

Intelligente Omgevingen: Ondersteunen van Interacties en Terugkijken op Activiteiten

Anton Nijholt
Universiteit Twente
Center of Telematics and Information Technology (CTIT)
Postbus 217, 7500 AE Enschede, Nederland
anijholt@cs.utwente.nl

Samenvatting

Intelligente omgevingen (smart environments) bevatten sensor- en AI-technologie die het enerzijds mogelijk maken in real-time ondersteuning te geven aan activiteiten die zich in de slimme omgeving afspelen en anderzijds, door op een of andere wijze wat is waargenomen en wellicht geïnterpreteerd op te slaan, geïnteresseerden de mogelijkheid te geven te grasduinen in wat is opgenomen, er vragen over te stellen en wellicht door het opgeslagene automatisch geattendeerd te worden op interessante zaken. In dit artikel zullen we invulling geven aan deze gezichtspunten door te kijken naar een Europees 6de Kader Project waaraan onze onderzoeksgroep deelneemt. Dit project, gestart in 2004, is het AMI project, waarbij AMI staat voor Augmented Multi-party Interaction. Uitgangspunt bij dit project zijn interacties die plaatsvinden tijdens vergaderingen. Dit maakt het begrijpelijk dat er niet alleen aandacht wordt besteed aan computerondersteuning van activiteiten, maar ook aandacht aan ondersteuning bij het terugkijken op activiteiten. Raadplegen van tijdens een vergadering geschreven notulen is een voorbeeld van dit terugkijken waarbij echter op geen enkele wijze gebruik wordt gemaakt van de huidige mogelijkheden van sensor- en AI-technologie. In dit artikel zal ingegaan worden op deze mogelijkheden zoals ze vorm krijgen binnen het genoemde AMI project.

1, INLEIDING

Intelligente omgevingen maken het mogelijk om activiteiten die zich afspelen in die omgevingen te ondersteunen, deelnemers op afstand toegang te geven tot die omgevingen en de activiteiten (zowel passief als actief), en de observaties en interpretaties van de activiteiten ook naderhand toegankelijk te laten zijn. Met andere woorden, we kunnen spreken over deelnemers die fysiek aanwezig zijn in een fysieke omgeving, we kunnen spreken over een vertaling van een fysieke omgeving naar een virtuele omgeving, we kunnen spreken over deelnemers (mogelijkerwijze afkomstig van ver van elkaar verwijderde locaties) die op een of andere wijze gerepresenteerd worden in een virtuele omgeving, en we kunnen spreken van ‘deelnemers’ die achteraf kennis nemen van activiteiten die zich afgespeeld hebben in een bepaalde omgeving. Dit laatste soort deelnemers kunnen de loop van de activiteiten niet beïnvloeden, ze kunnen wel de activiteiten meemaken vanuit een bepaald gezichtspunt.

Bij intelligente omgevingen (‘smart environments’) wordt vaak uitgegaan van een ondersteuning door de omgeving die gericht is op de bevordering van efficiëntie van de (menselijke) activiteiten. Om dat te bereiken wordt de omgeving uitgerust met sensoren die waarnemen. Dat kunnen microfoons en camera’s zijn, maar er zijn ook tal van mogelijke gespecialiseerde sensoren die in groten getale aanwezig kunnen zijn en meehelpen om informatie te verzamelen, te

integreren, te interpreteren en op grond van die interpretaties relevante interpretatieresultaten beschikbaar te stellen ten behoeve van de activiteiten in de omgeving.

In dit artikel gaan we in op het onderzoek dat door ons en anderen wordt verricht binnen een Europees project op het gebied van multimodale interactie in intelligente omgevingen. Het project heet het Augmented Multi-party Interaction (AMI) project¹. Het modelleren van de interactie (verbaal en non-verbaal) tussen meerdere mensen is een belangrijk doel van het project en om toch enige beperking aan te brengen aan de te modelleren interacties is gekozen voor interacties binnen een vergaderruimte. Wanneer we in staat zijn (aspecten van) die interacties te modelleren dan zijn we ook in staat (aspecten van) interacties te begrijpen, te voorspellen en er op in te spelen. Dat wil zeggen, we kunnen op meer of minder beperkte wijze ondersteuning geven vanuit de omgeving aan een vergadering en we kunnen deelnemers of anderen intelligente algoritmen aanbieden voor toegang achteraf tot de verzamelde vergaderinformatie. In beide gevallen kunnen we een onderscheid maken in het gebruik van direct observeerbare data en geïnterpreteerde data. Er is echter geen duidelijke grens aan te geven tussen observatie en interpretatie.

Real-time ondersteuning van de activiteiten vereist technologie die bespeurt waar zich activiteiten plaatsvinden, die activiteiten kan volgen en interpreteren. Bij vergaderen gaat het vooral om de verbale en non-verbale interactie tussen mensen, maar er kunnen ook presentaties gehouden worden, mensen komen binnen en moeten geïdentificeerd worden, men verandert van plaats, etc. In een bijeenkomst van ontwerpers zullen mogelijke alternatieve versies van een product op tafel komen en besproken worden. Multimodale interactie tussen mensen onderling en met (objecten in) hun omgeving dient dus gemodelleerd te worden. Ook de te geven ondersteuning zal van verschillende modaliteiten en media gebruik maken. Toegang tot geregistreerde vergaderingen kan door middel van multimedia retrieval en browsing technieken en informatie kan met behulp van multimedia gepresenteerd worden.

In de volgende sectie gaan we wat dieper in op zaken die spelen bij alom aanwezige intelligentie in huiskamer en kantoor situaties. Sectie 3 is gewijd aan een inleiding in twee van onze vergaderprojecten en het onderzoek dat daar gebeurt wat betreft het herkennen en modelleren van vergadersituaties. In sectie 4 gaan we wat dieper in op de ideeën die binnen het huidige project ontwikkeld en voor een deel gerealiseerd worden op het terrein van on-line en off-line ondersteuning, respectievelijk tijdens en na een vergadering. Sectie 5 van dit artikel is gewijd aan een invulling van onze kant wat betreft het toegankelijk maken van vergadermateriaal, zowel on-line als off-line, en gebruikmakend van virtual reality technologie. Dit artikel wordt besloten met enige conclusies en observaties over toekomstig onderzoek.

2. ALOM AANWEZIGE INTELLIGENTIE: WENSEN EN VEREISTEN

Huiskamers, werk- en recreatieruimten, en publieke ruimten kunnen worden uitgerust met camera's, microfoons en andere waarnemingsapparatuur (sensoren). Dat kan op vrijwel onzichtbare wijze. De waarnemingsapparatuur wordt onzichtbaar en de intelligentie die zowel in de waarnemingsapparatuur als in de achterliggende computersystemen (zorgdragend voor de integratie van informatie uit verschillende mediabronnen en de interpretatie ervan) aanwezig is, heeft geen direct zichtbare interface meer. De computer is uit het zicht verdwenen en de omgeving is de interface.

Een inwoner, bezoeker of gebruiker kan zich wenden tot de omgeving voor ondersteuning bij zijn of haar activiteiten. De omgeving kan hulpmiddelen bieden om voor de gebruiker duidelijk te

¹ AMI startte op 1 januari 2004 en heeft een duur van drie jaar. Het project wordt gefinancierd uit het Europese 6^{de} IST Raamwerk Programma.

maken wat geadresseerd kan worden. Dat kunnen objecten zijn (smart objects), maar ook displays die ingebouwd zijn in objecten (stoelen, kleding, gereedschap) en omgevingen (wanden en tafels bijvoorbeeld). Ook in dit soort slimme omgevingen kan de inwoner/bezoeker/gebruiker gebruik maken een PDA, een notebook of een PC, of meer in het algemeen een persoonlijke assistent die meereist van de ene naar de andere omgeving. Die persoonlijke assistent, maar eigenlijk geldt dat ook voor de interactie met andere omgevingen, kan vorm krijgen als een synthetische belichaamde persoonlijkheid (embodied agent, talking face, virtual human, avatar).

Gegeven deze observaties is het zinvol de volgende gezichtspunten te onderscheiden.

- **Waarnemen en Interpretatie van Gebeurtenissen en Activiteiten in een Omgeving.** Dit waarnemen betreft vooral de sociale en intelligente interacties tussen mensen, tussen mensen en objecten, tussen mensen en autonome belichaamde agents (virtuele mensen, robots) en interacties met de omgeving in het algemeen (niet noodzakelijkerwijs gericht tot een bepaald object of agent). Invoer kan worden verkregen van sensoren voor geluid, beeld en aanraking. De interactie die waargenomen moet worden omvat niet alleen alle aspecten van gefocuste interactie, maar ook die van ongefocuste. Interpretatie vereist de fusie van al de modaliteiten die door de omgeving waargenomen kunnen worden op verschillende niveaus van syntactische, semantische en pragmatische representatie van waaruit de verkregen informatie verder verwerkt kan worden.
- **Verzorgen van Real-time Ondersteuning van Gebeurtenissen en Activiteiten in een Omgeving.** Van slimme omgevingen verwachten we real-time ondersteuning van onze activiteiten. Die ondersteuning zal gebaseerd moeten zijn op de verkregen computationele representatie en de mogelijkheid daarover te redeneren. De ondersteuning wordt gegeven door de omgeving, objecten in de omgeving (verplaatsbaar, mobiel) en virtuele belichaamde agents die zich tonen op displayfaciliteiten in de omgeving. Waar bij het waarnemen de integratie van de signalen uit de verschillende mediabronnen het belangrijkste onderzoeksonderwerp is, is bij het verschaffen van ondersteuning aan de gebruiker(s) van de omgeving de selectie van de informatie en de meest toepasselijke multimedia display combinatie, rekening houdend met de karakteristieken van de gebruiker, een belangrijk onderzoeksonderwerp. Er kan daarbij sprake zijn van expliciete verzoeken aan de omgeving, zowel wat betreft inhoud als vorm, er kan ook sprake zijn van beslissingen van de omgeving wat betreft het aanbieden van informatie en de vorm ervan, het tijdstip ervan en, wanneer nodig geoordeeld, het ongevraagd aanbieden van deze informatie.
- **Gebeurtenissen en Activiteiten in een Omgeving: Terugblikken, Grasduinen en Gericht Zoeken.** Na afloop terugkijken naar wat er gebeurd is in een slimme omgeving is een volgend onderwerp. Informatie die ten behoeve van real-time ondersteuning bijeen is gebracht en is geïnterpreteerd kan ook naderhand worden bevraagd of anderszins getransformeerd naar een (multimedia) presentatie van voor een gebruiker relevante informatie. In de beschikbare informatie kan gebrowseed worden, er kunnen vragen over worden gesteld en er kunnen samenvattingen of alternatieve presentaties worden gegenereerd, bijvoorbeeld in virtual reality. Natuurlijk kan ook voor on-line ondersteuning dergelijke technologie een belangrijke rol spelen.

