete overzicht.</p>
---

---