



KUNGLTEKNISKA HÖGSKOLAN

Royal Institute of Technology
Numerical Analysis and Computing Science

TRITA-NA-D9901 • CID-41, KTH, Stockholm, Sweden 1999

Återkoppling med gester och ansiktsuttryck i virtuella mötesmiljöer

Martin Jonsson



CID
Centre for
User Oriented IT Design

Martin Jonsson

Återkoppling med gester och ansiktsuttryck i virtuella mötesmiljöer

Report number: TRITA-NA-D9901, CID-41

ISSN number: ISSN 1403-073X

Publication date: February 1999

E-mail of author: e92_mjo@elixir.e.kth.se

Reports can be ordered from:

CID, Centre for User Oriented IT Design

Nada, Dept. Computing Science

KTH, Royal Institute of Technology

S-100 44 Stockholm, Sweden

telephone: + 46 8 790 91 00

fax: + 46 8 790 90 99

e-mail: cid@nada.kth.se

URL: <http://www.nada.kth.se/cid/>

Sammanfattning

Återkoppling med gester och ansiktsuttryck i virtuella miljöer

På senare tid har applikationer utvecklats där tredimensionella datorgenererade miljöer, eller *virtuella miljöer*, används som mötesmiljöer. En allvarlig brist hos dessa system, särskilt i jämförelse med videokonferenssystem är bristen på stöd för kommunikation med visuella uttryck. Kroppsspråk, inklusive gester och ansiktsuttryck, är en viktig del av vanlig kommunikation mellan människor. Gester kan användas för att förmedla *återkoppling* till talaren, något som är viktigt för bl.a. turtagning. Återkoppling med gester överför också information om attityder och uppmärksamhet till talaren.

I ett försök att lösa problemet med förmedlingen av återkoppling, utvecklades och implementerades en metod, där användare förmedlar visuell återkoppling med naturliga handgester. Handgesterna fångas av en sensorhandske och tolkas sedan av ett neuralt nätverk för att slutligen översättas till en gest eller ansiktsuttryck hos användarens representation i den virtuella miljön. En utvärdering av denna metod visar att den är snabb och kräver liten uppmärksamhet från användaren.

En användarstudie med en testmiljö visar att gester i virtuella miljöer har effekt i huvudsak på upplevelsen av uppmärksamhet, och i viss mån på upplevelsen av hur väl personer förstår vad som sägs.

Abstract

Gestural feedback in virtual environments

Lately applications have been developed where three-dimensional computer generated environments, or *virtual environments*, are used as meeting rooms. A serious disadvantage of these systems, especially compared to videoconferences, is the lack of support for non-verbal communication. Body language, including facial expressions and gestures, is an important part of common peer to peer communication. Gestures can be used to convey *feedback* to the speaker, something that is important for the process of turn taking.

In an attempt to solve the problem with conveying visual feedback, a method was developed and implemented, where users convey visual feedback with natural hand gestures. The hand gestures are mapped by a sensor-glove and then interpreted by a neural network to finally be transformed to a gesture or facial expression on the representation of the user in the virtual environment. An evaluation of this method shows that it is fast and demands little attention from the user.

A user study with a test environment shows that gestures in virtual environments have effect on the experience of attention, and also to some extent on the experience of comprehension.

Förord

Detta arbete är ett examensarbete i datalogi vid institutionen för Numerisk Analys och Datalogi (NADA), vid Tekniska Högskolan i Stockholm. Examensarbetet som omfattar tjugo högskolepoäng har utförts vid Telia Research AB, avdelning Kommunikationstjänster.Handledare på NADA har varit Sören Lenman och Kerstin Frenckner. Handledare på Telia har varit Ulf Winroth och Bengt Serenius.

Stort tack till mina handledare och särskilt till Bengt Serenius och Ulf Winroth, för all hjälp med användarstudier med mera. Tack också till min rumskamrat Per Nystedt som ofta fått agera bollplank och uppslagsverk. Ett särskilt tack till min flickvän Hanna Sundqvist för hjälp med kritisk granskning och korrekturläsning.

Innehållsförteckning

1	Inledning.....	9
1.1	Arbetets syfte och upplägg.....	9
1.2	Definitioner och förklaringar.....	10
2	Kroppsspråkets funktioner.....	13
2.1	Återkoppling och turtagning.....	13
2.2	Visuella uttryckssätt.....	14
2.2.1	Gestikulation.....	14
2.2.2	Fokus.....	15
2.2.3	Ansiktsuttryck.....	15
2.2.4	Övriga uttryck.....	15
2.2.5	Medvetna och omedvetna gester.....	16
2.3	Kroppsspråkets funktioner i virtuella miljöer.....	16
3	Interaktionsmetoder.....	19
3.1	Direkt överföring av uttryck.....	19
3.2	Tolkad överföring.....	20
3.3	Medveten förmedling av uttryck.....	20
3.4	Automatiskt beteende.....	21
3.5	Förmedla gester med gester.....	21
3.5.1	Gestigenkänning.....	23
3.6	Summering av interaktionsmetoder.....	24
4	En testmiljö med stöd för visuell återkoppling.....	25
4.1	Val av uttryck och interaktionsmetod.....	25
4.2	Implementation av uttryck.....	26
4.2.1	Teknisk plattform.....	26
4.2.2	Utformning av avatar.....	26
4.2.3	Animationen av gesterna.....	27
4.2.4	Ansiktsuttryck.....	28
4.3	Implementation av interaktionsmetoder.....	29
4.3.1	Knappsats.....	29
4.3.2	Sensorhandske med gestigenkänning.....	29
4.3.3	Gestigenkänning med neuronnät.....	30
5	Studie 1: Användarstudie. Upplevelsen av visuell återkoppling i virtuella miljöer....	33

5.1	<i>Hypoteser / frågeställningar</i>	33
5.2	<i>Studiens genomförande</i>	33
5.2.1	Datainsamling.....	34
5.2.2	Betingelserna.....	34
5.2.3	Försökspersonerna.....	35
5.2.4	Utrustning och försöksledare.....	35
5.3	<i>Förmedlingen av återkoppling</i>	35
5.3.1	Förhållningssätt.....	36
5.3.2	Gester.....	36
5.4	<i>Resultat från studien</i>	36
5.4.1	Uppmärksamhet.....	36
5.4.2	Förståelse och attityd.....	38
5.4.3	Naturlighet.....	40
5.4.4	Transparens.....	40
5.4.5	Acceptansen av systemet.....	41
5.4.6	Observationer.....	41
5.4.7	Kommentarer från användarna.....	42
5.4.8	Förmedlingen av gesterna.....	42
5.5	<i>Problem i studien</i>	42
5.5.1	Tolkningar av frågor.....	42
5.5.2	Brister i testmiljön.....	43
5.5.3	Långtidseffekter.....	43
5.6	<i>Slutsatser</i>	43
6	Studie 2: Utvärdering av interaktionsmetoder	47
6.1	<i>Studiens genomförande</i>	47
6.2	<i>Resultat från studien</i>	48
6.2.1	Gränssnitt med knappats.....	48
6.2.2	Gränssnitt med sensorhandske.....	48
6.3	<i>Förmedla omedvetna gester medvetet</i>	49
6.4	<i>Slutsatser</i>	49
7	Summering och fortsatt arbete	51
7.1	<i>Artificiella uttryck</i>	51
7.2	<i>Visa fokus genom blickriktningsbestämning</i>	52
7.3	<i>Användarnas förväntningar</i>	53
7.4	<i>Slutord</i>	53
8	Referenser	55
	Appendix	57
	A. <i>Frågeformulär</i>	59
	B. <i>En översikt av gester och ansiktsuttryck</i>	63

1 Inledning

I och med utvecklingen av Internet och avancerad tredimensionell grafik, har ett antal applikationer utvecklats, där användare kan mötas och kommunicera med varandra i en tredimensionell miljö. I dessa miljöer kan användarna presenteras på olika sätt och kommunicera med varandra med tal eller text. Denna typ av tredimensionella fleranvändarmiljöer kan användas som samarbetsmiljöer, där möten kan äga rum och gemensamt arbete utföras. Tekniken kan vara ett intressant alternativ till såväl telefonkonferenser som videokonferenser. Fördelarna gentemot dessa system är att man kan använda systemets rumsliga egenskaper till att öka känslan av närvaro, känslan av att befinna sig på samma plats som de andra deltagarna. De tredimensionella miljöerna lämpar sig väl för möten med en stor mängd deltagare då det är möjligt att välja vem man vill tala med eller inte tala samt dela in sig i mindre grupper. De virtuella mötesmiljöerna är inte heller bredbandskrävande, d.v.s. man behöver inte skicka en stor mängd information mellan användarna.

Det finns dock några allvarliga nackdelar med dessa system jämfört med videokonferenser. En videokonferens möjliggör nämligen förmedling av kroppsspråk genom gester och ansiktsuttryck, vilket de flesta applikationer med virtuella miljöer som finns idag endast gör i mycket begränsad omfattning. Kroppsspråket fyller en rad funktioner vid kommunikation mellan människor. En viktig del av kroppsspråket är återkopplingen, som har stor betydelse t.ex. för hur ett samtal utvecklas, vem som talar när och hur vi gör för att inte tala i munnen på varandra (Duncan 1977). Denna återkoppling förmedlar också information om åhörarens uppmärksamhet och attityder till talaren (Poyatos 1983). I virtuella fleranvändarmiljöer uppstår problem på grund av de bristande uttrycksmöjligheterna. Bowers (1997) har i en studie visat på problem med bl.a. turtagning. Han efterlyser i studien större möjligheter till ickeverbala uttryck i virtuella miljöer.

Telia Research AB har utvecklat ett prototypsystem för gruppbaserad distansutbildning, *Learning Studio*, där en virtuell mötesmiljö används. Systemet är byggt på plattformen DIVE som utvecklats av SICS, (Swedish Institute of Computer Science). Varje deltagare i miljön representeras av en människoliknande avatar. Miljön stöder simultan överföring av tal och det är möjligt att aktivera samarbetsverktyg för att t.ex. dela dokument. Detta arbete har i huvudsak utgått från de specifika förutsättningar och problem denna miljö har. Miljön har också legat till grund för de prototyper som utvecklats inom ramarna för arbetet.

1.1 Arbetets syfte och upplägg

Detta arbete behandlar huvudsakligen två frågor:

- Σ Påverkar möjlighet att uttrycka kroppsspråk i virtuella miljöer kvaliteten på kommunikationen?
- Σ Kan man hitta en metod att indirekt förmedla visuella uttryck som kräver liten medveten ansträngning för användaren?

För att undersöka dessa områden har en testmiljö tagits fram. I testmiljön ingår en virtuell mötesmiljö där mötesdeltagarna har möjlighet att kommunicera med hjälp av gester och ansiktsuttryck. Detta system har använts i en användarstudie för att utröna hur användare reagerar på att bli bemötta av återkoppling i virtuella miljöer. Två metoder har också implementerats för att låta användaren förmedla återkoppling via systemet. Dessa metoder har sedan utvärderats.

Rapporten börjar med en litteraturstudie av kroppsspråkets olika funktioner och uttryckssätt. Under rubriken "Interaktionsmetoder" följer sedan en genomgång av möjliga metoder för låta användaren förmedla kroppsspråk. Nästa kapitel i rapporten, "En testmiljö med stöd för visuell återkoppling", beskriver en implementerad testmiljö där användare har möjlighet att förmedla *återkoppling* med kroppsspråk. I kapitlet redovisas och motiveras valen av de uttryck som ingår i miljön samt valen av metoder för användaren att interagera med systemet för att förmedla gester. I detta kapitel redovisas också den tekniska delen av införandet av gester. Under "Studie 1: Upplevelsen av visuell återkoppling i virtuella miljöer" beskrivs en användarstudie som försöker svara på om visuell återkoppling i virtuella miljöer har effekt på kvaliteten på kommunikationen. I kapitlet redovisas studiens upplägg och genomförande, samt resultat och slutsatser. Nästa kapitel är en utvärdering av de interaktionsmetoder som implementerats samt diskussion och slutsatser kring dessa. I "Summering och fortsatt arbete" ges en kort summering av resultaten samt en genomgång av arbete som påbörjats men som kräver fortsatta studier.