Een tweetal observaties moet hier nog aan toegevoegd worden. De eerste betreft het eigendom en de toegang tot de informatie die door slimme omgevingen wordt opgeslagen. De tweede observatie betreft de rol van autonome en semi-autonome agents die een rol spelen in de slimme omgevingen.

- **Autonome en semi-autonome embodied agents.** Autonome (belichaamde) software agents kunnen deel uitmaken van een ambient intelligence omgeving. We kunnen echter ook embodied agents in de omgeving hebben die meer of minder expliciet door anderen,

bijvoorbeeld niet fysiek aanwezig, real-time gecontroleerd worden. Bijvoorbeeld doordat hun gedrag gekoppeld is aan het (fysieke) gedrag van iemand op afstand, waardoor ze in feite deze persoon representeren. Een dergelijk virtueel personage kan ook meer of minder autonoom gedrag vertonen. Als het minder interessant is om aanwezig te zijn kan de eigenaar tijdelijk meer autonomie verlenen (het virtuele personage gaat bijvoorbeeld enkel luistergedrag vertonen of krijgt de rol van een intelligente camera die de eigenaar gaat attenderen op activiteiten zodra ze weer interessant worden.

- **Toegang tot en Beheer van Opgeslagen Waarnemingen.** Waarnemen van activiteiten in een slimme omgeving kan on-line en is niet noodzakelijkerwijs voorbehouden aan deelnemers (fysiek of virtueel aanwezig). Degene die de omgeving beheert of eigenaar is kan ook on-line waarnemen en off-line toegang hebben tot de opgeslagen waarnemingen. Technologie wordt ontwikkeld om de opgeslagen waarnemingen toegankelijk te maken voor retrieval, browsing, samenvatten en multimedia presentatie. Is er nog enige privacy voor gezinsleden, kantoorpersoneel, vergaderders of stadionbezoekers? Er wordt gekeken naar de gebruikers en bezoekers van slimme omgevingen. Hoe beïnvloedt dit hun gedrag? Wetende dat er ogen en oren zijn die hun gedrag observeren en dat die observaties op technologisch geavanceerde wijze toegankelijk zal worden gemaakt voor onbekenden zal gedrag beïnvloed worden en minder natuurlijk, wellicht minder effectief en minder plezierig zijn. Onbespied zijn, mag dat?

Een aantal van deze onderwerpen hebben we eerder bediscussieerd, bijvoorbeeld in de context van interactieve voorstellingen waarbij menselijke performers moeten interacteren met objecten en virtuele mensen in een virtual reality omgeving, in de context van sociaal intelligente embodied agents [13], en in de context van aanwezigheid (presence), vervreemding en privacy [14]. In het bijzonder onze betrokkenheid bij twee Europese projecten over vergaderomgevingen (M4 in het 5^{de} kaderprogramma en AMI in het 6^{de} kaderprogramma hebben geleid tot verdere ontwikkeling van ideeën en generalisatie van ideeën naar intelligente omgevingen in het algemeen (thuis-, kantoor-, vergader- en publieke omgevingen) [16]. Mede van daar willen we in de volgende secties aandacht besteden aan deze projecten, en in het bijzonder onze inbreng daarin.

3. Het Augmented Multi-party Interaction (AMI) Project

3.1 Inleiding

De belangrijkste doelen van het AMI (Augmented Multi-party Interaction) project zijn het ontwikkelen van technologie voor het toegankelijk maken van tijdens een vergadering gegenereerde informatie. Dat betreft informatie die traditioneel toegankelijk gemaakt wordt in notulen. Enerzijds is er de vraag hoe we die notulen kunnen vervangen door een multimedia presentatie en generatie van de waargenomen activiteiten met bijbehorende zoek, vraag en browse faciliteiten. Anderzijds is er de vraag hoe we ondersteuning kunnen verrichten aan de deelnemers van een vergadering tijdens de vergadering. De vragen zijn niet essentieel verschillend, ze zijn verschillend wat betreft mogelijkheden om on-line of off-line toegang te krijgen tot de gewenste informatie en de mogelijkheid om real-time geattendeerd te worden op interessante activiteiten die zich afspelen in de (in dit geval) vergaderomgeving.

In beide gevallen zal er sprake moeten zijn van een begrip van het hoe en waarom van de verbale en non-verbale interacties in de omgeving. De context aangeboden door de omgeving is natuurlijk in belangrijke mate bepalend voor de interpretatie van de interacties en andere gebeurtenissen in de omgeving.

Het AMI project is een voortzetting van het M4 (Multi-Modal Meeting Manager) project. In dit project lag de nadruk voornamelijk op het herkennen van low-level signalen (spraak en beeld) en veel minder op de interpretatie ervan. In het AMI project is er een bescheiden verschuiving in de richting van de interpretatie. Een onderwerp dat daardoor actueel wordt is de ontwikkeling van vergaderassistenten die met hun tijdens een vergadering verkregen kennis een rol kunnen spelen tijdens de vergadering. Dergelijke assistenten kunnen hulp verschaffen aan individuele deelnemers, aan deelnemers met een bepaalde rol of aan alle deelnemers en voor een ieder waarneembaar. Die verschuiving naar interpretatie maakt het ook gemakkelijker te denken aan deelname op afstand, waarbij deelnemers aan activiteiten (in dit geval vergaderen) niet fysiek aanwezig zijn. Er moet dan desondanks deelgenomen worden aan verbale en non-verbale interactie. Kijk- en beurtgedrag (turn-taking) dient gemedieerd te worden om soepele interactie mogelijk te maken. Aan dergelijk gedrag liggen modellen voor multi-party interactie ten grondslag en ten behoeve van zowel real-time ondersteuning als van deelname op afstand dient modelgebaseerd gereedschap ontwikkeld te worden.

3.2 Datacollectie en Data-annotatie

Zowel binnen het inmiddels beëindigde M4 als binnen het huidige AMI project is op grote schaal vergaderdata verzameld. Dit is gebeurd in vergaderruimten in het IDIAP Instituut (Martigny), op de Universiteit van Edinburg en bij TNO Human Factors in Soesterberg. Figuur 1 geeft een beeld van de vergaderruimte bij IDIAP.

Het vergaderen wordt waargenomen, opgenomen en opgeslagen. Dat gebeurt door middel van camera's, microfoons en informatie die wordt verkregen uit notities ('smart pens') en uit PowerPoint presentaties. In het bijzonder spraak- en beeldverwerking, de integratie van informatie uit deze twee mediabronnen en het presenteren en toegankelijk maken krijgen aandacht. Het hoofdcorpus bestaat uit honderd uur video/audio registratie van vergaderactiviteit, opgenomen met enkele videocamera's voor overzichtsbeelden, camera's voor afzonderlijke deelnemers, globale microfoonopstellingen en reversmicrofoons voor de deelnemers.

Om data te verzamelen en vergaderactiviteit te bestuderen zijn vergaderingen geregeld. Bij het M4 project was sprake van 'mock-up' vergaderingen waarbij de deelnemers zorgvuldig een script volgden waarin ieders acties binnen een globaal vergaderonderwerp waren vastgelegd. Per activiteit waren er geen beperkingen, met andere woorden, er was niet vastgelegd hoe iemand inhoud gaf aan zijn bijdrage. Deze vergaderingen zijn gebruikt om te kijken hoe sequenties van vergaderactiviteiten herkend kunnen worden, om in kaart te brengen wat er gemodelleerd dient te worden en welke theorie daarvoor aanwezig is of dient te worden ontwikkeld. In het huidige AMI project worden meer natuurlijk vergaderingen geregistreerd. Scenario's zijn bedacht die een meer natuurlijke interactie tussen vergaderdeelnemers toelaten en waarbij de verschillende deelnemers gevraagd worden verschillende rollen te spelen. De scenario's hebben betrekking op ontwerpbijschotten en de rollen die worden onderscheiden zijn die van een projectmanager, een ontwerper van gebruikers-interfaces, een industrieel ontwerper en een



Figuur 1. Vergaderzaal bij IDIAP

marketing deskundige. Hoe vergadert een dergelijk team over het ontwerpen van een afstandsbediening?

Hoe komt een vergadering tot een beslissing? Welk verbaal en non-verbaal interactiegedrag speelt daarbij een rol en wat ligt ten grondslag aan dat gedrag? Er bestaan globale theorieën over (sociaal) interactiegedrag en hoewel er ook in dit project veel aandacht is voor die theorieën is de benadering binnen het project meer technologisch. Welke signalen kunnen we herkennen, samenvoegen, en interpreteren? Echter, het herkennen, samenvoegen en interpreteren wordt mede geleid door deze theorieën. Aangezien de theorieën, zeker vanuit een computationeel oogpunt, verre van perfect en volledig zijn kan men op dit moment niet verwachten dat volautomatisch alle aspecten van vergadergedrag herkend en geïnterpreteerd kunnen worden. Vandaar dat in het project er zowel sprake is van de ontwikkeling van gereedschap gebaseerd op tamelijk rudimentaire herkenning en interpretatie, als van de ontwikkeling van annotatiegereedschap dat voor een belangrijk deel uitgaat van handmatige herkenning en interpretatie en dat meehelpt om materiaal te verzamelen en te analyseren ten behoeve van de verdere ontwikkeling van theorieën en daarop gebaseerd gereedschap. Wat betreft dit laatste kan gewezen worden op annotatieschema's en annotatiegereedschap dat helpt bij het beschrijven van eigenschappen van dialoogbijdragen, van kijk- en beurtwisselingsgedrag van vergaderdeelnemers, van lichaams-houdingen en gebaren, etc. Evenzo, op een iets hoger niveau, kunnen pogingen ondersteund worden om te beschrijven wat het effect is van hiërarchische verhoudingen, hoe dominant gedrag herkend kan worden of welk non-verbaal gedrag gepaard gaat met het veranderen van onderwerp tijdens een discussie.