1.2 Definitioner och förklaringar

Virtual Reality (VR) eller virtuell verklighet är ett begrepp som kan innefatta allt från tredimensionella objekt i datormiljö till virtuella världar som användarna ser genom bildskärmar monterade på huvudet. När VR nämns i denna rapport menas framförallt en miljö där användaren sitter framför en datorskärm där en tredimensionell datorgenererad miljö visas, så kallad desktop-VR. Arbetet fokuserar också på virtuella miljöer där det är möjligt att överföra tal mellan användarna. De representationer av användarna som befolkar de virtuella världarna kallas för *avatarer* och är ofta men inte alltid människoliknande figurer.

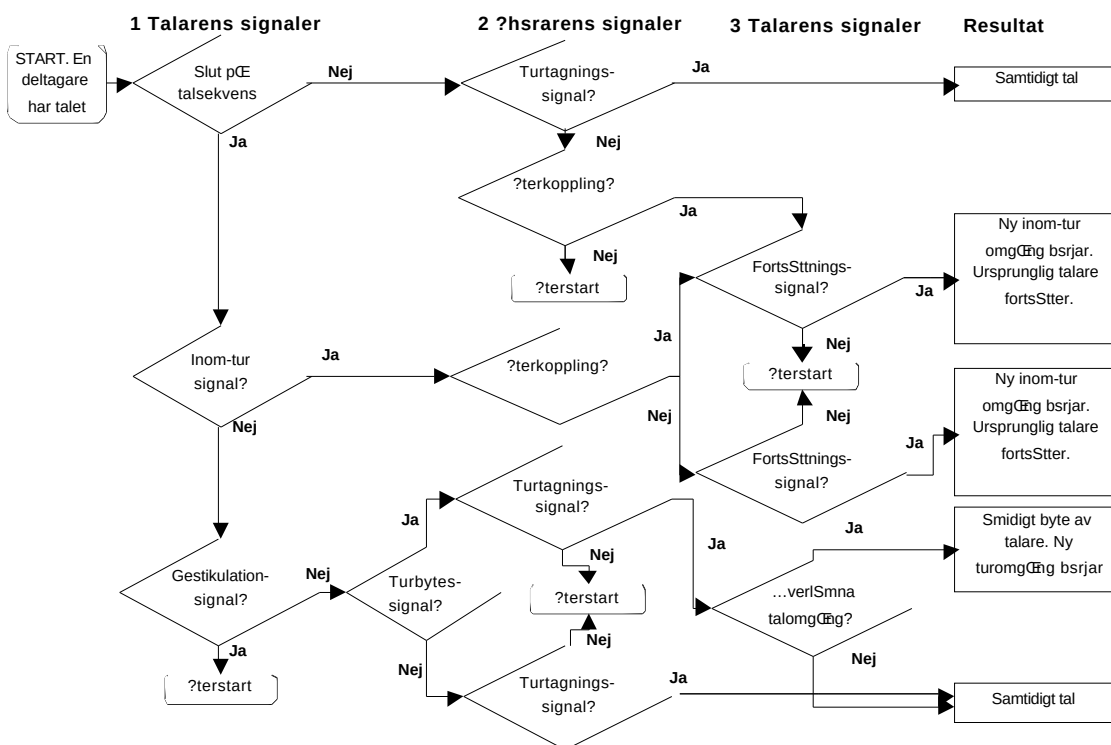
När ordet *användare* används i denna rapport menas personer som på något sätt styr en applikation. Ordet används ofta för att beskriva en person som styr en avatar och avatarens gester och ansiktsuttryck i en virtuell miljö

I detta arbete används ordet *kroppsspråk* för att beteckna alla typer av gester och ansiktsuttryck som används i kommunikation. *Visuell återkoppling* betecknar endast den del av kroppsspråket som utförs av åhöraren i en konversation. *Ickeverbalt återkoppling* omfattar visuell återkoppling men kan också inbegripa vissa röstuttryck som inte är ord. I denna rapport ligger tyngdpunkten på den visuella återkopplingen. En utförligare beskrivning av kroppsspråkets funktioner ges i följande kapitel.

2 Kroppsspråkets funktioner

2.1 Återkoppling och turtagning

Duncan och Fiske (1977) har i en omfattande studie undersökt vilka faktorer som styr så kallad *turtagning*. Med turtagning menas de regelverk och konventioner som styr vem som talar när i en konversation. I studien har de särskilt undersökt betydelsen av olika former av återkoppling för kommunikationens utveckling. Återkopplingen beskrivs som *reaktioner hos åhöraren som uppfattas av talaren utan att åhöraren övergår till att bli talare*. Denna återkoppling kan förmedlas samtidigt som talet utan att påverka kvaliteten hos den verbala kommunikationen (Poyatos 1983). Flödesschemat i figur 2.1 är en översatt och tolkad modell av Duncans logiska modell av ett turtagningsförlopp.



Figur 2-1. Figuren, som är översatt och något förenklad, visar en logisk modell av ett turtagningsförlopp, och hur det beror av återkoppling (Duncan, S. m.fl., 1977).

Återkopplingen kan bestå av korta eller längre talade uttryck, där de korta kan bestå av yttranden som "mm", "mhm" eller "ja", och de längre av flera ord som "jag förstår". Den kan också bestå av visuella signaler som nickningar, huvudskakningar och leenden. Till återkopplingen kan man också räkna

blickbeteende, hos Duncan definierat som frekvens och längd hos blickar riktade mot samtalspartnerns ansikte. Nedan följer några exempel på hur visuell återkoppling kan förmedlas och i vilket syfte (Poyatos, 1983).

Visa uppmärksamhet: Nickningar, höjda ögonbryn, luta sig framåt, leenden, ögonkontakt, handen mot hakan, luta sig tillbaka med korsade armar.

Visa intresse: Ökad frekvens av nickningar och blinkningar, förlängd ögonkontakt.

Be om klargörande: Höjda ögonbryn, rynkad panna, kupad hand vid örat om talaren talar för tyst.

Duncans och Fishers studie visar att det finns klara samband mellan förekomst och frekvens av återkoppling gentemot talomgångarnas längd, och frekvensen av byte av talare. Studien visar på ett komplext mönster av samband mellan variabler som styr interaktionen i en konversation. Studien utmynnar i utvecklingen av en metod som visar hur konversationsdeltagare inför varje samtal väljer en konvention att kommunicera efter som samtalspartnern måste acceptera eller förkasta. Denna konvention är kulturellt färgad och erbjuder ett regelverk för hur konversationen skall gå till med avseende på turtagning och användningen av återkoppling.

2.2 Visuella uttryckssätt

Turtagning och återkoppling är *funktioner* där människor använder visuella uttrycksmedel. Nedan följer en beskrivning av olika uttryckssätt, som kan användas för de funktioner som beskrivits ovan.

2.2.1 Gestikulation

Gestikulation kan definieras som gester som uttrycks i samband med tal. Om denna typ av gester kan sägas att de är intimt förknippade med det semantiska innehållet i talet. Många hävdar (bl.a. Armstrong m.fl.1995) att talet och gesterna är produkter av en gemensam representation i hjärnan. Det är därför svårt att hantera dessa gester utan att också ta hänsyn till talet. Gestikulationen kan delas upp i fyra olika typer av gester med olika egenskaper, ikoniska gester, metaforiska, deiktiska och betonande. Denna indelning används bl.a. av McNeill (1992) och Cassel (1997).

Ikoniska Gester: Gestens form avbildar någon egenskap hos målet för talet. Den kan beskriva en fysisk egenskap hos ett objekt, eller beskriva hur en handling skall utföras. Att be om notan kan illustreras med en skrivande gest mot andra handens handflata.

Metaforgester: Också här representeras en egenskap hos målet för talet, men en abstrakt sådan. Meningen "Mötet bara fortsatte och fortsatte" kan följas av en cirkulerande rörelse med handen, som symboliserar ett hjul som går runt.

Deiktiska (utpekande) gester: Deiktiska gester pekar ut punkter i rummet. Gesterna kan vara referenser till objekt eller personer i rummet, men de kan också vara referenser till personer eller objekt som inte är närvarande, men som ingår i ett spatialt sammanhang. Man kan introducera objekt i rummet mellan talare och lyssnare för att sedan referera till dem under samtalet. Dessa objekt behöver inte vara konkreta utan kan bestå av idéer eller händelser. Ett exempel på detta kan vara att man vid orden "Å ena sidan..." kan göra en gest med handen åt höger, för att sedan gestikulera åt vänster när man kommer till "Å andra sidan...".

Betonande gester: Dessa gester är formlösa gester som medvetet eller omedvetet används för att betona och ge eftertryck åt ord och meningar och har inget samband med det semantiska innehållet i talet. Gesternas funktion är i huvudsak grammatisk och kan jämföras med användningen av skiljetecken i skrift. En betonande gest kan vara en del av en annan gest som en flackande eller slående rörelse i den större gesten.

2.2.2 Fokus

Vår blick och fokus är främst ett sätt att samla information från omvärlden men vårt *blickbeteende* sänder också meningsfulla signaler till andra människor, t.ex. hur intresserade vi är av personen vi lyssnar på. Ett antal aspekter av hur vi styr våra ögon kan beaktas: (Argyle 1988)

- Σ Vår blickriktning berättar för andra var vi har vår uppmärksamhet.
- Σ Längden och frekvensen av ögonkontakt mellan personer kan avslöja vilken relation som råder mellan dem, samt attityder gentemot varandra.
- Σ Att vara iakttagen av någon kan kännas obehagligt och ge upphov till självmedvetenhet, en känsla av att vara 'observerad', medan ömsesidiga blickar kan ge en känsla av intimitet, attraktion och öppenhet.
- Σ Ögonen är uttrycksfulla på så sätt att uppspärade ögon kan uttrycka rädsla eller ilska. Storleken på pupillen kan påverkas bl.a. av rädsla och sexuell upphetsning.
- Σ Frekvensen av blinkningar varierar bl.a. med upphetsning och koncentration.

2.2.3 Ansiktsuttryck

Ansiktet är den viktigaste kommunikationskanalen för visuella uttryck. Det är särskilt viktigt för att uttrycka känslor och attityder gentemot andra människor. Ansiktsuttryck ändras snabbt och spelar en viktig roll i social interaktion. De iaktas noga av blickar riktade mot ansiktet. Med ansiktet kan vi uttrycka en mängd olika miner, men försök har visat att vi bara kan särskilja åtta olika känslouttryck ur dessa miner. Sex av dessa kan tydligt identifieras: glädje, förvåning, rädsla, sorg, ilska och avsmak/förakt. Det finns också ett antal ansiktsuttryck som inte räknas som känslouttryck som intresse, förvåning, tveksamhet, skam och smärta. Det är möjligt att dela upp ansiktet i ett antal uttryckselement och med deras hjälp beskriva ansiktsuttryck. Dessa element kan vara positioner för ögonbryn, ögon, ögonlock, mun, m.m. (Argyle 1988).

2.2.4 Övriga uttryck

Närhet: Det fysiska avståndet mellan personer i en social interaktion kan ge tydliga signaler. Avståndet mellan personer kan ange vilket förhållande de har till varandra. Avstånden kan uppfattas som publika, sociala, personliga eller intima (Poyatos 1983). Dessa avstånd varierar mellan olika kulturer samt mellan olika situationer. Genom att minska eller öka avstånden till varandra (t.ex. luta sig fram eller tillbaka) förmedlar vi bl.a. intresse, uppmärksamhet eller avståndstagande.

Symbolgester: Symbolgester är medvetna gester med ett semantiskt innehåll som i stor utsträckning inte beror av kontexten. Dessa gester är ofta knutna till kulturen, både gesternas utseende och antalet gester varierar mellan olika

länder. Italien och Frankrike har t.ex. betydligt fler än USA (Kendon, 1981). Oavsett nationella skillnader upptar symbolgester en liten del av det totala kroppsspråket, aldrig mer än 10%. Symbolgester kan vara: V-tecken, OK-tecken, tumme upp eller tecken för pengar (gnugga fingerspetsarna).