3.3 Herkennen en Interpretatie van Vergaderactiviteiten

Zoals gezegd, volautomatisch vergaderactiviteiten en situaties herkennen en interpreteren is niet aan de orde. Spraakherkenning alleen al, in een situatie met meerdere sprekers, zal verre van perfect zijn, laat staan de interpretatie van wat gezegd wordt. Tal van andere te gebruiken technologie zou om volmaakt te moeten zijn uit moeten gaan van het hebben van oplossingen voor alle artificiële intelligentie problemen. Echter ook onvolmaakte technologie kan nuttig gereedschap opleveren dat kan meehelpen om de vergaderomgeving intelligent te maken of om achteraf de verzamelde informatie toegankelijk te maken.

Het is zinvol onderscheid te maken tussen het herkennen en interpreteren van groepsactiviteiten en activiteiten van individuele deelnemers. Bij beiden zijn modellen nodig voor de integratie van de informatie afkomstig van afzonderlijke bronnen en modaliteiten. Deze modellen vormen de basis van browsing-, retrieval-, extractie- en samenvattingstechnologie. Vaak zijn deze modellen probabilistisch van aard. Binnen het (inmiddels afgelopen) M4 project werd uitgebreid gekeken naar het herkennen van groepsacties tijdens vergaderingen (momenten van discussie, monologen, presentaties, het maken van aantekeningen, stemmen, etc.). Voor dat doel zijn allerlei varianten van Hidden Markov Modellen in het leven geroepen [10]. In het huidige AMI project is er meer aandacht voor het herkennen van de activiteiten van de afzonderlijke deelnemers en het integreren van de herkenningresultaten op een niveau waar ook interpretatie kan plaatsvinden.

Het herkenningsonderzoek binnen AMI wordt opgehangen aan de volgende reeks vragen die moeten kunnen worden beantwoord door een intelligente omgeving.

- “Wat is gezegd tijdens de vergadering?”
- “Wie zijn aanwezig? Wie heeft het woord?”
- “Welke emoties zijn bij de deelnemers merkbaar?”
- “Waar bevindt een ieder zich?”
- “Welke activiteiten zin te onderscheiden?”
- “Wat gebeurt er tijdens de vergadering? Hoe gedragen mensen zich?”

- “Waar is de aandacht van de deelnemers op gericht?”

De basis van herkenning is het audio-video verwerkingsproces, dat wil zeggen, spraak- en beeldherkenning en vooral de integratie van informatie afkomstig van die twee bronnen. Vooral dat laatste is een onderwerp van nieuw onderzoek. Robuuste spraakherkenning in een situatie waar verschillende mensen het woord voeren en niet altijd even gemakkelijk van elkaar te onderscheiden zijn is een bijzonder moeilijk probleem. Hetzelfde geldt voor beeldherkenning als in een omgeving verschillende personen door elkaar heen lopen. Personen moeten door de omgeving gevolgd en geïdentificeerd worden. Non-verbaal communicatiegedrag helpt mee aan het betekenis geven aan uitingen tijdens een bijeenkomst. Ook dit moet dus herkend worden: gebaren, verandering van lichaamshouding, gezichtsuitdrukkingen, kijkgedrag en hoofdbewegingen spelen daarbij essentiële rollen. Zodra we modaliteiten gaan combineren is er sprake van een interpretatie die gestuurd wordt door de context van de interactie, de doelen die nagestreefd worden door de vergadering en/of de individuele deelnemers of die van de waarnemer.

Zinvolle informatie kan verkregen worden door het meten en herkennen van direct observeerbaar gedrag. Op grond van dergelijke informatie kan een voorzitter of kunnen de aanwezigen hun gedrag aanpassen in een richting die meer bijdraagt aan het doel van de bijeenkomst. Die informatie kan ook gevoed worden aan virtuele vergaderassistenten die reactief en proactief invloed kunnen hebben op het verloop van de vergadering. In sectie 4 komen we hier op terug. Onze eigen inbreng in het project wat betreft deze herkenningsonderwerpen hebben vooral betrekking op kijkgedrag en beurtwisseling, lichaamshouding en gebaren. Ook zijn er experimenten gedaan, gebruikmakend van op de hoofden van de vergaderdeelnemers bevestigde sensoren, met herkenning van hoofdoriëntatie als een benadering van kijkrichting [20]. Weten naar wie door een spreker wordt gekeken of voor wie zijn of haar uiting in het bijzonder bestemd is, is nuttig voor vergaderdeelnemers en kan daarom ook een belangrijke rol spelen bij on- en offline situaties waar dit niet zo direct kan worden waargenomen (zie ook sectie 5). In aan AMI verwante projecten van ons wordt overigens ook gekeken naar gezichtsuitdrukkingen en emotieherkenning om een omgeving daar sociaal en intelligent op te laten reageren.

Hoewel er zeker sprake is van een toenemende kwaliteit van spraak- en beeldherkenning en ook de verwerking van de herkenning tot interpretaties en daarop gebaseerde acties steeds meer aandacht krijgt in het onderzoek, kan er echter in slechts beperkte mate sprake zijn van real-time ondersteuning vanuit de intelligente omgeving aan de vergaderdeelnemers. Pogingen om herkenning en interpretatie op een hoger niveau te brengen, dat wil zeggen, meer allesomvattende en diepergaande theorieën en daarop gebaseerde modellen te ontwikkelen, baseren zich vaak op een analyse van het corpus materiaal en het op systematische wijze, deels automatisch en deels handmatig, voorzien van een dergelijk corpus van annotaties die door (te ontwikkelen of te leren) interpretatiealgoritmen als voorbeeld en/of als invoer gebruikt kunnen worden.

3.4 Annotaties en Annotatiegereedschap

Annotaties voegen syntactische, semantische en pragmatische informatie toe aan waarnemingsdata. Zowel annotaties als de theoretische modellen waarop ze zijn gebaseerd kunnen vergaderactiviteiten op verschillende niveaus beschrijven. Het herkennen van namen vergt niet al teveel kennis en kan daarom ook automatisch gebeuren, het automatisch identificeren van de deelnemers aan een vergadering is al moeilijker, de interpretatie van een gebaar zal vaak nog meer kennis vereisen, laat staan het uitleggen van stemgedrag of emoties. Annotaties kunnen ook betrekking op het beschrijven van fenomenen die van toepassing zijn op een enkel individu (bijvoorbeeld intonatie of gezichtsuitdrukking) of op een groepsactiviteit (een verandering van

onderwerp, een stemming). In de vorige sectie hebben we gezien dat er vanuit de computationele linguïstiek en vanuit de beeldherkenning methoden en technieken beschikbaar zijn voor het automatisch herkennen van vooral syntactische informatie (sleutelwoordherkenning, naamherkenning, spraakherkenning, naamwoordelijke en zinsontleding, het herkennen van vormen en objecten, persoonidentificatie, lokaliseren en volgen van personen, het herkennen van sequenties van gebeurtenissen, etc.) en er kan gereedschap ontwikkeld worden, gebaseerd op modellen die aspecten van non-verbale multi-party interactie beschrijven (hoofdoriëntatie, kijkgedrag, lichaamshoudingveranderingen, het laten zien van emoties), die informatie verschaffen voor real-time ondersteuning of voor het achteraf toegankelijk maken van de verzamelde informatie. Diepergaande annotaties (niet zozeer herkennen als wel interpreteren) kunnen handmatig worden toegevoegd, maar spelen dan echter geen rol bij mogelijke ondersteuning tijdens een vergadering. Dergelijke annotaties vergen ook meer tijd dan ‘slechts’ waarnemen.

Handmatig annoteren kan natuurlijk wel een rol spelen bij het toegankelijk maken van het verzamelde audio en video materiaal en zeker ook voor de analyse van het materiaal met als doel modellen te ontwikkelen die automatische herkenning een stap dichterbij brengt. Die analyse van het materiaal kan ook met ‘machine learning’ technieken aangepakt worden [10,19,20].

Handmatige annotatie gaat met behulp van annotatiegereedschap. Op grond van verworven kennis kan annotatiegereedschap steeds intelligenter worden en kunnen de taken van de menselijke annotator verlicht worden. Door ons ontwikkeld gereedschap vergemakkelijkt het annoteren van dialogen en van het observeren van videomateriaal. Dit gereedschap is ontwikkeld met gebruikmaking van de NITE XML Toolkit (NXT), gereedschap dat annotatie, vertoon, en analyse van multi-modale corpora mogelijk maakt. In NXT kan geannoteerde data ook bevraagd worden. In het videomateriaal kunnen bijvoorbeeld zaken als kijkrichting, lichaamshouding, hoofdoriëntatie, en attentieoriëntatie geannoteerd worden. Duidelijk is, dat het nuttig is onderscheid te maken tussen annotaties gebaseerd op observaties (bijvoorbeeld hoofdbewegingen) en annotaties gebaseerd op interpretaties (bijvoorbeeld dialoogacties en emoties). Het is ook nuttig lagen van annotaties te introduceren. In afzonderlijke lagen worden bijvoorbeeld modaliteiten zoals gebaren, gezichtsuitdrukkingen of kijkrichting geannoteerd en in een andere laag, bijvoorbeeld bedoeld voor annotatie van dialoogacties, worden dit soort annotaties gebruikt om de goede dialoogactie te bepalen. Annotatielabels kunnen mede gekozen worden uit een ontologie die de belangrijkste concepten van de toepassing beschrijft.