Funktionella gester: Vid funktionella gester används kroppen som ett verktyg för att påverka omvärlden så att den bättre passar våra behov. Vi kan flytta, vrida och deformera objekt i vår omgivning. Funktionella gester har inte nödvändigtvis någon kommunikativ funktion för den som utför dem, däremot sänder de ut sådana signaler till andra. Då funktionella gester ofta kräver stor uppmärksamhet för att utföras, indicerar de att personen inte är tillgänglig för kommunikation. Exempel på funktionella gester kan vara att rätta till håret eller klädseln.

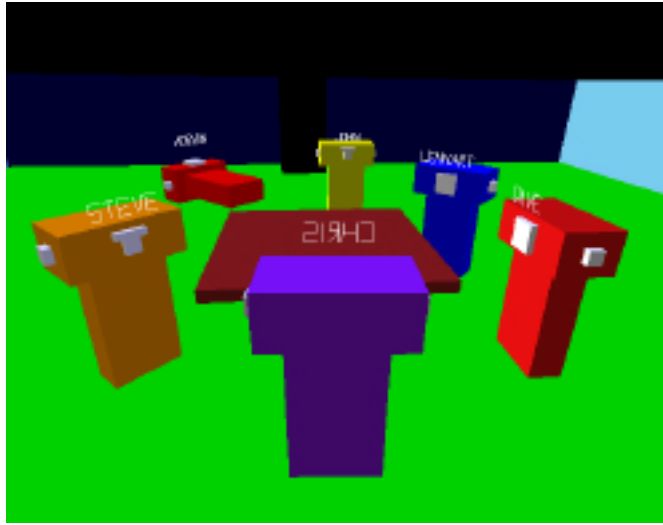
Pantomim: Pantomimgester är gester som används när man av någon anledning är förhindrad att tala, t.ex. när den ena samtalspartnern sitter i telefon med någon annan. Med denna typ av gester kan man förmedla budskap på liknande sätt som tal. Pantomimen består till stor del av ikoniska gester som efterliknar verkliga objekt.

2.2.5 Medvetna och omedvetna gester

Man bör skilja på huruvida man är medveten om att man utför en gest och om man utför en gest avsiktligt. Vissa gester som att skratta är vi klart medvetna om att vi utför men vi skrattar sällan avsiktligt. Det omedvetna kroppsspråket står för mer än 90% av totala kroppsspråket. Större delen av gestikulationen (ikoniska, metaforiska, deiktiska och betonande gester) sker omedvetet (Cassel 1997). Till de medvetna gesterna hör bl.a. de symboliska gesterna som tumme upp m.m. Man använder sig ibland också medvetet av gester för att ersätta ord.

2.3 Kroppsspråkets funktioner i virtuella miljöer

Bowers (1997) har studerat hur den sociala interaktionen fungerar under möten i virtuella miljöer. I dessa studier har man tittat på turtagning och hur den kroppsliga representationen av deltagarna påverkade den sociala interaktionen. Studien utfördes på ett system som heter MASSIVE (Greenhalgh 1995). Systemet stöder överföring av tal och användarna representeras av enkla figurer uppbyggda av klossar. Tanken med det enkla utseendet på avatarerna är att de skall spegla begränsningarna i funktionaliteten. Avataren har öron och mun för att användaren kan tala och höra, men saknar armar för att de inte skulle fylla någon funktion i det aktuella systemet. De enda kroppsliga uttrycksformer som finns är navigering och vridning på kroppen (vilket indicerar fokus) samt möjlighet att vifta på öronen(!).



Figur 2-2. I fleranvändarsystemet MASSIVE representeras användarna av enkla blockliknande figurer.

Bowers identifierar en rad problem i dessa mötesmiljöer, framförallt med turtagning där långa pauser uppstår mellan varje byte av talare. I studien undersöktes också om användarna utnyttjade de begränsade formerna av kroppsspråk som stod till buds. Det visade sig att nästan inga användare utnyttjade möjligheterna att uttrycka sig med visuella uttryck när de talade eller för att begära ordet, situationer som i vanliga fall åtföljs av ett intensivt kroppsspråk. En förklaring till detta menar Bowers kan vara att användaren måste ägna en stor del av sin uppmärksamhet åt att föra fram sitt budskap genom den dåliga ljudkanalen. Kroppsspråket användes dock desto mer av åhörarna som genom att vrida avatarerna mot varandra ”utbytte rörelser”, något som Bowers menade motsvarade att utbyta blickar i verkliga världen. Även talaren reagerade på denna återkoppling på ett sätt som är naturligt i en vanlig konversation. En av slutsatserna av studien var att användarna inom vissa gränser kunde använda avatarerna för att interagera enligt deras invanda interaktionsmönster. Bowers efterlyser dock större uttrycksmöjligheter för avatarerna. Bland annat menar han att man skall kunna titta sig omkring utan att behöva vända någon ryggen (vilket man gör i MASSIVE).

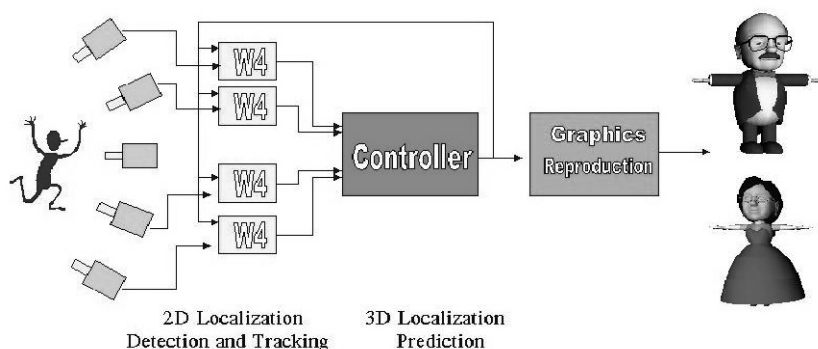
3 Interaktionsmetoder

Ett av huvudsyftena med detta arbete är att undersöka hur användaren skall interagera med ett system för att förmedla visuella uttryck. Skall användaren medvetet förmedla uttrycken, eller skall man försöka fånga användarens omedvetna kroppsspråk? Hur skall användaren förmedla uttrycken på ett enkelt sätt, utan dyr och klumpig teknik och utan alltför stor belastning för användaren? Nedan diskuteras några möjliga lösningar på problemet.

3.1 Direkt överföring av uttryck

Det är möjligt att med hjälp av videokamera eller sensorer av olika slag läsa av användarens beteende framför skärmen för att sedan överföra detta direkt till den virtuella miljön. Man kan t.ex. låta en avatar röra sig på samma sätt som användaren, eller visa videobilder på användaren direkt i den virtuella miljön. Denna metod lämpar sig för att överföra de uttryck som vi tidigare kallat för *gestikulation*. Gemensamt för dessa gester är att de i stor utsträckning är omedvetna och är intimt knutna till det semantiska innehållet i talet. De är också individuella och samma gester kan betyda olika saker i olika sammanhang. De kan bestå av fingerrörelser såväl som av arm och torsorörelser. Detta gör att gester av denna typ är svåra att förmedla på annat sätt än att direkt överföra detaljerad information om användarens kroppshållning.

Ett exempel på en metod där användarens kroppsspråk översätts direkt till en avatar är ett system som utformats i samarbete mellan MIT och University of Maryland. I detta system fångas användarens kroppsspråk av fem videokameror placerade runt användaren. Ur videobilderna från dessa kameror genereras en tredimensionell modell av användaren. Denna modell används sedan för att styra kroppen hos en avatar, se fig. 3-1, (Horprasert 1998).



Figur 3-1. Användarnas kroppsspråk fångas med fem videokameror.

Joshua Bers (1995) har utvecklat ett system där användare kan *regissera* rörelsemönster hos animerade varelser. Användaren kan t.ex. säga "flyg så här" till ett animerat bi och visa med armarna hur det skall flyga. Rörelsemönstret

översätts sedan till biet. Användarens rörelser fångas av en dräkt med magnetsensorer. Rörelserna översätts sedan till en datormodell av en kropp där hastighet och kinetisk energi räknas ut för varje kroppsdel och rörelserna utförs slutligen av den animerade modellen.

En fördel med denna metod är att användarens medvetande inte belastas. Metoden har dock några allvarliga nackdelar. Om man väljer att använda sig av kroppssensorer för att bestämma positionen på användarens kroppsdelar måste användaren sätta på sig någon form av dräkt, vilket förmodligen upplevs som besvärligt. Sensorer av denna typ är dessutom mycket dyra i dagsläget. Videolösningen är både billig och enkel för användaren, men kräver stor bandbredd vid överföringen. Den största nackdelen med att använda sig av användarens naturliga beteende som indata till systemet är att användarens omgivning inte är densamma som omgivningen i den virtuella miljön. Det kan t.ex. tänkas att användarens avatar står upp i den virtuella miljön medan användaren förmodligen sitter framför datorn. Användaren kanske också gestikulerar mot en annan person i rummet som inte är en deltagare i den virtuella mötesmiljön. Hur ska användarens kroppsrörelser överföras i det fallet? Med denna metod har användaren heller ingen kontroll över vad som förmedlas. Detta leder till att det förmedlade kroppsspråket blir 'ärligt', om kroppsspråket förmedlas på rätt sätt, men om uttrycken inte är ämnade för miljön kan förvirring uppstå.

3.2 Tolkad överföring

Man kan också tolka information om användarens beteende från sensorer och video, för att sedan återskapa motsvarande beteende i den virtuella miljön. Denna metod kräver liten bandbredd vid överföringen och innebär ingen belastning för användaren. Metoden lämpar sig för att överföra omedvetet beteende hos användaren där uttrycken är möjliga att generalisera till en begränsad uppsättning uttryck. Exempel på sådant beteende är blickbeteendet, d.v.s. var användaren har sitt fokus, och ansiktsuttryck.

Ett exempel på tolkad överföring är ett system som har utvecklats av Vertegaal med flera (1997). Där undersöker man var på skärmen användaren tittar vid användningen av en virtuell mötesmiljö. Punkter används för att visa var användarna har sin uppmärksamhet i den virtuella miljön.

Det är också möjligt att identifiera ansiktsuttryck ur videosignaler (Mase 1991). Motsvarande ansiktsuttryck kan sedan visas av en avatar i en virtuell miljö. Också denna metod begränsas av att användarens miljö inte är densamma som den virtuella miljön. Metoder av denna typ är ofta beräkningskrävande och kan också vara känsliga för feltolkningar.

3.3 Medveten förmedling av uttryck

Man kan också medvetet förmedla visuella uttryck genom att t.ex. trycka på knappar som utlöser olika gester. Detta ger användaren total kontroll över vad som förmedlas, på bekostnad av ärligheten i uttrycken. Medveten kontroll av det förmedlade kroppsspråket leder också till en kognitiv belastning för användaren. Vid utformningen av interaktionsmetoder kan man dock sträva efter att minimera belastningen. Ett sätt att minimera belastningen kan vara att försöka *automatisera* interaktionen. Automatiserade processer kräver liten eller ingen uppmärksamhet. Automatisering kan uppnås med hjälp av träning men lämpar sig bäst för situationer där beteendet kan ses som direkta responser på givna stimuli (Anderson 1995).

Många applikationer med virtuella miljöer har någon form av stöd för medveten förmedling av uttryck. I den virtuella fleranvändarmiljön Active Worlds (Active Worlds 1998) finns en uppsättning gester tillgängliga som användaren utlöser genom att trycka på knappar. Med dessa gester kan man bl.a. vinka till personer, se glad ut och dansa.