Al deze aspecten moeten geadresseerd worden bij het onderzoek handmatige annotatie te vervangen door halfautomatische en uiteindelijk volledig automatische annotatie van de belangrijke aspecten van verbale en non-verbale vergaderinteractie. Naast de mogelijkheid om modellen en regels af te leiden uit de geobserveerde en geannoteerde data bestaat er de mogelijkheid om algoritmen te trainen, gebruikmakend van deze data, om belangrijke kenmerken nodig voor interpretatie van interacties en gebeurtenissen automatisch te herkennen.

4. Off- en On-line Ondersteuning van Deelname en Kennisname

4.1 Inleiding

In welke situatie dan ook, als ‘ambient intelligence’ op een of andere wijze ondersteuning kan geven aan onze activiteiten, dan kunnen we daar blij mee zijn. De uitvoering van de activiteiten kan efficiënter gebeuren of kan op een meer aangename wijze plaatsvinden. Willen we terug kunnen kijken op activiteiten die in het verleden hebben plaats gevonden, willen we informatie kunnen opvragen over die activiteiten of willen we die activiteiten als het ware opnieuw beleven, eventueel vanuit een ander gezichtspunt of in iemand anders huid?

Ons uitgangspunt is dat er om tal van redenen behoefte kan zijn om terug te kijken op gebeurtenissen in het verleden waar we op meer of mindere mate bij betrokken zijn. Bij een vergadersituatie ligt dit voor de hand, om de simpele reden dat we niet anders gewend zijn. Er zijn publiek toegankelijke (en vertrouwelijke) notulen opgesteld van vergaderingen, deelnemers hebben vaak hun eigen notities en er is ander materiaal (agenda, deelnemerlijsten, stukken, presentaties) dat een vergadering begeleidt. Door deze zaken onderscheidt een vergadering zich ook van spontane bijeenkomsten, ontmoetingen en gezinssituaties. Dat wil zeggen, bijeenkomsten waar noch een agenda, noch belangen en noch in te vullen rollen opvallend belangrijk zijn. Er wordt niet noodzakelijkerwijs toegewerkt naar informatieoverdracht of een te nemen beslissing.

Een vergadersituatie, een kantooromgeving of een leeromgeving wijkt belangrijk af van een 'ambient intelligence' thuisomgeving. Maar ook die thuisomgeving kan vragen om real-time ondersteuning of om de mogelijkheid op gebeurtenissen terug te kijken (of om er op afstand op een of andere wijze aan deel te nemen). We willen kunnen terugkijken op ervaringen, op een of andere wijze eerdere ervaringen opnieuw beleven, of ook als we niet fysiek aanwezig kunnen zijn, op een of andere wijze deelnemen aan voor ons belangrijke gebeurtenissen in onze thuisomgeving. Dagboeken, knipsel- en fotoalbums, dia's en videomateriaal, e-mails, chat-deelname en telefoontjes geven de behoefte weer. Een ieder heeft zijn eigen tekst/foto/video/web archief over wat hij of zij in een afgelopen periode heeft meegemaakt en webproviders maken het mogelijk die informatie met anderen te delen. Privé-archieven worden openbaar gemaakt. Persoonlijke notities verschijnen als blogs (chronologische publicatie van persoonlijke gedachten en belevenissen) op het Web. Registratie van beeld en geluid in sociale omgevingen (zie bijvoorbeeld [2] en het vervolgens via het web toegankelijk maken van dergelijk materiaal is een eerste stap naar het continu registreren en het tegelijkertijd of daarna toegankelijk maken van deze multimedia informatie [26].

In professionele omgevingen (bijvoorbeeld een kantooromgeving) ligt de nadruk in eerste instantie op andere 'toepassingen'. Wat draagt de intelligente omgeving bij aan het gemak om een kopie of een scan te maken van een document, om de voor mijn collega relevante informatie in een voor hem aantrekkelijke visualisatie toegankelijk te maken, om gezamenlijk aan een document te werken of om gezamenlijk aan een tafel te zitten waarbij mijn collega echter enkel virtueel aanwezig is? Ook hier, beeld en geluid, maar ook informatie van andere bronnen, dienen geïntegreerd te worden in een betekenisvolle representatie van waaruit gebruikers reactief of proactief van informatie kunnen worden voorzien of op grond waarvan de karakteristieken van de omgeving worden aangepast aan de wensen van de gebruiker(s).

Dat laatste zal minder het geval zijn bij publieke ruimten, maar ook daar zal de voor de gebruiker toegankelijke virtuele publieke ruimte aangepast kunnen worden aan zijn of haar wensen. In het algemeen zal in een publieke ruimte minder sprake zijn van directe interactie tussen de aanwezigen, zoals in een thuissituatie of in een kantooromgeving. Er is niet noodzakelijkerwijs een relatie tussen aanwezigen in een straat, op een plein of in een bibliotheek die ondersteund dient te worden. Die kan er natuurlijk wel zijn. Het kan interessant zijn te weten of een kennis op weg is naar hetzelfde evenement of misschien naar iets wat veel interessanter is dan een oorspronkelijk doel. Ook in een museum is het de moeite waard te weten waar de belangstelling van andere bezoekers of van gelijkgezinden naar toe gaat. Het is de omgeving die deze informatie kan verzamelen, kan interpreteren en manipuleren, toegankelijk kan maken en op grond van dergelijke informatie proactief kan optreden. Non-verbaal gedrag in publieke ruimten is bijvoorbeeld onderwerp van onderzoek in [7]. Het door Goffman bestudeerde gedrag is niet zozeer gericht op interactie, maar er kan wel sprake zijn van het afgeven van signalen aan de omgeving of andere aanwezigen. En daarop kan natuurlijk gereageerd worden. Bespeurt de omgeving onzekerheid in het gedrag van een aanwezige? Moet de aanwezige ergens op attent gemaakt worden?

Er is een keerzijde en die heeft niet zozeer te maken met ondersteuning als wel met toezicht en veiligheid. De omgeving is in staat aanwezig te behoeden voor onveilige of anderszins ongewenste situaties. Als je die straat ingaat, houd er dan rekening mee dat je met die en die situatie kan worden geconfronteerd. Natuurlijk, ook van de kant van de beheerder kan sprake zijn van het detecteren en voorkomen van ongewenste situaties in intelligente publieke ruimten. In het algemeen, maar dit is ook al aan de orde gekomen in sectie 2 van dit artikel, is het zinvol verschillende gezichtspunten bij het beschouwen van activiteiten in intelligente omgevingen te onderscheiden: bezoekers, gebruikers, aanwezigen, beheerders, onderhouders, eigenaars, etc.

In de volgende subsectie gaan we terug naar vergaderingen. Vrijwel alle in deze inleiding genoemde aspecten van verbale en non-verbale interactie en de rol van omgevingen zijn terug te vinden in intelligente vergaderomgevingen. Aan de andere kant, vergaderomgevingen zijn beperkter en doelgerichter dan thuis-, kantoor- en publieke omgevingen. In een tamelijk onontgonnen gebied als intelligente omgevingen bieden ze meer houvast aan onderzoekers wat betreft het analyseren van gedrag, het opstellen van modellen en het ontwerpen van proactief gedrag door de omgeving. Vragen die in de volgende subsecties aan de orde komen hebben betrekking op het geven van vergaderassistentie, het willen teruggrijpen op eerder verzamelde kennis en het verzamelen en op multimediale wijze toegankelijk maken van dergelijke kennis.

4.2 Vergaderinformatie: Grasduinen, Vragen en Presentatie

Het is uitermate interessant om er achter te komen wat de deelnemers aan een vergadering na afloop terug willen 'zien' van die vergadering. Waar zou belangstelling voor kunnen bestaan? Notulen opgesteld door een waarnemer (secretaris), een video van de bijeenkomst, de mogelijkheid vragen te stellen (inclusief het vragen van een samenvatting die samengesteld wordt op grond van interesses van de bevrager), een heropvoering van de vergadering rekening houdend met de belangstelling van de bevrager? Erg veel onderzoek naar wat voor soort informatie vergaderdeelnemers of andere belangstellenden na afloop van een vergadering en eventueel ter voorbereiding van een volgende vergadering zouden willen vragen of graag ter beschikking zouden willen hebben is nog niet gedaan. Dit soort gebruikerswensen is een belangrijk onderwerp in het huidige AMI project (zie ook Sectie 2).

Gebruikerswensen kunnen enigszins achterhaald worden door te kijken naar de huidige praktijk van hun activiteiten en door hun te confronteren met de nieuwe technologie en een gebruikersevaluatie te houden (gebruikmakend van use cases, interviews, het laten uitvoeren van taken). De use cases die in het AMI project worden gebruikt hebben betrekking op het vinden van informatie uit in het verleden bezochte vergaderingen, het checken van een niet bezochte vergadering, het tijdens een vergadering geïnformeerd worden over relevante informatie van een vorige vergadering en het 'bijgepraat' worden over het eerste deel van een vergadering waarvoor je te laat bent. Ook zijn er use cases ontworpen voor het op afstand bijwonen van een vergadering.

De interface die beschikbaar wordt gesteld aan gebruikers van vergaderinformatiesystemen moet het mogelijk maken te browsen door vergadermateriaal. Verschillende typen browsers kunnen worden onderscheiden [25]: Browsers met een focus op audiomateriaal, met een focus op videomateriaal, 'artefact' browsers (persoonlijke notities, documenten, sheets) en discourse browsers. Bij het laatste, waarbij ook de binnen het AMI project ontwikkelde Ferret browser [28] hoort, is sprake van afgeleide data die het mogelijk maakt op discourse niveau audiovisuele vergaderinformatie te onderzoeken. Voor de vergelijking van de kwaliteiten van browsers is een zogenaamde 'browser evaluatie test' [19] geïntroduceerd.

Naast het browsen moet het natuurlijk mogelijk zijn om informatie op te vragen (zie ook de vragen genoemd in sectie 3.3). Jaimes et al. [8] beschrijven bijvoorbeeld een implementatie van

een systeem dat gebruikers helpt om op een aansprekbare wijze vage herinneringen over een voorbije vergadering te formuleren. Binnen AMI is het uitgangspunt om vragen in termen van de multi-modale invoermedia te kunnen formuleren en antwoorden ook via multimedia uitvoer te kunnen presenteren. Het automatisch genereren van een multimodale samenvatting van een vergadering [6] is ook binnen AMI onderwerp van onderzoek. Dit is niet wezenlijk verschillend van het construeren van een antwoord op een vraag. Het gezichtspunt van de gebruiker dient richting te geven aan het te presenteren antwoord. AMI-onderzoekers in Brno proberen samenvattingen te genereren die zijn samengesteld uit opgenomen videomateriaal. Evenals bij tekstsamenvattingen is hierbij sprake van het niet altijd gladjes samenvoegen van fragmenten die op een of andere wijze gerelateerd zijn aan de vraagelementen.

4.3 Real-time Ondersteuning van Vergaderdeelnemers

Natuurlijk kan al het gereedschap dat het mogelijk maakt achteraf een vergadering nog eens door te nemen en er vragen over te stellen ook zinvol gebruikt worden tijdens een vergadering. Dat wil zeggen, multimedia notulen van een vorige vergadering zijn beschikbaar en kunnen geraadpleegd worden. Evenzo, ervan uitgaande dat de multimedia notulen op zijn minst deels real-time vervaardigd kunnen worden, kan ook teruggekeken worden op een voorgaand gedeelte van een lopende vergadering. Daarnaast kan natuurlijke proactief informatie en hulp worden geboden aan de deelnemers tijdens de vergadering. In het lopende AMI project is er, afgezien van ons eigen onderzoek, geen directe aandacht voor deze mogelijkheid. De daarvoor te ontwikkelen intelligentie wijkt niet essentieel af van wat nodig is voor het achteraf intelligent bevragen en doorzoeken van vergaderinformatie. Intelligente observaties over het verloop van een vergadering kunnen zowel na afloop als tijdens een vergadering een rol spelen bij het begrijpen van het verloop van de discussie, het hoe en waarom van beslissingen, en de waarde die gehecht moet worden aan de beslissingen. Tijdens een vergadering kan dergelijke informatie een rol spelen bij de loop die de vergadering neemt. Automatisch verkregen informatie over de lopende vergadering kan beschikbaar komen voor enkel de voorzitter van de vergadering en voor afzonderlijke deelnemers (exclusieve informatie of algemene informatie gefilterd op voorkeuren), maar ze kan ook ter beschikking komen voor de gehele vergadering.

Relevante informatie die aangewend kan worden om de loop van de vergadering te beïnvloeden wordt verkregen door waarnemen en interpreteren. Dit dient in het ideale geval te gebeuren in de context van kennis van het domein en het doel van de vergadering en kennis van de afzonderlijke deelnemers en hun doelen. Met andere woorden, de intelligente omgeving dient te kunnen beschikken over die contextkennis. We kunnen ons ook wat bescheidener opstellen. Sommige zaken kunnen gemeten worden en het beschikbaar maken van die gegevens (zelfs het feit dat ze beschikbaar gemaakt kunnen worden) op zich kan de vergadering beïnvloeden en mede in goede banen leiden.

Automatische vergaderassistenten kunnen proberen vanuit verschillende gezichtspunten informatie te verzamelen. Zinnige, maar nog ontbrekende kennis tijdens een vergadering kan bijvoorbeeld van het Web gehaald worden, op grond van tijd en vergadermoeheid kunnen suggesties gegeven worden betreffende een pauze of een verandering van de agenda, gegevens over de inbreng van de verschillende deelnemers kunnen verzameld worden en ook, maar dan zitten we op het niveau van de interpretatie van de inhoudelijke bijdragen, gegevens over tekortkomingen in het redeneerproces dat moet leiden tot beslissingen.

Een voorzitter kan worden geholpen met informatie over de inbreng van de verschillende deelnemers. Iemand overschrijdt zijn tijd, niet iedereen komt aan het woord, er wordt teruggekeerd naar een onderwerp dat al afgehandeld is, een bijdrage levert geen nieuwe informatie, een onderwerp dreigt niet aan de orde te komen, etc. Ook dit zijn waarneembare

zaken die door assistenten waargenomen kunnen worden. Voor het waarom van wat waargenomen wordt is meer kennis nodig, maar ook kan een voorzitter of de vergadering zelf met deze informatie haar voordeel doen en er een eigen interpretatie aan geven. Een voorzitter kan dergelijke informatie gebruiken om het vergaderproces te sturen en te verbeteren. Hij kan, om een voorbeeld te noemen, selectief omgaan met het geven van het woord aan deelnemers of het toestaan van interrupties.

In het voorgaande is het uitgangspunt dat het nemen van beslissingen op grond van automatisch of andere wijze verkregen informatie wordt overgelaten aan de voorzitter of aan de vergaderdeelnemers. De informatie verzameld door de vergaderassistenten kan geraadpleegd worden. Bijvoorbeeld, in DiMicco [3,4] wordt een systeem beschreven dat tijdens een vergadering via spraakherkenning de verbale uitingen (en dus ook de frequentie ervan) van de deelnemers zichtbaar maakt voor iedereen. Het systeem heeft een voorkeur voor uitingen van deelnemers die minder prominent aan het woord zijn en voor uitingen die nieuwe informatie inbrengen. Experimenten hebben duidelijk gemaakt dat een dergelijk systeem het gedrag van vergaderdeelnemers beïnvloedt. De meer dominante deelnemers reduceren hun inbreng, en de meer passieve deelnemers gaan over tot het vaker nemen van het woord.

Vergaderassistenten kunnen natuurlijk ook zelf de bevoegdheid krijgen om informatie te melden (eventueel gefilterd naar interesses) of om op grond van beschikbare informatie beslissingen te nemen die het vergaderproces beïnvloeden. Op deze wijze kan er ook sprake zijn van een ontwikkeling in de richting van een intelligente, virtuele, voorzitter die de taken van een menselijke voorzitter ondersteunt en uiteindelijk overneemt.

In het vergaderproces kunnen we inhoud, organisatie en sociale aspecten onderscheiden. Als we kijken naar mogelijke virtuele vergaderassistenten dan kunnen we ook assistenten met dit soort specialisaties onderscheiden. Neem is een project van de University of Colorado [5] waarin drie agents voor het geven van assistentie worden onderscheiden: een informatie-agent (o.a. voor het assisteren bij het verkrijgen van de nodige informatie, bijvoorbeeld door een webzoekactie), een sociale agent (o.a. verantwoordelijk voor sfeerbevoorgerende acties) en een organisatorische agent (o.a. voor het bewaken van de tijd en de agenda). Het gedrag van deze agents is mede gemodelleerd aan de hand van de 'Social Interaction Systems' theorie van Robert Freed Bales. Het is de moeite waard na te denken over de wijze waarop en aan wie dergelijke agents dienen te rapporteren. Alle door dergelijke agents beschikbaar gekomen informatie kan doorgegeven worden aan de voorzitter. De informatie kan echter ook gedeeld worden door de vergaderdeelnemers, waardoor ze een indruk krijgen van hun eigen gedrag en dat van anderen. Wanneer men zich bewust wordt van dit gedrag laat men zich er ook door beïnvloeden. Zie opnieuw het werk van DiMicco [4].

Zonder uitpuddend te zijn, geven we in het kort de onderwerpen aan waaraan op dit moment onderzoek wordt verricht en waarvan de resultaten gebruikt kunnen worden om de omgeving en de vergaderassistenten te voorzien van relevante informatie. Met andere woorden, waar gaat op dit moment onze aandacht, dat wil zeggen, van de Human Media Interaction onderzoeksgroep van de Universiteit Twente, naar toe als het gaat om het 'waarnemen en interpreteren'?

Veel kan gemeten worden en sommige zaken kunnen direct geobserveerd worden zonder dat sprake hoeft te zijn van een 'diep' onderliggend model van herkenning of interpretatie. Dit komt voor een deel terug in de volgende sectie van dit artikel. 'Talkativity', de tijd die iemand aan het woord is, is te meten (minder eenvoudig wellicht dan op het eerste gezicht lijkt; er kan overlap zijn tussen verschillende sprekers en de sprekers dienen geïdentificeerd te worden door het spraaksysteem). Het herkennen van sleutelwoorden tijdens spraakherkenning op grond waarvan retrievalacties gepleegd kunnen worden is een ander voorbeeld van direct herkenbare en bruikbare observaties. Iets dergelijks geldt voor beurtwisselingen. Niet direct observeerbaar

gedrag vereist interpretatie van de observatie en vaak ook het combineren van kenmerken afkomstig van verschillende modaliteiten. In plaats van herkennen van sleutelwoorden komen we dan bijvoorbeeld terecht op het herkennen van gespreksonderwerpen en, kijkend naar inhoud en non-verbaal gedrag van deelnemers, van wisseling van onderwerp. In het bijzonder hebben we in ons onderzoek bij dit niet direct observeerbaar gedrag aandacht besteed aan onderwerpen als het bepalen van de spreker [20], bepalen van de aangesproken persoon of personen en het detecteren van dominantie [19]. Bayesiaanse netwerken werden veelal aangewend om uit de beschikbare gegevens in de AMI corpora niet direct observeerbaar gedrag af te leiden.