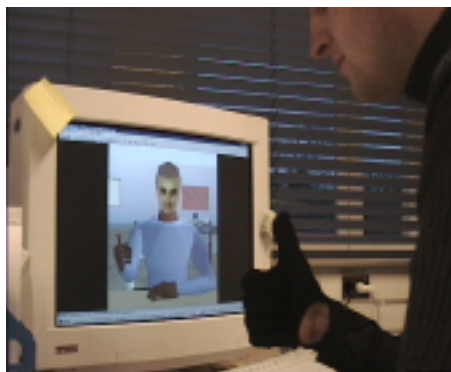
3.4 Automatiskt beteende

Kroppsspråket vid konversation följer mönster och regler som i viss mån är möjliga att generalisera. Vilhjalmsen (1997) på MIT har tagit steget fullt ut och byggt en prototyp för ett system där avatarerna i en virtuell värld sköter kroppsspråket på egen hand. Systemet är en utveckling av de s.k. MUD-världarna (Multi User Domains) där användarna representeras av en avatar i en grafisk miljö och där kommunikationen sker genom text. I Vilhjalmsens prototyp sköter avatarerna på egen hand större delen av kroppsspråket, inklusive gester, utväxling av blickar och munrörelser. Det enda användaren behöver göra för att reglera kroppsspråket är att styra några parametrar som avslöjar dennes intentioner. En sådan parameter kan vara *tillgänglighet*, om användaren välkomnar att andra söker kontakt eller inte. Om användaren vill avbryta en pågående konversation kan parametern *Avbryt konversation* användas, och leder då till ett passande kroppsspråk hos avataren. Kroppsspråket som uttrycks i kommunikationen utlöses av olika händelser. I och med att kommunikationen sker genom text kan munrörelser och ansiktsuttryck knytas till särskilda ord och tecken som uttrycks (som "?"). Andra utlösande faktorer har att göra med interaktionen med andra avatarer t.ex. vid öppning av konversation (hälsningar, den andres tillgänglighet), under konversationen (begära och ge återkoppling, betoning, ge ordet) och vid avsked.

Inte heller denna metod belastar användarens uppmärksamhet nämnvärt. Inga kameror eller sensorer behövs heller, vilket gör denna metod till en billig lösning. I denna metod har användaren dock mycket liten kontroll över vilka uttryck som förmedlas, vilket i detta fall inte är samma sak som att uttrycken blir ärligare. Uttrycken som förmedlas är hårt knutna till vissa specifika ord eller händelser, något som borde leda till feltolkningar hos systemet. Ett annat problem med metoden är att systemet måste känna till en del av innehållet i talet, något som är svårt att genomföra när kommunikationen mellan användarna inte sker genom skriven text, utan genom överföring av tal.

3.5 Förmedla gester med gester

Ett sätt att förmedla visuella uttryck utan att påtagligt höja den kognitiva belastningen kan vara att använda naturliga gester för att interagera med den virtuella miljön. Nedan beskrivs ett förslag till en ny metod där användaren med hjälp av handgester kan ge återkoppling till andra personer i en virtuell miljö. Handgesterna överförs via en sensorhandske och tolkas av systemet som kan känna igen en uppsättning gester. I den virtuella miljön kan man sedan välja att låta användarens gest motsvaras av en likadan gest, eller alternativt av ett uttryck med samma innebörd. Ett OK-tecken kan t.ex. motsvaras av ett leende. Figur 3-1 visar hur det kan se ut. Här översätts en *tumme upp* till en liknande gest hos avataren. Denna metod kan ses som en kombination av tolkad överföring och medveten överföring.



Figur 3-2. Gesterna användaren utför tolkas och omvandlas till gester med liknande semantiskt innehåll.

Kan man motivera valet av denna metod för att interagera med systemet? Tanken med de naturliga gesterna är att de skall vara lätta att förknippa med uttrycket i miljön och därför lätta att komma ihåg. Hanteringen av handsken borde också kräva liten uppmärksamhet bl.a. för att man inte behöver titta på den för att kunna använda den.

Sturmans (1991) doktorsavhandling "Whole hand input" är en studie av datorgränssnitt där man använder sig av händerna som interaktionsverktyg. Sturman går igenom för och nackdelar med sådana system samt utarbetar en designmetod för handbaserade gränssnitt. Den första delen av denna metod går ut på att genom att svara på en serie frågor avgöra om ett system passar att styras med ett handbaserat gränssnitt. Nedan följer ett försök att svara på frågorna från Sturmans avhandling för att utröna om ett handbaserat gränssnitt lämpar sig för att uttrycka gester i virtuella miljöer. Det bör sägas att metoden är utvecklad främst för system där objekt eller system skall *påverkas* på ett eller annat sätt, t.ex. inom industrin. Frågorna är kursiverade och markerade med ett "F". Svaren är inte kursiverade och är markerade med ett "S".

Naturlighet

F: Är följande egenskaper viktiga för att styra systemet?

a) Tidigare inlärda motoriska färdigheter (som minskar den kognitiva belastningen)

S: Förmedlingen av gester bör ske så intuitivt som möjligt, helst med redan inlärda gester.

b) Vedertagna gester inom området (minskar inlärningstiden)

S: Lämpligen använder man sig av gester som är naturliga och väl invanda, som tumme upp. En gest av användaren behöver dock inte leda till samma gest hos avataren. Om användningen kräver en större repertoar gester än vad som står till buds uppstår inlärningsproblem.

c) Frånvaro av mellanliggande teknik (direktmanipulation ökar känslan av närvaro)

S: En av huvudorsakerna till att förmedla gester i virtuella miljöer är att öka känslan av närvaro.

Anpassningsmöjlighet

F: Används flera sätt att styra systemet vid användning, (kan man utnyttja handgränssnittets möjlighet till flexibilitet)?

S: Om handen endast används till att förmedla gester utnyttjas nog inte möjligheterna fullt ut. Om man däremot använder handen för att navigera och manipulera objekt i miljön kan nog detta tänkas stämma. Man kan t.ex. ersätta musen.

Koordination

F: Kräver uppgifterna koordination av flera indatavariabler? (Kan man utnyttja handens förutsättningar)?

S: Förmedlingen av gester kräver nog ingen större precision och noggrannhet.

Styrning i realtid, *Människan har ett avancerat system för återkoppling mellan öga-hand som ger möjlighet till snabba reaktioner och kontinuerlig styrning*

F: Hur viktigt är kontinuerlig övervakning och justering?

S: Troligtvis inte så viktigt om man inte använder handen för navigering t.ex. för att följa blicken i en diskussion.

F: Hur viktigt är det med snabba reaktioner hos användaren?

S: Mycket viktigt, kommunikation bör ske utan fördröjningar. Återkoppling bör komma omedelbart.

Vilka slutsatser kan man dra av svaren på dessa frågor? Man kan se att det finns risk för att handens möjligheter inte utnyttjas fullt ut. Man kan också ana problem med inlärning. Samtidigt ser man fördelar som snabb återkoppling och ökad närvarokänsla. Några nackdelar som inte kommit fram är svårigheten att hantera ett tangentbord eller andra ting i omgivningen med en handske på handen.

3.5.1 Gestigenkänning

För att kunna använda metoden med gester som indata, krävs det att man har ett system som kan identifiera gester ur utdata från en sensorhandske. En hel del forskning har gjorts inom gestigenkänning, bl.a. för att kunna översätta teckenspråk för döva. Den vanligaste tekniken som använts är sensorhandskar som ger stabila och noggranna indata till systemet. Video har också använts framgångsrikt (Starner 1995). Neurala nätverk eller mönstermatchning kan användas för att klassificera gesten. Nedan följer ett par implementerade system för gestigenkänning

Kadous (1996) har utvecklat ett system där man med en billig sensorhandske designad för TV-spel (Mattels Powerglove) kan tolka teckenspråk. Handsken ger information om handens form och position. Ett antal attribut används för att definiera ett tecken. Maxvärdena från positionssensorerna används för att generera en låda, vars storlek varierar mellan olika tecken. Andra attribut som används är fingerböjningar och handvridningar.

Glove-Talk är ett system utvecklat av Fels m.fl. (1998) Detta system används för att med hjälp av gester (ej teckenspråk) *generera* tal. En sensorhandske ger

indata från handens form och position, där formen styr genereringen av konsonanter och positionen styr vokalerna. Gesterna tolkas med neurala nät.

3.6 Summering av interaktionsmetoder

Detta kapitel har behandlat några allmänna metoder för att förmedla visuella uttryck i virtuella miljöer. En avgörande faktor som skiljer de olika metoderna åt är huruvida användaren aktivt förmedlar uttryck eller om systemet fångar användarens rörelser automatiskt. Kapitlet har också gått in lite djupare på en metod där användaren förmedlar gester med gester. I tabell 3-1 sammanställs några av de olika metodernas för- och nackdelar.

Tabell 3-1. De olika metodernas för- och nackdelar

	Direkt överföring	Tolkad överföring	Medveten förmedling	Automatiskt beteende	Gester med gester
Kräver bandbredd	Ja	Nej	Nej	Nej	Nej
Kräver datorkraft	Ja	Ja	Nej	Nej	Nej
”Ärliga” uttryck	Ja	Oftast	Nej	Nej	Ibland
Risk för feltolkningar	Ibland	Ja	Nej	Ja	Nej
Uppmärksamhets- krävande	Nej	Nej	Ja	Nej	Något
Användaren har kontroll	Nej	Nej	Ja	Nej	Ja
Begränsat antal uttryck	Nej	Ja	Ja	Ja	Ja

4 En testmiljö med stöd för visuell återkoppling

Ett av syftena med detta arbete är att undersöka om möjligheten att förmedla kroppsspråk i virtuella miljöer höjer kvaliteten på kommunikationen. För att kunna undersöka detta har en testmiljö implementerats där det är möjligt att uttrycka kroppsspråk. Testmiljön har sedan använts i två studier. En användarstudie för att undersöka om kroppsspråk i virtuella miljöer höjer kvaliteten på kommunikationen och en studie för att utvärdera de metoder med vilka användaren förmedlar kroppsspråk i testmiljön.

Hur skall en testmiljö som stöder förmedling av kroppsspråk utformas? I ett idealiskt system kan användarna kommunicera på samma sätt som människor kommunicerar ansikte mot ansikte. I den testmiljö som utvecklats måste dock några begränsningar göras. Dessa begränsningar görs utifrån svaren på två grundläggande frågor.

Σ Vilken del av kroppsspråket är väsentligt att införliva i testmiljön?

Σ Hur skall användaren förmedla uttrycken i testmiljön?

4.1 Val av uttryck och interaktionsmetod

I avsnitt 2-3 redovisades i en studie av Bowers m.fl. (1997) att problem uppstår vid turtagning i virtuella miljöer med begränsade visuella uttrycksmedel, samt att deltagarna i miljöerna i större utsträckning utnyttjade kroppsspråk när de var åhörare än när de talade. Dessa slutsatser pekar på att stöd för *återkoppling* är viktigt att införa i testmiljön. Återkopplingen har inte bara stor betydelse för turtagningsprocessen, i virtuella miljöer fyller återkopplingen också funktionen att visa personen som talar att avatarerna representerar människor, att de inte är tomma skal (vilket kan vara fallet om användaren tillfälligt har lämnat tangentbordet).

En annan faktor som gör återkopplingen lämplig att implementera i testmiljön är att återkopplingens uttrycksformer går att generalisera till en begränsad uppsättning uttryck, något som är svårt att göra t.ex. med de uttryck som ryms under benämningen gestikulation. Återkoppling förmedlas till stor del genom ansiktsuttryck, nickningar och blickbeteende, men kan också ges genom handgester och förändringar i kroppshållning (Argyle 1988). Ett urval av dessa uttryck har implementerats i testmiljön. En noggrannare beskrivning av uttrycken och hur de implementerades i miljön följer under *Implementation av uttryck*.

I kapitel tre beskrevs några metoder för att förmedla kroppsspråk. De flesta av dessa metoder är möjliga att använda för att förmedla återkoppling. Den tekniskt enklaste av dessa metoder är att låta användaren medvetet förmedla återkopplingen till exempel genom att trycka på knappar. Men är det möjligt att förmedla återkoppling medvetet? I och med att man inte är upptagen av att prata när man förmedlar återkoppling, har man större möjlighet att ägna en del

av sin uppmärksamhet till att förmedla uttryck, vilket kom till uttryck i Bowers (1997) studie där åhörarna ”utbytte gester”. I testmiljön har två metoder att förmedla återkoppling implementerats: En metod där användaren förmedlar återkoppling genom att trycka på knappar och en metod där användaren förmedlar återkoppling med handgester via en sensorhandske. Tankarna bakom metoden med sensorhandsken har beskrivits i avsnitt 3-5, ”Förmedla gester med gester”. Den tekniska implementationen av dessa metoder beskrivs i avsnitt 4-3, och metoderna utvärderas i kapitel 6, ”Utvärdering av interaktionsmetoder”.