5 Vertaling naar Virtual Reality en Regeneratie in Virtual Reality

Technologie voor het herkennen van verbale en non-verbale vergaderactiviteit, voor het herkennen van plaats en plaatsverandering, voor het identificeren van personen, etc., maakt het mogelijk, ook in onvolmaakte vorm, ondersteuning te verlenen aan menselijke activiteiten in vergadersituaties. Het niveau van ondersteuning zal toenemen wanneer theorie en modellen verder tot ontwikkeling worden gebracht. Mogelijke real-time ondersteuning houdt in dat multimodale input voor de omgeving real-time getransformeerd wordt naar multimedia output die beschikbaar komt voor de vergaderdeelnemers. Natuurlijk kan die multimedia presentatie ook beschikbaar gesteld worden aan deelnemers op afstand of aan toeschouwers. En gedurende die transformatie kan de informatie die uit het waarnemen is verkregen verrijkt worden met kennis beschikbaar vanwege de geschiedenis van de interactie, kennis over de deelnemers aan de interactie, en kennis ondergebracht in ontologieën die ten grondslag liggen aan het vergaderen en aan vergaderthema's.

Een mogelijke representatie die op deze wijze verkregen kan worden is een virtual reality weergave van de vergaderactiviteiten. Met andere woorden, er is een (bij voorkeur real-time) omzetting van gebeurtenissen en interacties in de fysieke vergaderzaal naar de virtuele representaties van gebeurtenissen en interacties in een virtuele vergaderzaal. De mogelijkheid om dit te doen en het nut hiervan wordt hieronder besproken. Het idee past in eerdere observaties over het delen van fysieke en virtuele ruimten en het past ook in meer algemene observaties over 'augmented' en 'mixed reality' [15]

Aangezien bovenstaande inhoudt dat we in staat moeten zijn mensen te representeren in de virtuele vergaderruimten die we creëren, dienen we ook hun virtuele replica's (avatars, virtuele personages, belichaamde conversationele agents, 'talking faces', etc.) in onze beschouwingen te betrekken. Het kan daarbij gaan om replica's van de actuele vergaderdeelnemers, maar het is natuurlijk ook mogelijk om belichaamde agents te ontwerpen die een specifieke taak hebben in de omgeving en die met vergaderdeelnemers of hun replica's kunnen communiceren. Voordat we mogelijke rollen van virtuele mensen in virtuele vergaderruimten gaan bediscussiëren beschouwen we eerst de volgende twee vragen.

- Waarom zouden we virtuele representaties van gebeurtenissen en interacties willen hebben?
- Hoe kunnen we die virtuele representaties realiseren?

Zodra we deze twee vragen beantwoord hebben keren we terug naar de rol die virtuele mensen in de virtuele vergaderomgeving kunnen spelen.

5.1 Waarom Willen we een Virtuele Weergave?

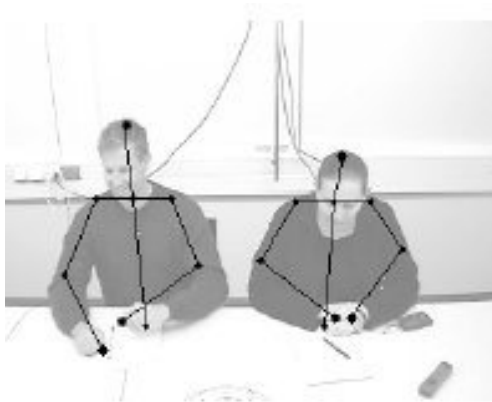
Er zijn tal van redenen om geïnteresseerd te zijn in het realiseren van een transformatie van vergaderactiviteiten naar een virtuele weergave van die activiteiten in een virtuele vergaderruimte (zie ook [18]).

- In de eerste plaats, een dergelijke transformatie maakt een 3D presentatie en een ‘replay’ mogelijk van multimedia informatie die verkregen is door het opnemen via video en audio van een vergadering. Afhankelijk van de ‘state of the art’ van spraak en beeldverwerking (herkenning en interpretatie) kan gedacht worden aan een replay die gebaseerd is op een handmatige annotatie, een replay waaraan deels een handmatige en deels een automatische annotatie en interpretatie aan ten grondslag ligt, en een replay die puur gebaseerd is op volledig automatisch verkregen interpretaties en in real-time de vergadering volgt. Vanzelfsprekend, als de vergaderomgeving de intelligentie heeft om de gebeurtenissen te interpreteren, kan het ook gebeurtenissen transformeren en manipuleren op een voor de deelnemers zinvolle wijze en de verkregen kennis op andere nuttige wijzen ter beschikking stellen (antwoord geven op vragen, samenvattingen opstellen, extra informatie - bijvoorbeeld van vorige vergaderingen – toevoegen, metadata visualiseren, etc.);
- In de tweede plaats, het transformeren van annotaties, of ze nu handmatig of automatisch zijn verkregen, naar gedrag in een virtual reality omgeving kan gebruikt worden voor de evaluatie van de annotaties en de annotatieschema’s en voor de evaluatie van resultaten verkregen uit methoden voor machinaal leren. Computationale modellen voor verbale en non-verbale interactie, multi-party interactie, sociale interactie, groepsinteractie en, kijkend naar ons domein van vergaderactiviteiten, modellen van vergaderinteractie op het niveau van een individu of de volledige vergadering zijn niet beschikbaar of slechts beschikbaar voor het beschrijven van tamelijk oppervlakkige fenomenen van groepsactiviteiten. De virtuele vergaderruimte die wij introduceren kan dienen als testbed voor het ontwerpen en valideren van modellen van sociale interactie aangezien we in deze representatie het waarneembaar maken van de verschillende factoren in de verbale en non-verbale interactie (stem, kijkrichting, gebaren, gezichtsuitdrukkingen, etc.) tussen vergaderdeelnemers kunnen sturen en daardoor die afzonderlijke factoren en combinaties ervan onderwerp van studie laten zijn voor het begrijpen van sociale interactie en sociaal gedrag;
- In de derde plaats, een virtual reality vergaderomgeving kan gebruikt worden om participatie op afstand op een natuurlijke wijze mogelijk te maken. Ervan uitgaande dat we niet alle details van natuurlijke interactie virtueel waarneembaar kunnen maken, blijft de vraag welke elementen van multi-party interactie gedurende een vergadering gerepresenteerd moeten worden in een virtuele vergadering om die zo natuurlijk mogelijk te laten plaatsvinden. De bovengenoemde testbed-functie van een virtuele vergaderruimte zal meehelpen om er achter te komen welke verbale en non-verbale signalen belangrijk genoeg zijn om op een of andere wijze gemedieerd te moeten worden.

5.2 Hoe een Virtuele Representatie te Realiseren?

De informatie die nodig is om een virtuele representatie van vergaderactiviteiten te realiseren kan in real-time verkregen worden van de registratie van activiteiten in plaatsvindende vergaderingen (real-time tracking van alle activiteiten van vergaderdeelnemers), van handmatige annotaties, combinaties van handmatige en automatisch verkregen annotaties en volledig automatisch verkregen herkenning en interpretatie van activiteiten. De moeilijkheid en vrijwel onmogelijkheid om allesomvattende modellen te ontwerpen nodigt uit om machinale leeralgoritmen toe te passen die vanuit handmatige of machinaal verkregen kenmerken proberen vergaderinteracties te generaliseren tot vergaderinteractiemodellen.

Uitgaande van perfecte (niet-ambigue en voldoende gedetailleerde) annotatieschema’s en perfecte annotatoren mogen we een perfecte regeneratie van een vergadering verwachten of een perfecte input voor leeralgoritmen. In de praktijk is die annotatie niet perfect en voldoende gedetailleerd en bovendien zijn we in de eerste plaats geïnteresseerd in een automatisch verkregen annotatie die



Figuur 2. Afbeelden op avatarskelet

zichtbaar te kunnen maken via animaties van 3D avatar representaties in een virtuele vergaderruimte. Waargenomen lichaamshoudingen, gebaren en afgeleide gewrichtsrotaties worden afgebeeld op een 3D H-Anim avatarskelet (Figuur 2). Dit kan in real-time gebeuren en kan leiden tot voldoende overtuigende animaties van lichaamshoudingen en gebaren in een virtual reality omgeving.

Ten behoeve van de uitwerking van onze ideeën over kijkrichting tijdens een vergadering hebben we ook gekeken naar de omzetting van hoofdoriëntatie van vergaderdeelnemers naar hoofdoriëntatie van avatars. Deze omzetting speelde een rol bij het bestuderen van kijkgedrag van vergaderdeelnemers. Ten behoeve van deze omzetting werden de vergaderdeelnemers uitgerust met 'flock-of-birds' sensoren (Figuur 3) om hun hoofdbewegingen te kunnen registreren. De omzetting naar de virtuele vergaderruimte, gebruik makend van de automatische herkenning van lichaamshouding, gebaren en hoofdbewegingen is weergegeven in Figuur 4. In deze figuur is de virtuele vergaderruimte ingebed in een van onze annotatietools. Informatie over de huidige spreker (zie aanduiding boven het hoofd) en de persoon die geadresseerd wordt (zie pijl van spreker naar geadresseerde) is gevisualiseerd.

Real-time herkenning van vergaderactiviteiten en omzetting naar een virtuele vergaderruimte staat toe dat geografisch verspreide deelnemers via audio en video deel gaan uitmaken van een vergadering waarin ze virtueel gerepresenteerd worden. Hoewel in het AMI project ook karakteristieken van individuele vergaderdeelnemers geregistreerd worden (bijvoorbeeld via camera's die recht tegenover iedere deelnemer op de vergadertafel staan), is er op dit moment in ons onderzoek nog geen sprake van de omzetting van dergelijke gedetailleerde informatie naar de virtuele wereld.

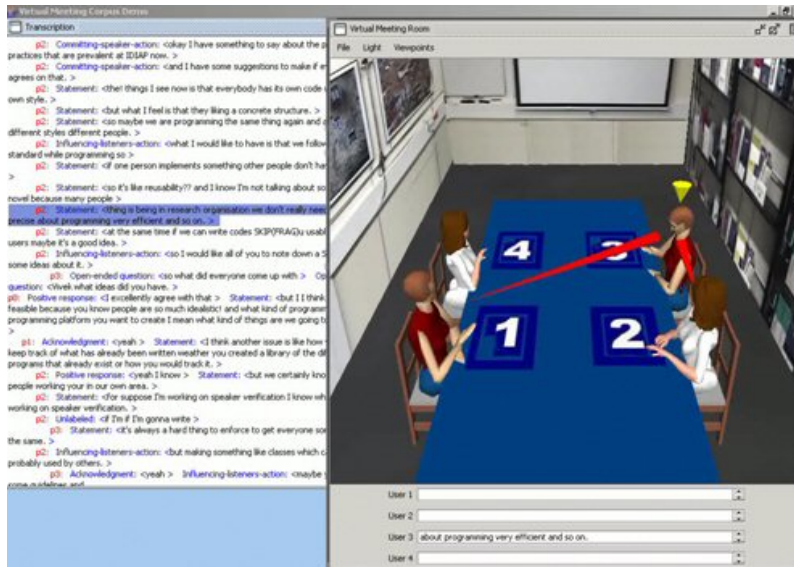


Figuur 3. 'Flock-of-Birds' sensoren

in meer of mindere mate een voldoende gedetailleerde regeneratie van een vergadering in een virtuele weergave mogelijk maakt. Hoe vollediger een automatisch verkregen annotatie kan zijn, hoe vollediger een real-time regeneratie in virtual reality mogelijk is. In ons onderzoek hebben we gebruik gemaakt van beeldextractie-technieken (silhouetextractie, het volgen van gebieden in het beeld met huidskleur) voor het herkennen van lichaamshoudingen, gebaarherkenning en het volgen van bewegingen. Dat wil zeggen, algoritmen worden aangeroepen om lichaamshoudingen en gebaren van vergaderdeelnemers (en zelfs van voordrachtgevers) te halen uit video-opnamen. Het doel daarbij is hun gedrag

5.3 Over de Rol van Virtuele Mensen

De hierboven omschreven technologie laat toe dat geografisch verspreide personen virtueel, verbaal en non-verbaal aanwezig



Figuur 4. Virtuele vergaderdeelnemers

correcte weergave van het gebeuren kunnen we ook kijken hoe we een mogelijke weergave kunnen aanpassen tot een meer wenselijke weergave. We kunnen de animaties manipuleren, gebaren toevoegen of versterken, een lichaamshouding veranderen van ongeïnteresseerd naar geïnteresseerd, etc. Dergelijke gemanipuleerde animaties kunnen het vergadergedrag van deelnemers die niet de fysieke signalen van de interacties ontvangen beïnvloeden. Zo wordt bijvoorbeeld in [17] gerapporteerd over de beïnvloeding van sprekers door een virtueel publiek dat via non-verbale signalen meer of minder positief reageert op een gegeven voordracht (zie Figuur 5 voor een publiek dat uiterst geïnteresseerd is (rechts) of zich tamelijk verveelt (links)). Ook belangrijk, bijvoorbeeld in het geval van het virtueel deelnemen aan een vergadering, is de mogelijkheid om de verkregen signalen te veranderen voordat we ze vertalen naar een virtuele omgeving. We komen daarmee op het terrein van ‘Transformed Social Interaction’ (TSI) [1], het filteren en modificeren van non-verbaal sociaal gedrag. In het bijzonder kan daarbij naar signalen gekeken worden die bijdragen aan efficiënt vergadergedrag. Bijvoorbeeld, aanpassingen aan het kijkgedrag kan het beurtwisselingsproces versoepelen en het kan duidelijker maken wie aangesproken wordt. Dit moet het systeem dan wel weten, bijvoorbeeld met behulp van methoden voor het automatisch bepalen van de geadresseerde op grond van kenmerken die onder meer komen uit het beurtwisseling- en aankijkgedrag.

De vergaderassistentie kan natuurlijk ook toebedeeld worden aan virtuele, belichaamde agenten. In dat geval zijn ze zichtbaar aanwezig in de omgeving en ‘vragen’ er om aangesproken te worden. Een dergelijke agent kan zelfs een leidende rol krijgen door hem als virtuele voorzitter of als hulp van een menselijke voorzitter op te voeren. Misschien niet alles begrijpend, maar wel



Figuur 5. Geïnteresseerd en ongeïnteresseerd publiek

kunnen zijn in een gedeelde virtuele ruimte in real-time. De virtuele vergaderdeelnemers kunnen we laten imiteren wat er gebeurt in de fysieke vergaderruimte. Dit is wat wij met onze transformaties realiseren. Er is niet altijd een noodzaak om dit te doen en bovendien laat state-of-the-art technologie en de wenselijkheid om de omzetting real-time te doen een volledige omzetting in alle details niet toe. Aan de andere kant, in plaats van te kijken naar een

duidelijk de tijd en de afhandeling van de agenda in de gaten houdend. Vergaderdeelnemers die niet de volledige vergadering kunnen bijwonen kunnen besluiten hun avatar in een neutrale positie of in een luisterpositie te zetten op momenten dat ze niet zelf de vergadering volgen. Onderzoek naar (non-verbaal) luistergedrag wordt bijvoorbeeld gerapporteerd in [9]. In zo'n geval kan de eigenaar van de avatar een waarschuwing krijgen op het moment dat er een onderwerp aan de orde komt dat zijn aandacht vereist en daarmee de controle van de avatar overnemen.

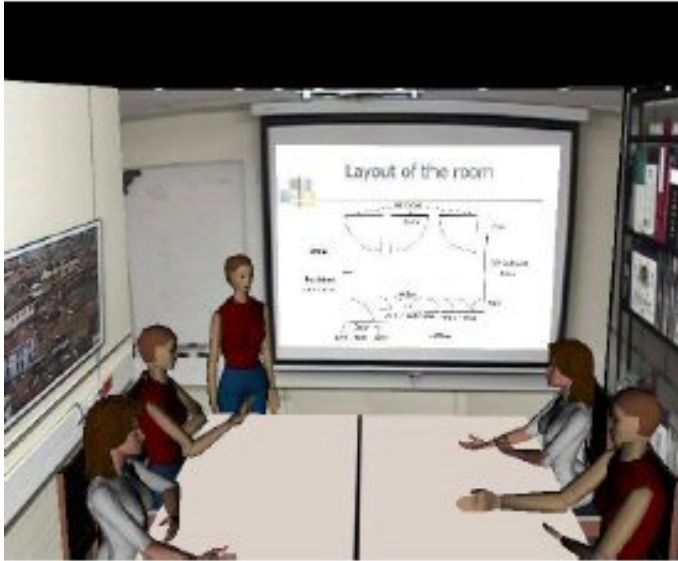
6 Conclusies en Verder Werk

Er zijn zeker meer onderzoeksprojecten waar men zich bezig houdt met het computationeel modelleren van vergaderingen, of, meer bescheiden, het ontwikkelen van gereedschap ter ondersteuning van vergaderingen, het voorbereiden van vergaderingen en het achteraf toegankelijk maken van geregistreerde vergaderactiviteiten. Een bekend voorbeeld is het ICSI project [12], maar in tegenstelling tot de meer recente projecten die uitgaan van multimodale registratie, herkenning en interpretatie wordt bij ICSI enkel audio opgenomen.

Een project dat tamelijk gerelateerd is aan M4 is het Meeting Room project van Carnegie Mellon University [22]. Het houdt zich bezig met het opnemen en browsen van vergaderinformatie gebruikmakend van audio- en videodata. Nauw gerelateerd aan AMI is ook het werk dat gedaan wordt aan de University of California in San Diego. Er is zowel onderzoek op het gebied van intelligente vergaderzalen [11] als op het terrein van intelligente omgevingen in het algemeen (AVIARY: Audio-Video Interactive Appliances, Rooms and sYstems) [23].

Een ander 6^{de} Raamwerk IP gerelateerd aan AMI is het CHIL (Computers in the Human Interaction Loop) project. Ook in dit project staat het modelleren van multimodale interacties centraal, maar in plaats van te concentreren op vergaderingen richten de huidige activiteiten zich vooral op het bestuderen van presentaties en de interactie tussen een presentator en zijn publiek [27]. Wellicht daarom is CHIL ook meer gericht op het geven van real-time ondersteuning. Beide projecten (AMI en CHIL) vallen onder de EU pogingen om onderzoek op het gebied van multimodaliteit te bevorderen. Binnen CHIL wordt een uitgebreid corpus van voordrachten verzameld. Dat corpus moet meehelpen om een aantal software modules te ontwikkelen. Zo is er sprake van een Attention Cockpit, een agent die de aandacht van het publiek moet volgen en feedback dient te geven aan de voordrachtgever. Zijn er nog mensen geïnteresseerd, is er iemand die een vraag wil stellen? Het monitoren van het publiek (oogcontact, gezichtsuitdrukkingen, lichaamstaal, het stellen van vragen) is onderdeel van het onderzoek. Informatie afkomstig van spreker (spraaksignaal) en publiek (vragen) maakt het mogelijk te denken aan het automatisch veranderen van de volgorde van sheets gedurende een PowerPoint presentatie.

Vergaderassistenten worden ook ontwikkeld binnen het CALO (Cognitive Agent that Learns and Organizes) project. Net als AMI houdt ook dit project, met als deelnemende universiteiten Stanford University, Carnegie Mellon University, MIT, Oregon Health State University en Georgia Tech, zich bezig met datacollectie, multimodale perceptie en multimodale integratiemethoden. De te ontwikkelen agent wordt verondersteld multimodale mens-mens en mens-computer interactie te begrijpen, waarbij mensen en computers (agents, robots, intelligente objecten) kantoorgerelateerde werkzaamheden uitvoeren, inclusief het bijwonen van vergaderingen. De agent kan daarbij een rol als vergaderassistent op zich nemen, de vergadering coördineren, real-time ondersteuning geven aan deelnemers en op zodanige wijze informatie verzamelen dat naderhand, onder andere gebruikmakend van visualisatiegereedschap, de vergaderinformatie toegankelijk kan worden gemaakt [<http://www.ai.sri.com/project/CALO>]. Ook wordt een poging ondernomen om een dergelijke assistent te laten leren van zijn ervaringen



Figuur 6. Presentaties in de virtuele vergaderzaal

bij het integreren van informatie die vanuit verschillende bronnen bij elkaar wordt genomen voor interpretatie [30].