4.2 Implementation av uttryck

De uttryck som implementerats i prototypen är huvudsakligen gester och ansiktsuttryck som kan användas för att ge återkoppling, men även symboliska gester samt en betonande gest har implementerats. De gester som är ämnade för återkoppling delades in i tre grupper:

- Σ För att ge möjlighet att förmedla uppmärksamhet till talaren implementerades följande gester och ansiktsuttryck: Nickning, huvudskakning, leende.
- Σ För att ge möjlighet att förmedla förståelse: Frågande min.
- Σ För att ge möjlighet att förmedla attityd: Tumme upp, avvärande gest, tveksam min samt tvekande gest.

I övrigt implementerades två symbolgester: en hälsningsgest och en stoppgest samt en betonande gest (ett slag med handen för att betona vissa ord).

Några uttryck som sträcker sig över en längre tidsperiod implementerades också i systemet. Dessa uttryck bygger på kroppshållning och ansiktsuttryck. En normal kroppshållning implementerades, samt två kroppshållningar som representerar intresse respektive ointresse. Alla gester som implementerats redovisas med bilder i appendix B.

4.2.1 Teknisk plattform

Telia Research AB har utvecklat ett prototypsystem för gruppbasad distansutbildning, *Learning Studio*, som är en applikation byggd på plattformen DIVE. DIVE är utvecklat av SICS (Swedish Institute of Computer Science) och är en plattform för tredimensionella fleranvändarmiljöer. Den testmiljö som utvecklats inom ramarna för detta arbete är en modifikation av Learning Studio-miljön. Animeringen av gesterna och ansiktsuttrycken har programmerats i programspråket TCL (Tool Command Language), vilket är ett textbaserat programspråk som ofta används för att förmedla kommandon till interaktiva program. DIVE har definierat egna TCL-kommandon för bl.a. animering av objekt.

4.2.2 Utformning av avatar

Avatarer fyller flera funktioner i virtuella fleranvändarmiljöer. En viktig funktion är att visa för andra användare av systemet var i rummet denne befinner sig. Med en avatar kan man också förmedla var användaren har sitt fokus med hjälp av huvud eller blickriktning. Avataren kan också användas för att visa användarens identitet, genom namn eller bild på personen. Ett synsätt som kan appliceras på avatarens design är att det är bra om ett objekts utseende speglar begränsningarna i funktionaliteten. Om man t.ex. inte kan använda

armar till någonting i den virtuella miljön bör avataren inte ha några armar. Detta synsätt har bl.a. använts vid utformningen av de avатарer som används i systemet MASSIVE (Greenhalgh 1995).

De avатарer som tidigare utvecklats i Learning Studio är förhållandevis detaljerade. De har både armar, ben och en bild på användarens ansikte som är placerad över avatarens ansikte. Avатарerna har händer men inga fingrar. Trots avатарernas relativa detaljrikedom gör de ett något klumpigt intryck. Dessa avатарer bedömdes inte vara tillräckligt uttrycksfulla för uppgiften. En viktig orsak till detta beslut var avsaknaden av fingrar hos avатарerna. Fingrar är nödvändiga för att forma vissa symbolgester som tumme upp.

Den avатар som utformats för detta examensarbete skiljer sig något från de tidigare avатарerna. Förutom att den har händer med fingrar är den också något spensligare samt något mer detaljerad. Som de tidigare avатарerna har den en bild av användarens ansikte placerad över avatarens ansikte. Avataren har utformats utifrån en sittande grundposition, detta för att de gester som sedan implementerats är tänkta att utföras från den positionen. Utformningen av sittställningen bygger på betraktelser av människor i verkligheten och en viktig tanke har varit att avataren skall ge ett avslappnat intryck.

Avataren som har skapats inom ramarna för detta arbete har utformats i programmet 3D-Studio Max från Kinetix, och kroppsdelar har lånats från exempelavатарer i insticksprogrammet Character Studio. Avataren har sedan exporterats i VRML 1.0-format för att slutligen omvandlas i DIVE till DIVEs egna format för att beskriva objekt.

4.2.3 Animationen av gesterna

Ambitionen med animationen av gesterna var att de skulle vara mjuka och så lika som möjligt motsvarande gester hos en människa. Det visade sig dock att möjligheterna till animationer i systemet DIVE är begränsade.



Figur 4-1. Skärmdump av avатар som utför en gest.

För att animera ett objekt i DIVE flyttar man helt enkelt objektet i ett antal små rörelser till en slutposition. Vid animationen av gesterna består dessa rörelser som regel av rotationer kring objektets lokala centrum, t.ex. en överarms rotation kring axeln. Efter varje liten förflyttning lämnas kommandot över till DIVE som lägger kommandona i en s.k. händelsekö. Hastigheten med vilken

DIVE hanterar kommandona i kön varierar kraftigt med renderingshastigheten (hastigheten med vilken bilderna på skärmen ritas upp). Om det är många objekt på skärmen som skall renderas tar det längre tid för kommandona att bli utförda. Detta gör att hastigheten på gesterna varierar. Hastigheten varierar också med snabbheten på datorn.

4.2.4 Ansiktsuttryck

En viktig del av den ickeverbala kommunikationen sker med ansiktet. Med ansiktet förmedlar vi känslor och attityder gentemot personer i vår omgivning, samt attityder gentemot vad som sägs. Ansiktsuttrycken fyller därför en viktig roll vid återkoppling.

I det implementerade systemet ligger ett fotografi av användarens ansikte placerad som en textur över avatarens ansikte. När ett annat ansiktsuttryck skall visas byts bilden på användaren helt enkelt ut mot en annan bild. Bilderna med de olika ansiktsuttrycken (totalt 11 stycken) är tagna med en enkel digital videokamera. Bilderna har sedan behandlats i ett bildbehandlingsprogram så att de blivit exakt lika stora och så att ögonen befinner sig på samma plats i de olika bilderna. En del av bilderna har sedan kombinerats till animerade bildserier.



Figur 4-2. Bilden på ansiktet är projicerad på avatarens ansikte, som har en mänsklig form.

Bilden som ligger över avatarens ansikte när ingen gest förmedlas är ett neutralt ansikte utan något utmärkande uttryck. Med jämna mellanrum blinkar avataren, detta för att avataren skall upplevas som mer levande. Det neutrala blinkande ansiktet är skapat som en animerad bildserie bestående av endast två bilder. Det neutrala ansiktet kan bytas ut mot ett mer intresserat ansiktsuttryck. Detta ansiktsuttryck används i samband med den mer framåtlutade kroppshållning vilken avataren kan inta under längre perioder.

De övriga ansiktsuttryck som används uttrycks oftast i samband med en gest. Bilden som ligger som textur över ansiktet byts ut mot en bild med det uttryck som söks. Efter en bestämd tidsperiod byts texturen tillbaks till den ursprungliga bilden.

Ett försök till animering av munrörelser implementerades också i systemet. I DIVE registreras när ljud sänds genom systemet. Denna information användes

för att vid dessa tillfällen byta ansiktsbilden mot ett ansikte med öppen mun, för att strax bytas tillbaka mot ursprungsbilden. Denna metod fallerade dock då ljuden som sändes genom systemet inte alltid registrerades.

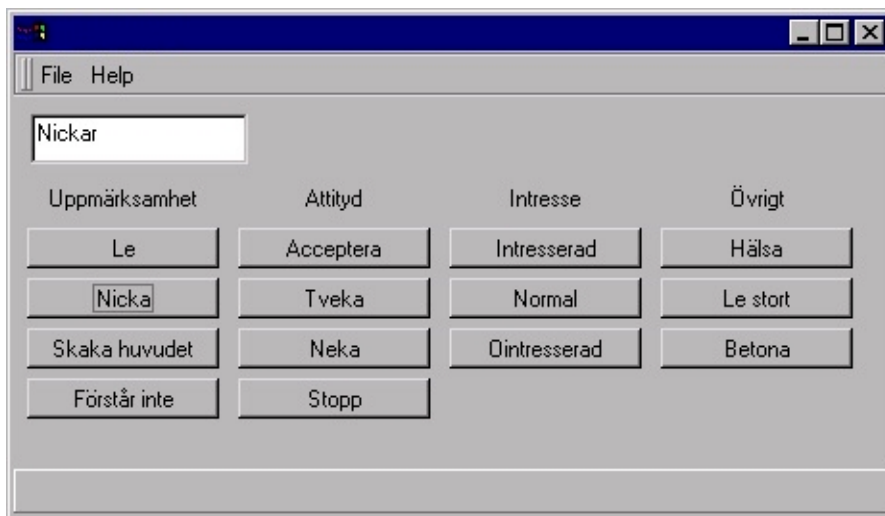
Lösningen med att använda fotografier för att förmedla ansiktsuttryck har både fördelar och nackdelar. Det är bra att användarna kan identifieras med hjälp av utseendet, och ansiktsuttrycken blir förhållandevis naturliga. Problem uppstår dock om denna lösning skall användas i ett kommersiellt system. Användarna måste då ha tillgång till en digital kamera, samt kunna utföra den bildbehandling som beskrevs tidigare i avsnittet. Det är troligt att bildbehandlingen till stor del kan automatiseras och byggas in i applikationen, men kravet på tillgång till kamera består. De ansiktsuttryck som implementerats i systemet redovisas med bilder i appendix B.

4.3 Implementation av interaktionsmetoder

I testmiljön har två olika metoder för att förmedla visuell återkoppling implementerats. En knappsats och en sensorhandske med gestigenkänning.

4.3.1 Knappsats

Det tekniskt enklaste sättet att förmedla gester genom ett system är att låta användaren trycka på knappar i ett datorgränssnitt. I detta arbete implementerades en knappsats med vilken användaren kan förmedla gester. Applikationen skrevs i Java i utvecklingsmiljön Borland Jbuilder och fungerar enligt följande:



Figur 4-3. För att förmedla gester i gränssnittet med knappsatsen trycker man på knappen med motsvarande gest. En bekräftelse på att gesten uttrycks visas i teckenfönstret uppe till vänster.

När användaren trycker på en knapp i knappsatsen skickas ett meddelande via en socket till en applikation i DIVE. I applikationen finns ett antal funktioner för att animera gester. Meddelandet som skickas från knappsatsen används för att initiera en animering av den gest som användaren valt.

4.3.2 Sensorhandske med gestigenkänning

Tidigare i detta arbete lades ett förslag fram till en ny metod för att förmedla gester. I denna metod nyttjar användaren en uppsättning naturliga gester för att

kommunicera med systemet. Gesterna tolkas av systemet och avataren i den virtuella miljön utför en gest med en liknande innebörd (inte nödvändigtvis samma gest som användarens). Denna metod har införlivats i testmiljön. Nedan följer en beskrivning av den tekniska implementationen av metoden.

Sensorhandsken som används i den implementerade applikationen heter 5th Glove från 5DT (Fifth Dimension Technologies). Handsken mäter böjningen för varje finger med hjälp av optiska fibrer. En lutningssensor känner också av lutningen på handen kring två axlar i horisontalplanet. Upplösningen för varje sensor är åtta bitar och antalet utdatavariabler är sju (fem fingersensorer och två lutningssensorer).



Figur 4-4. Sensorhandsken mäter böjningen av varje finger samt lutning och vridning av handen.

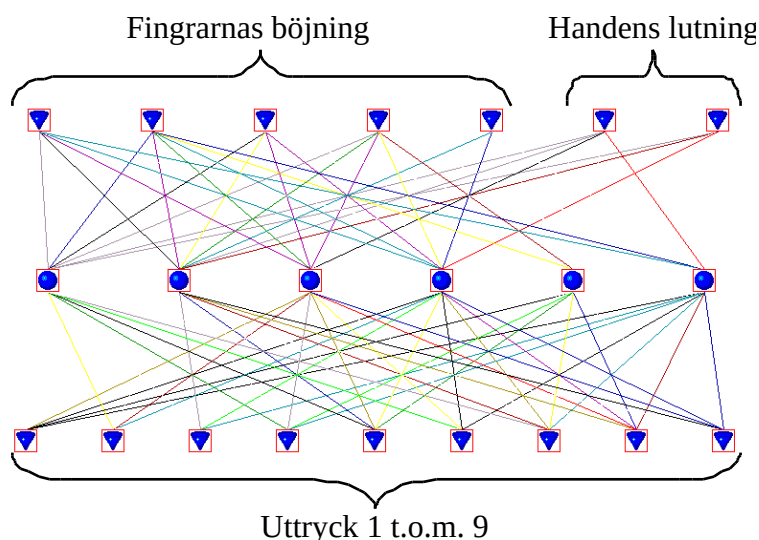
För att behandla värdena från handsken användes en betaversion av en utvecklingsplattform i Java som heter NetGlove SDK, från General Reality. Denna plattform innehåller ett kommunikationsgränssnitt för att skicka utdata från handsken till andra applikationer. Plattformen innehåller också en gestigenkänningsmodul med vilken utdata från handsken kan tolkas i form av gester. Gestigenkänningen i plattformen bygger på så kallad mönstermatchning, där en gest detekteras om utdata från varje finger ligger inom bestämda intervall.