Zoals uit dit (onvolledig) overzicht mag blijken, vergaderingen - of meer in het algemeen situaties waar mensen elkaar ontmoeten en met elkaar communiceren – krijgen veel aandacht in hedendaags Europees en Amerikaans onderzoek op het gebied van multimodale interactie, intelligente en sociale omgevingen en multi-party interactie. Initiatieven in Nederland waar wij op een of andere wijze betrokken zijn hebben betrekking op het registreren en ondersteunen van

vergaderingen van gemeenteraden, vergaderingen van provinciale staten en tweedekamer vergaderingen. Ook wordt gekeken naar mogelijkheden om de ontwikkelde AMI technologie in te zetten bij rechtszittingen. Hoewel ook duidelijk is dat de ontwikkelde technologie ook ingezet kan worden in intelligente thuisomgevingen, educatieve omgevingen en recreatieve omgevingen, dit uitgangspunt is nog niet geëxploreerd.

In toekomstig onderzoek van de Human Media Interaction groep van de Universiteit Twente zal vooral gekeken worden naar formele modellen voor multi-party interactie en de implementatie ervan in virtual reality omgevingen. Die modellering en implementatie heeft ook betrekking op het gedrag van de autonome, semi-autonome en volledig aangestuurde belichaamde agenten die deelnemen aan een vergadering. Ook zal daarbij aandacht geschonken worden aan mogelijke rollen van presentatie-agenten, dat wil zeggen, virtuele agenten die tijdens een bijeenkomst een presentatie geven (zie Figuur 6).

Verantwoording

In dit artikel is in ruime mate gebruik gemaakt van een aantal eerdere (Engelstalige) artikelen geschreven vanuit de onderzoeksgroep HMI (Human Media Interaction) van de Universiteit Twente. Met andere woorden, er is veel dank verschuldigd aan alle leden van HMI die bijdragen aan haar vergaderprojecten. Totstandkoming van dit artikel werd in belangrijke mate ondersteund door het 6^{de} Raamwerk IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-??) van de Europese Unie.

6 Referenties

- [1] J.N. Bailenson, A.C. Beall, J. Loomis, J. Blascovich & M. Turk. Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *Presence*, Vol. 13, No. 4, August 2004, 428–441.
- [2] M. Deutscher, P. Jeffrey & N. Siu. Information capture devices for social environments. In : *Proceedings EUSAI 2004*, LNCS 3295, Springer-Verlag Berlin Heidelberg, 2004, 267-270.

- [3] J.M. DiMicco. Designing interfaces that influence group processes. In *Doctoral Consortium Proceedings of the Conference on Human Factors in Computer Systems (CHI 2004)*, April 2004, 1041-1042.
- [4] J.M. DiMicco. Changing Small Group Interaction through Visual Reflections of Social Behavior. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2001, May 2005.
- [5] C. Ellis & P. Barthelmeß. The Neem dream. *TAPIA '03*, October 2003, Atlanta, Georgia, USA, 23-29.
- [6] B. Erol, D. Lee & J. Hull. Multimodal summarization of meeting recordings. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2003)*, July 2003.
- [7] E. Goffman. *Behavior in Public Spaces: Notes on the Social Organization of Gatherings*. Glencoe: The Free Press, 1963.
- [8] A. Jaimes, K. Omura, T. Nagamine, & K. Hirata. Memory cues for meeting video retrieval. In: Proceedings: 1st ACM workshop on *Continuous Archival and Retrieval of Personal Experiences (CARPE'04)*. ACM Press, 2004, 74-85.
- [9] R. M. Maatman, J. Gratch & S. Marsella. Responsive Behavior of a Listening Agent. ICT Technical Report No: ICT TR 02.2005, 2005.
- [10] I. McCowan, D. Gatica-Perez, S. Bengio, D. Moore, H. Bourlard. Towards Computer Understanding of Human Interactions. In: Proceedings *Machine Learning for Multimodal Interaction (MLMI 2004)*, S. Bengio & H. Bourlard (Eds.), Lecture Notes in Computer Science 3361, Springer-Verlag, Heidelberg, 2005, 56 – 75.
- [11] I. Mikic, K. Huang & M. Trivedi. Activity monitoring and summarization for an intelligent meeting room. In: Proceedings *IEEE Workshop on Human Motion*, Austin, Texas, December 2000, 107-112.
- [12] N. Morgan, D. Baron, S. Bhagat, H. Carvey, R. Dhillon, J. Edwards, D. Gelbart, A. Janin, A. Krupski, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters. Meetings about meetings: Research at ICSI on speech in multiparty conversations. In *Proc of the ICASSP'03*, 2003.
- [13] A. Nijholt. Where computers disappear, virtual humans appear. *Computers and Graphics*, Vol. 28, No. 4, Elsevier, 2004, 465-476.
- [14] A. Nijholt, T. Rist & K. Tuinenbreijer. Lost in ambient intelligence. In: Proceedings *ACM Conference on Computer Human Interaction (CHI 2004)*, ACM, New York, 1725-1726.
- [15] A. Nijholt. Meetings in the Virtuality Continuum: Send Your Avatar. In: Proceedings *Cyberworlds 2005*, november 2005, Singapore, te verschijnen.
- [16] A. Nijholt, R. op den Akker & D. Heylen. Meetings and Meeting Modeling in Smart Environments. *AI & Society*, The Journal of Human-Centred Systems. Springer-Verlag London Ltd, 2005, te verschijnen.
- [17] D.-P. Pertaub, M. Slater & C. Barker. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments* 11 (1), 68-78. Volume 11, Issue 1, 2002, 68 – 78.
- [18] D. Reidsma, R. op den Akker, R. Rienks, R. Poppe, A. Nijholt, D. Heylen & J. Zwiers. Virtual Meeting Rooms: From Observation to Simulation. CD Proceedings: *Social Intelligence Design 2005*, Stanford University, USA, R. Fruchter (Ed.).
- [19] R.J. Rienks & D. Heylen. Dominance detection in meetings, using support vector machines. In: Proceedings of the 2nd Joint Workshop on *Multimodal Interaction and Related Machine Learning Algorithms (MLMI 2005)*. S. Renals & S. Bengio (Eds.), Springer-Verlag, Berlin, 2005. te verschijnen.

- [20] R.J. Rienks, R.W. Poppe & M. Poel. Speaker prediction based on head orientations. In: Proceedings of the Fourteenth Annual *Machine Learning Conference of Belgium and the Netherlands (Benelearn 2005)*, M. van Otterlo, M. Poel & A. Nijholt (Eds.), CTIT, Enschede, The Netherlands, 2005, 73–79.
- [21] R. Rienks, A. Nijholt, D. Reidsma. Meetings and Meeting Support in Ambient Intelligence. Hoofdstuk in: *Ambient Intelligence, Wireless Networking, Ubiquitous Computing*. Th. A. Vasilakos & W. Pedrycz (eds.), Artech House, Norwood, MA, USA, 2005, te verschijnen.
- [22] T. Schultz, A. Waibel, M. Bett, F. Metze, Y. Pan, K. Ries, T. Schaaf, H. Soltau, M. Westphal, Hua Yu and K. Zechner. The ISL Meeting Room System. Proc. of the Workshop on *Hands-Free Speech Communication (HSC-2001)*, Kyoto Japan 2001.
- [23] Trivedi, M, Mikic, I, Bhonsle, S (2000) Active Camera Networks and Semantic Event Databases for Intelligent Environments. *IEEE Workshop on Human Modeling, Analysis and Synthesis* (in conjunction with CVPR), Hilton Head, South Carolina.
- [24] S. Tucker & S. Whittaker. Accessing multimodal meeting data: Systems, problems and possibilities. In Proceedings *Machine Learning for Multimodal Interaction (MLMI 2004)*, S. Bengio & H. Bourlard (Eds.), Lecture Notes in Computer Science 3361, Springer-Verlag, Heidelberg, 2004, 1-11.
- [25] S. Tucker, S. Whittaker & R. Laban. Identifying user requirements for novel interaction capture. In: Proceedings *Measuring Behaviour 2005*, Wageningen, 2005, te verschijnen.
- [26] S. Vemuri & W. Bender. Next-generation personal memory aids. *BT Technology Journal*, Vol. 22, No 4, October 2004, 125-138.
- [27] A. Waibel, H. Steusloff, R. Stiefelhagen & the CHIL Project Consortium. CHIL - Computers in the Human Interaction Loop. 5th Intern. Workshop on *Image Analysis for Multimedia Interactive Services*, April, 2004, Lisbon, Portugal. Zie ook <http://chil.server.de/>.
- [28] P. Wellner, M. Flynn, and M. Guillemot. Browsing recorded meetings with ferret. In: Proceedings *Machine Learning for Multimodal Interaction (MLMI 2004)*, S. Bengio & H. Bourlard (Eds.), Lecture Notes in Computer Science 3361, Springer-Verlag, Heidelberg, 2005, 12-21.
- [29] P. Wellner, M. Flynn, S. Tucker & S. Whittaker. A Meeting Browser Evaluation Test, Proceedings of the *Conference on Human Factors in Computing Systems (CHI '05)*, 2005, 2021-2024.
- [30] M. Wesson & S. Kumar. Demo: A Multimodal Learning Interface for Sketch, Speak and Point Creation of a Schedule Chart. In: Proc. *International Conference on Multimodal Interfaces (ICMI '04)*. 2004. State College, PA.
- [31] S. Whittaker, R. Laban & S. Tucker. Analysing Meeting Records: An Ethnographic Study and Technological Implications. In: Proceedings *Machine Learning for Multimodal Interaction (MLMI 2005)*, Lecture Notes in Computer Science, Springer-Verlag, Heidelberg, te verschijnen.