Det visade sig dock att gestigenkänningen i plattformen inte fungerade tillfredsställande. Gestigenkänningsmodulen hade stora problem att skilja på t.ex. gesterna *tumme upp* och ett *OK-tecken*. En bidragande orsak till problemen kan vara att programmet inte utnyttjar informationen från lutningssensorerna på handsken, vilket bl.a. gör att det inte skiljer på gesterna *tumme upp* och *tumme ner*. På grund av dessa problem beslutades att en annan metod skulle prövas.

4.3.3 Gestigenkänning med neuronnät

I den andra metoden som implementerades användes ett neuronnät för att känna igen gester från utdata för handsken. I denna metod utnyttjas även informationen från lutningssensorerna på handsken. Det neurala nätverk som användes i implementationen skapades med hjälp av programmet Neural Bench från Vniigaz. Nätverket är av typen back-propagation, har tre lager neuroner och består av sju neuroner i indatalagret, sex i mellanlagret och nio i

utdatalagret. Detta nät klarar utan problem att känna igen nio olika gester. Hastigheten på gestigenkänningen är hög, programmet klarar att känna igen ca femtio gester i sekunden. Mellan varje förmedlad gest måste användaren gå tillbaka till ett viloläge, detta för att inte samma gest skall förmedlas flera gånger vid en utdragen gest av användaren. För mer information om neurala nätverk hänvisas till boken Neural Networks av Haykin (1994).



Figur 4-5. Figuren beskriver arkitekturen hos neuronnätet. Varje neuron är kopplad till alla neuroner i underliggande lager. Neuronerna i utdatalagret representerar var sitt uttryck.

Nedan följer en redovisning av de gester som användes för att kommunicera via handsken, och hur de presenterades i den virtuella världen.

Uttryck	Gest hos avatar	Inputgest för användaren
Leende	Leende	OK-tecken
Jakande , visa uppmärksamhet	Nickning	Pekande gest
Nekande	Skaka på huvudet	Pekande gest med två fingrar
Medhåll	Tumme upp + Nick	Tumme upp
Avslå	Avvärjande gest	Tumme ner
Hälsa, begära ordet	Höja handen	Höjd flat hand
Frågande	Frågande ansiktsuttryck.	Handen krökt som frågetecken
Stopp, vänta	Avvärjande gest	Pekande gest mot taket
Neutral	-	Avslappnad horisontell hand

På grund av ytterligare tekniska problem med programmet som följde med handsken, kunde inte handsken användas vid den användarstudie som utfördes som en del i detta arbete. Handsken har därför utvärderats i en separat studie. Resultat och utvärderingar från denna studie återges i kapitel sex.

5 Studie 1: Användarstudie. Upplevelsen av visuell återkoppling i virtuella miljöer

Fyller möjligheten att uttrycka visuell återkoppling någon funktion i virtuella mötesmiljöer? Förbättras kvaliteten på kommunikationen? För att undersöka detta gjordes en användarstudie där ett antal faktorer undersöktes. De huvudsakliga syftena med undersökningen var följande:

- Σ Mäta den visuella återkopplingens betydelse i virtuella miljöer, med avseende på närvarokänsla, uppmärksamhet och förmedling av attityder.
- Σ Mäta den visuella återkopplingens betydelse i virtuella miljöer, med avseende på acceptansen av systemet.

5.1 Hypoteser och frågeställningar

Vid en jämförelse mellan miljöer där åhöraren kan förmedla visuell återkoppling och miljöer där ingen sådan återkoppling kan förmedlas ställs följande hypoteser om det första fallet:

- 1) Talaren får en bättre uppfattning om:
 - a) åhörarens uppmärksamhet.
 - b) åhörarens attityder till vad som sägs.
 - c) huruvida åhöraren förstår vad som sägs.
- 2) Talaren upplever en ökad närvarokänsla, därför att:
 - a) systemet upplevs som mer transparent.
 - b) kommunikationen känns mer naturlig.

I denna undersökning genereras gesterna i den virtuella miljön medvetet, medan återkoppling i verkligheten för det mesta utlöses omedvetet. Det är därför också intressant att undersöka om gesterna upplevs som naturliga.

5.2 Studiens genomförande

För att kunna avgöra återkopplingens betydelse för kvaliteten på kommunikationen är det nödvändigt att skapa en kommunikationssituation där försökspersonerna utsätts för återkoppling. I denna studie bestod kommunikationssituationen av att försökspersonerna samtalade med en försöksledare. Denna försöksledare kommer i fortsättningen att kallas för *samtalspartner*. Varje försöksperson gick igenom tre delmoment där betingelserna ändrades mellan varje moment. Varje moment bestod av att samtalspartnern ställde frågor till försökspersonen där han bad denne att argumentera för eller emot en ståndpunkt i ett engagerande men inte alltför känsligt ämne, t.ex. ”Skall vi ha reklam i TV?”. Denna metod gav samtalspartnern en serie yttranden som lämpade sig att ge återkoppling på.

Samtalspartnern uppmanades försöka samtala naturligt, men att vara uppmärksam på sitt eget beteende. Efter varje fråga uppmanades försökspersonen att svara på frågor om upplevelsen av samtalspartnern, med avseende på uppmärksamhet, förståelse och huruvida denne höll med om vad försökspersonen sade. Samtidigt angav samtalspartnern i ett formulär vad han ansåg att han hade förmedlat. Varje försöksperson samtalade med försöksledaren ca tio minuter i varje betingelse. De olika betingelserna beskrivs nedan.

5.2.1 Datainsamling

Efter varje fråga som ställdes av samtalspartnern fyllde både försökspersonen och försöksledaren i varsitt formulär med tre frågor. Efter varje delmoment (som bestod av tre frågor) fyllde försökspersonen i ett något längre formulär och i slutet av det sista delmomentet svarade försökspersonen på ytterligare ett kort formulär. Försöket avslutades med en kort intervju där försöksledaren förde anteckningar. Frågeformulären bifogas i Appendix A.

5.2.2 Betingelserna

Testet utfördes under tre olika betingelser:

Virtuell miljö utan möjlighet till visuell återkoppling (inga gester eller ansiktsuttryck).

I testuppsättningen såg användaren följande av en tredimensionell datormiljö: En avatar som satt på andra sidan av ett bord, se figur 5-1. Denna avatar var en tämligen verklighetstrogen representation av en människa. I denna betingelse var avataren helt orörlig med undantag för ögonblinkningar. Försökspersonen samtalade med en försöksledare via en headset-mikrofon. Denna betingelse kommer i fortsättningen att benämnas "VR-miljö utan gester".



Figur 5-1. I de virtuella miljöerna ser försökspersonerna denna bild av samtalspartnern. I den ena miljön är avataren statisk, i den andra försedd med gester och ansiktsuttryck.

Virtuell miljö med begränsad möjlighet till visuell återkoppling.

Samma utseende på miljön som i den tidigare betingelsen. Försökspersonen blev i denna betingelsen bemött med begränsad visuell återkoppling. Återkopplingen bestod av en uppsättning gester och ansiktsuttryck hos avataren på andra sidan

bordet. Försöksledaren som agerade samtalspartner förmedlade återkopplingen genom att trycka på knappar i ett datorgränssnitt. Denna betingelse kommer i fortsättningen att benämnas "VR-miljö med gester".

Videokonferens, (större möjlighet till visuell återkoppling).

I denna betingelse såg försökspersonen försöksledaren via en videokamera. Huvudet och halva överkroppen är synliga.

5.2.3 Försökspersonerna

Tretton försökspersoner användes i testet. Alla försökspersoner arbetade på Telia i Farsta. Merparten av dem arbetade med teknik men även personer med administrativa arbetsuppgifter samt personer i ledande befattningar deltog i försöket. De flesta personerna hade god datorvana och liten kännedom om virtuella mötesmiljöer och videokonferensmiljöer. Ett par försökspersoner hade dock stor erfarenhet av att använda videokonferenser. Åldern hos försökspersonerna var något ojämnt fördelad med fem personer över 50 år och sju personer mellan 20 och 30 år. Den högsta åldern var 57 år, den lägsta 24 år. Fyra kvinnor deltog i försöket.

5.2.4 Utrustning och försöksledare

VR-miljön som användes är den testmiljö som beskrivits i kapitel fyra. Videokonferensmiljön som användes heter Easy Conference och nätverket som användes var Telias interna Ethernet.

I både videokonferensmiljön och den virtuella mötesmiljön satt försökspersonen framför en stor skärm där en bild från en dator projicerades med hjälp av en projektor placerad bakom skärmen. I båda miljöerna sändes ljudet genom videokonferenssystemet. Figur 5-2 visar en försöksperson framför skärmen som visar den virtuella miljön.



Figur 5-2. Både videokonferensen och VR-miljön visades på en stor skärm framför försökspersonen.

Två försöksledare deltog vid genomförandet av uppgiften, en som agerade som samtalspartner och talade med och förmedlade gester till försökspersonen, samt en försöksledare som skötte direktkontakten med försökspersonen och observerade denne via en videokamera från ett angränsande rum. Den senare var författaren till detta arbete.

5.3 Förmedlingen av återkoppling

I studien förmedlades återkopplingen till försökspersonerna genom att försöksledaren tryckte på knappar i ett datorgränssnitt. Nedan följer en

beskrivning av de gester som användes i undersökningen under betingelsen VR-miljö med gester.

5.3.1 Förhållningssätt

Denna typ av uttryck är sådana som är mer bestående över tid och förmedlar uppmärksamhet och attityder gentemot samtalspartnern. Två sådana *förhållningssätt* implementerades, en framåtlutad kroppshållning kombinerat med ett intresserat ansiktsuttryck, samt en mer upprätt hållning med ett neutralt ansiktsuttryck.

Tabell 5-1. *Uttryck som är bestående över en längre tid.*

<i>Uttryck</i>	<i>Gest hos avatar</i>	<i>Ansiktsuttryck</i>
Neutral	-	Neutralt
Intresse/Uppmärksamhet	Lutar sig framåt	Uppmärksam

5.3.2 Gester

En uppsättning mer tidsbundna uttryck användes i studien. Dessa uttryck är oftast direkt kopplade till något talaren sagt. Med hjälp av dessa uttryck kan man bl.a. uttrycka uppmärksamhet, attityder och att man inte förstår.

Tabell 5-2. *Tidsbundna uttryck som lämpar sig för återkoppling.*

<i>Uttryck</i>	<i>Gest hos avatar</i>	<i>Ansiktsuttryck</i>
Leende	-	Leende
Jakande	Nickning	-
Nekande	Skaka på huvudet	-
Frågande	-	Frågande ansiktsuttryck
Acceptera	Nickning + tumme upp	Leende
Tveka	'vicka på handen'	Tveksamhet
Avböja	Avvärjande gest	Blundar
Hälsning	Höja handen	Leende
Stopp, vänta	Höjd hand.	-

Bilder på gesterna och ansiktsuttrycken finns i appendix B.

5.4 Resultat från studien

Nedan presenteras en sammanställning av resultaten från de frågor som försökspersonerna svarade på under testet.

5.4.1 Uppmärksamhet

En av hypoteserna som sattes upp för testet var att det är lättare att förmedla uppmärksamhet om man har tillgång till gester och ansiktsuttryck. Eller med andra ord: Differensen mellan talarens uppfattning om förmedlad uppmärksamhet och åhörarens uppfattning om talarens uppmärksamhet bör vara mindre om man har möjlighet att förmedla gester. Figur 5-3 visar hur denna differens räknades fram från frågeformulären. Frågeformuläret var utformat så att försökspersonen fick ange sina uppfattningar längs en linje graderad från 0% till 100%. Detta val av gradering gör att resultaten från frågorna är redovisade i procent. Detta gäller även differenserna.

Fråga till försökspersonen:

Hur uppmärksam upplevde du samtalspartnern? 0% |-----X-----| 100%

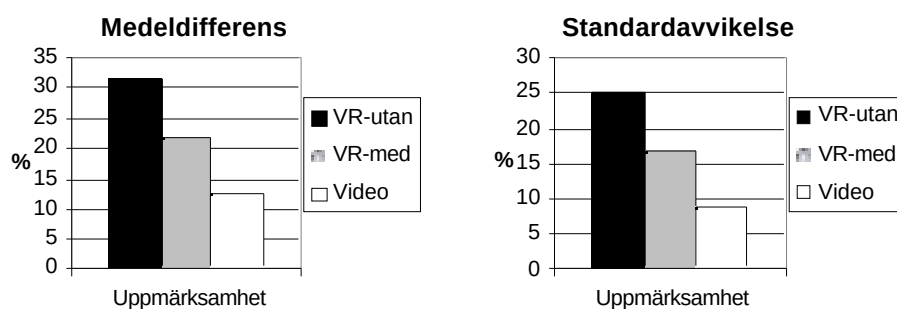
Fråga till samtalspartnern:

Hur uppmärksam var du? 0% |-----X-----| 100%

Differens

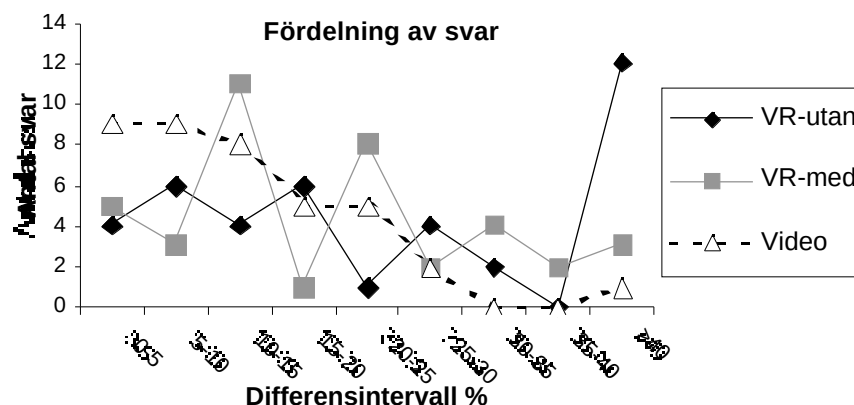
Figur 5-3. Differensen mellan försökspersonens och samtalspartnerns uppfattning om samtalspartnerns uppmärksamhet räknades ut som figuren visar.

Hypotesen om uppmärksamheten kan inte säkerställas statistiskt med resultaten av detta test då standardavvikelsen är alltför stor och antalet försökspersoner litet, men resultaten ger en antydning om att hypotesen är riktig. Figur 5-4 visar skillnaderna mellan medeldifferenserna för de olika systemen samt standardavvikelsen för differenserna.



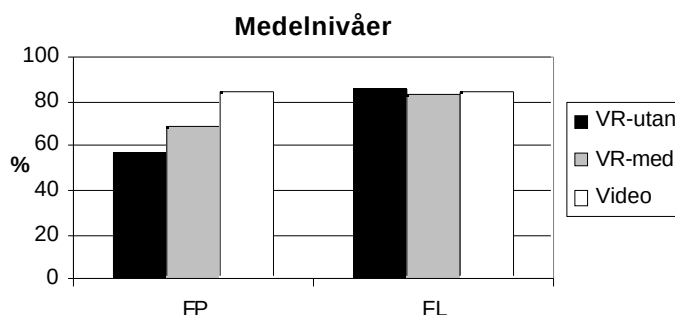
Figur 5-4: Diagrammet till vänster visar medelvärdet av differenserna mellan försökspersonens och samtalspartnerns uppfattning om samtalspartnerns uppmärksamhet. Diagrammet till höger visar standardavvikelsen för dessa differenser.

Ur dessa diagram kan vi utläsa en ganska tydlig skillnad mellan de tre betingelserna, vilken stöder hypotesen. Som regel hade försökspersonerna lättare att uppfatta vad samtalspartnern ansåg att han förmedlade när det fanns möjlighet att uttrycka gester. Det är intressant att också titta på skillnaden i standardavvikelse. Vi ser en tydlig trend med minskande standardavvikelse med ökande grad av tillgång till gester. Frågan är hur detta skall tolkas. Om man granskar resultaten för varje person ser vi att en ganska stor andel av försökspersonerna kraftigt har missbedömt uppmärksamheten i VR-miljön utan gester, vilket borde förklara både den stora medeldifferensen och standardavvikelsen (se fig. 5-5).



Figur 5-5: Diagrammet visar fördelningen av differenser mellan försökspersonens och samtalspartners uppfattning om samtalspartners uppmärksamhet. Till vänster i diagrammet redovisas de bästa överensstämmelserna och till höger de sämsta. Varje person har gett tre svar per betingelse.

Det är också intressant att titta på medelnivåerna på de värden försökspersonerna satt på samtalspartners uppmärksamhet. I Figur 5-6 ser vi att samtalspartnern (betecknad "FL" i diagrammet) har hållit en jämn medelnivå medan nivåerna för försökspersonerna varierar kraftigt mellan de olika betingelserna. Detta resultat tyder på att en miljö som stöder överföring av gester och ansiktsuttryck leder till att personerna i miljön upplevs som mer uppmärksamma. I och med att nivåerna för samtalspartnern som regel är högre än för försökspersonerna så vet vi inte om förändringarna i resultat mellan betingelserna är ett mått på bättre överföring av samtalspartners sanna uppmärksamhet eller om det bara är så att en gestikulerande avatar upplevs som mer uppmärksam.

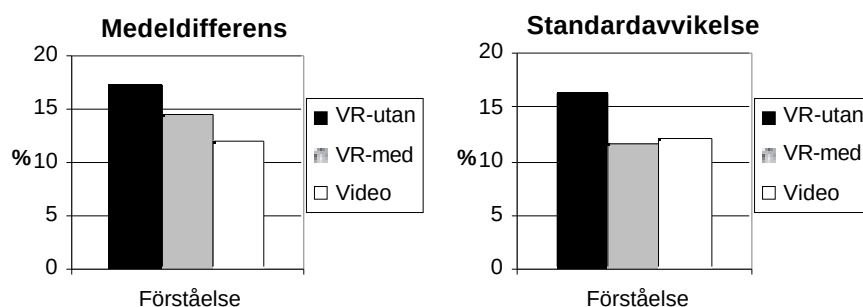


Figur 5-6. Diagrammet visar medelvärdet av försökspersonernas (FP) uppskattningar av samtalspartners uppmärksamhet i jämförelse med medelvärdet av samtalspartners (FL) uppskattningar av sin egen uppmärksamhet.

5.4.2 Förståelse och attityd

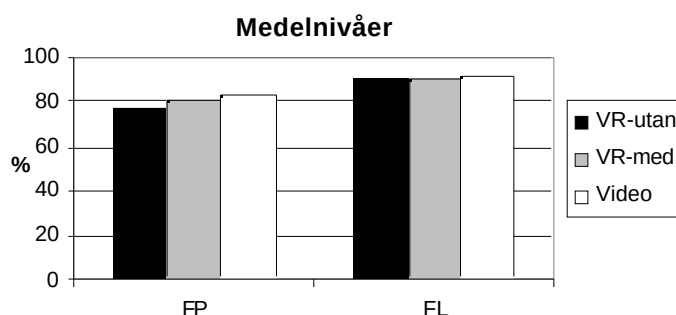
Undersökningen testade också hypotesen att det är lättare för åhöraren att förmedla huruvida denne *förstår vad som sägs* om miljön har stöd för visuell återkoppling. Detta undersöktes på samma sätt som uppmärksamheten, med hjälp av differensen mellan användarens och försökspersonens uppfattningar. I

figur 5-7 kan vi se en tendens till stöd för hypotesen då differensen minskar med ökande möjligheter till återkoppling. Skillnaderna är dock betydligt mindre än för uppmärksamheten. Standardavvikelsen är också här alltför stor för att man skall kunna dra några statistiska slutsatser, men även i detta fall kan vi se skillnader i standardavvikelse mellan de olika betingelserna.



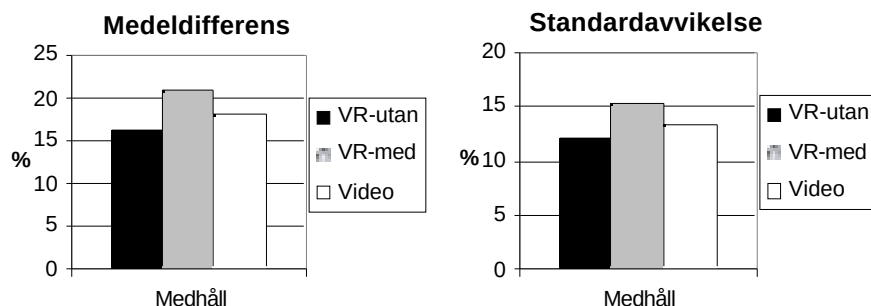
Figur 5-7. Diagrammet till vänster visar medelvärdet av differenserna mellan försökspersonens och samtalspartners uppfattning om hur väl samtalspartnern förstod vad försökspersonen menade. Diagrammet till höger visar standardavvikelsen hos dessa differenser.

Skillnaden i medelvärdet av försökspersonernas uppskattningar, som redovisas figur 5-8, är för liten för att man skall kunna dra några slutsatser om huruvida möjligheten att uttrycka återkoppling gör att åhöraren upplevs som att denne förstår mer.



Figur 5-8. Diagrammet visar medelvärdet av försökspersonernas (FP) uppskattningar av hur väl samtalspartnern förstod vad som sades i jämförelse med medelvärdet av samtalspartners (FL) uppskattningar av sin egen förståelse.

En tredje hypotes som skulle testas var att det är lättare att visa en talare *huruvida man håller med eller inte*, om man har tillgång till gester och ansiktsuttryck. Undersökningen gav inget stöd åt den hypotesen. VR-miljön utan gester fick här det lägsta medelvärdet på differenserna följt av videokonferensen och sist VR-miljön med gester. Storleken på standardavvikelsen följer samma trend.



Figur 5-9. Diagrammet till vänster visar medelvärden av differenserna mellan försökspersonens och samtalspartnerns uppfattning om i hur stor utsträckning samtalspartnern höll med försökspersonen. Diagrammet till höger visar standardavvikelsen hos dessa differenser.

En möjlig förklaring till dessa värden kan vara att samtalspartnern i undersökningen ofta avslöjade genom talet om han höll med eller inte genom att ge motargument. Detta kan ha lett till att försökspersonerna i mindre utsträckning sökte information från den visuella kanalen.

5.4.3 Naturlighet

För att få reda på huruvida försökspersonerna upplevde systemen som naturliga ställdes tre frågor: En fråga om det kändes onaturligt eller naturligt att kommunicera genom de olika systemen och en fråga om naturligheten hos samtalspartnerns visuella beteende. I slutet av testet ombads försökspersonerna också att rangordna de tre försöksuppsättningarna tillsammans med några andra vanliga kommunikationssystem efter stigande naturlighet längs en skala.

Försökspersonerna gav något högre betyg åt systemet med gester än systemet utan i de två första frågorna. Undantaget var frågan där försökspersonen ombads rangordna kommunikationsmedel. Här hamnade de båda VR-miljöerna lägst på skalan av naturlighet ungefär i nivå med e-post. VR-miljön med gester hamnade något lägre än VR-miljön utan gester. I alla tre frågorna fick båda VR-miljöerna tydligt negativa svar i fråga om naturlighet.

Vid intervjuerna framkom att försökspersonerna lade olika betydelse i ordet naturlighet. Några ansåg e-post och telefonsamtal vara naturliga på grund av att man använder dem ofta, andra satte låga värden på e-post och telefonsamtal. Att personerna tolkat betydelsen hos naturlighet på så olika sätt gör att det är svårt att dra några slutsatser om betingelsernas påverkan på uppfattningen av naturlighet.

5.4.4 Transparens

Tre frågor ställdes för att undersöka hur medvetna försökspersonerna var om gränssnitten i de olika fallen:

- Σ Upplevdes samtalspartnerns visuella beteende som störande?
- Σ Upplevdes gränssnittet som störande?
- Σ Hur medvetna var försökspersonerna om gränssnittet?

Uppfattningarna om huruvida den gestikulerande avataren upplevdes som störande varierade mycket. I intervjuerna uppgav någon att de "fladdrande rörelserna" var distraherande. Också resultaten från VR-miljön utan gester

varierade stort. Medelvärdena från de två VR-miljöerna är ungefär lika stora och något lägre än för videokonferensen.

En stor del av försökspersonerna ansåg att båda VR-miljöerna var störande i någon mån. Försökspersonerna ansåg som regel att de var mindre medvetna om gränssnittet i videokonferensen än i de två VR-miljöerna. Åsikterna om VR-miljöerna var ganska spridda. Förvånande är att man tänkte ungefär lika mycket på gränssnittet i VR-miljön utan gester som i VR-miljön med gester. Den förra miljön är ju nästan helt statisk och i intervjun jämförde många denna betingelse med ett telefonsamtal.

5.4.5 Acceptansen av systemet

Försökspersonens inställning till systemen under de olika betingelserna undersöktes genom frågor om hur försökspersonen kände inför att kommunicera genom de olika systemen, med avseende på naturlighet, hur roligt och trevligt det var, samt en fråga om försökspersonens inställning till att använda systemet.

Svaren på frågorna tyder på att försökspersonerna var något mer positiva till att kommunicera genom uppsättningen med gester än utan gester men att skillnaden mellan dessa system är liten i förhållande till värdena angivna för videokonferensen som upplevdes mycket mer positiv. Resultaten från de två VR-miljöerna ligger mestadels på den negativa halvan av svarsområdet, medan resultaten från videokonferensen ligger en bra bit in på den positiva halvan.

I intervjuerna framkom ett antal åsikter som berör acceptansen av systemen. En del personer tyckte att avataren såg onaturlig ut. I VR-miljön med stöd för gester upplevde många avatarens rörelser som ryckiga och störande.

5.4.6 Observationer

Under testets gång videofilmades försökspersonerna av försöksledaren. Ett antal faktorer söktes:

- Σ Hur stor del av tiden försökspersonen tittade på skärmen.
- Σ Hur livligt försökspersonen gestikulerade under de olika betingelserna.
- Σ Om försökspersonen 'sökte återkoppling', d.v.s. vid vissa tillfällen tittade mot skärmen för att få bekräftelse på något denne just sagt.

Användarens fokusering på skärmen varierade mellan de olika systemen. Ett klart samband kunde ses mellan graden av rörelser på skärmen och försökspersonernas fokusering på skärmen. I VR-miljön utan stöd för gester kopplade många försökspersoner bort fokus från skärmen efter ett tag, och tittade i stället ner i bordet eller ut genom fönstret. Dessa försökspersoner upplevdes som helt koncentrerade på ljudkanalen. I VR-miljön med stöd för gester observerades vid ett flertal tillfällen att plötsliga rörelser på skärmen fångade försökspersonens uppmärksamhet.

Gestikulation med händer och armar var som regel liten under alla tre betingelserna. Ingen större skillnad kunde upptäckas mellan betingelserna, gestikulation hos försökspersoner observerades även i VR-miljön utan stöd för gester. Stora skillnader kunde dock observeras vad gäller ansiktsuttrycken under de olika betingelserna. I de båda VR-miljöerna visade majoriteten av försökspersonerna ganska neutrala ansikten medan ansiktena upplevdes mycket mer livfulla och uttrycksfulla i videokonferensen, med märkbart högre frekvens

av leenden. Det upplevdes som att det fanns en visuell kontakt mellan försökspersonen och samtalspartnern som saknades i VR-miljöerna.

I undersökningen observerades också i hur stor grad försökspersonerna sökte återkoppling efter ett yttrande genom att söka ögonkontakt med samtalspartnern. Denna företeelse observerades under alla betingelserna men i olika stor utsträckning. Ett förlopp som observerades vid flera tillfällen var att försökspersonen sökte återkoppling men inte fick det, försökspersonen tenderade då att sluta söka återkoppling. Detta förlopp var tydligast i VR-miljön utan gester där en stor del av försökspersonerna var mer fokuserade på skärmen i början av testet. När de senare sökt återkoppling ett par gånger utan att få det slutade de intressera sig för skärmen. Försökspersonen fick inte alltid sökt återkoppling i de andra betingelserna heller, detta på grund av att samtalspartnern inte var fokuserad på försökspersonen hela tiden. Då det var svårt att identifiera denna företeelse tydligt, är det svårt att göra en jämförelse mellan de olika betingelserna. Samtalspartnern hade möjlighet att se försökspersonen under alla betingelser via videokonferenssystemet, men då hanteringen av gesterna i VR-miljön sköttes från en annan skärm var samtalspartnern inte alltid fokuserad på försökspersonen.

5.4.7 Kommentarer från användarna

Under den avslutande intervjun ställdes ett antal öppna frågor om försökspersonernas uppfattningar om de olika systemen, om de upplevde några fördröjningar av gesterna samt om de hade några förslag till förbättringar för VR-miljön med gester. De flesta hade åsikter om utformningen av gesterna. En stor del av försökspersonerna upplevde gesterna som ryckiga, samt upplevde en fördröjning av gesterna. Några försökspersoner upplevde ansiktsuttrycken och gesterna som onaturliga. Vid en jämförelse mellan de olika systemen föredrog nästan alla videokonferensen som ett möjligt arbetsverktyg. En person tyckte bäst om VR-miljön med gester när frågorna var av mer känslig typ, personen kände sig då mer anonym. En annan person ansåg att VR-miljön med gester lämpar sig bäst för situationer som inte är känsliga, typ planeringsmöten.

5.4.8 Förmedlingen av gesterna

Försöksledaren, som agerade samtalspartner under testet, förmedlade gester genom att trycka på knappar i ett datorgränssnitt. Denna metod att förmedla gester visade sig inte vara idealisk. Försöksledaren upplevde att denne ej kunde använda hela sin uppmärksamhet till samtalet, då denne med jämna mellanrum var tvungen att kontrollera att rätt knapp trycktes vid rätt tidpunkt.

5.5 Problem i studien

Ett antal faktorer kan ha påverkat användarna och deras svar på frågorna. Nedan följer en redovisning av några problem som identifierats i samband med studien.

5.5.1 Tolkningar av frågor

Frågorna i testet handlar mestadels om subjektiva värderingar av upplevelser hos försökspersonerna. Det är sannolikt att försökspersonerna har tolkat frågorna olika och har olika uppfattningar om vilka egenskaper som definierar ett värdeomdöme. Vilka egenskaper krävs t.ex. för att något skall få kallas naturligt? Några personer kan ha definierat naturlighet hos kommunikationsmedier utifrån hur ofta de används och hur accepterade de är i

samhället, medan andra också kan ha vägt in konkreta likheter med ”naturlig” kommunikation öga mot öga.

5.5.2 Brister i testmiljön

VR-miljön med stöd för gester som implementerats för testet har ett antal brister som kan ha påverkat resultaten från studien. Dessa brister har inte passerat obemärkta förbi försökspersonernas kritiska ögon.

- Σ Avatarens rörelser var något ryckiga, vilket enligt försökspersonernas omdömen har upplevts som störande. Det verkar också ha inverkat på upplevelsen av naturlighet.
- Σ En viss fördröjning fanns i systemet, vilken upplevdes som störande av nästan alla försökspersoner. Detta kan också ha påverkat resultaten vid mätningarna av uppmärksamhet, förståelse och attityd.
- Σ Den begränsade uppsättningen gester samt att varje gest utförs likadant varje gång kan ha uppfattats som störande.
- Σ Försökspersonen saknar möjlighet att själv uttrycka medveten återkoppling. Om försökspersonerna hade haft denna möjlighet är det möjligt att de i större utsträckning accepterat denna kommunikationsform och dess begränsningar.

5.5.3 Långtidseffekter

Det faktum att de flesta försökspersonerna var obekanta med virtuella mötesmiljöer kan ha påverkat deras inställning till dem. I och med att VR-miljöer är ett nytt medium saknar användare av systemet förväntningar på systemet. Att avatarerna har mänskliga ansikten och förhållandevis detaljerade kroppar kan ha lett till att användarna vid den första kontakten med VR-miljön fått höga förväntningar på visuell kommunikation. Om inte systemet uppfyller dessa förväntningar påverkas acceptansen av systemet. När användaren lärt sig systemets begränsningar minskar dessa effekter. I videokonferensmiljön kan man anta att förväntningarna stämde bättre överens med begränsningarna i systemet.

5.6 Slutsatser

Visuell återkoppling i virtuella miljöer har effekt på hur uppmärksam representationen av användaren upplevs. Man kan dock ej utifrån testet avgöra om möjligheten att uttrycka visuell återkoppling leder till bättre överensstämmelse mellan upplevd och förmedlad uppmärksamhet eller om det endast är så att en gestikulerande avatar upplevs mer uppmärksam. Man kan också se tendenser till att gester i virtuella miljöer ger bättre information om huruvida åhöraren förstår vad talaren säger eller ej.

Studien kan inte påvisa att stöd för visuell återkoppling i virtuella miljöer ökar närvarokänslan och känslan av naturlighet. Dessa faktorer verkar bero mycket på kvaliteten hos animeringarna av gester och ansiktsuttryck. Detta är dock inte verifierat, men det får anses som mycket intressant för en fortsatt utveckling av området att kartlägga hur kvaliteten på utformningen av uttrycken påverkar faktorer som acceptans, transparens och upplevelsen av naturlighet.

6 Studie 2: Utvärdering av interaktionsmetoder

I detta arbete har två interaktionsmetoder för att låta användare förmedla återkoppling implementerats. För att undersöka för- och nackdelar med metoderna har en utvärderande studie gjorts.

Det viktigaste kravet ett system för att förmedla gester måste uppfylla är att det i så liten utsträckning som möjligt skall belasta användarens medvetande och uppmärksamhet. Sådan belastning kallas *kognitiv belastning*. I det ideala fallet kan användaren ägna hela sin uppmärksamhet åt konversationen, utan att behöva tänka på förmedlingen av gester. Det är också intressant att undersöka om den kognitiva belastningen minskar om man tränar på att använda systemet. Om man kan göra användningen mer intuitiv genom träning, hur lång träning behövs, och hur låga nivåer på den kognitiva belastningen kan man nå?

Andra viktiga krav som bör ställas på systemet är:

Snabbhet: Återkopplingen bör förmedlas så snabbt som möjligt, för att undvika missförstånd.

Säkerhet: Det är olyckligt om gesterna som förmedlas inte stämmer överens med talet.

6.1 Studiens genomförande

Två interaktionsmetoder implementerades för att förmedla återkoppling i virtuella miljöer. En knappsats, där användaren ger återkoppling genom att trycka på knappar, och en sensorhandske med gestigenkänning. Dessa interaktionsmetoder har testats var för sig vid skilda tillfällen och under något skilda omständigheter. Knappsatsen har testats som interaktionsmetod såtillvida att den använts av en och samma person under användarstudien som beskrevs i kapitel 5. Metoden med sensorhandsken har använts i ett mindre test, i en situation som i så stor mån som möjligt liknade användarstudien. Personen som hanterade handsken i testet var samma person som skötte knappsatsen i användarstudien. Denna studie består i huvudsak av resultaten från en intervju med denna person. Personen kommer från och med nu att omnämnas som *intervjupersonen*.

Intervjun gjordes i samband med testet av sensorhandsken, och intervjupersonen ombads jämföra upplevelsen av att förmedla återkoppling med knappsatsen gentemot förmedling med sensorhandsken. Frågorna som ställdes syftade i huvudsak till att undersöka hur väl systemen levde upp till de krav som diskuterats. Nedan utvärderas de två interaktionsmetoderna var och en för sig utifrån resultaten från intervjun.

6.2 Resultat från studien

6.2.1 Gränssnitt med knappats

Knappatsen som implementerats har ett mycket intuitivt användargränssnitt. Användaren ser en uppsättning knappar. På varje knapp står skrivet vilken gest som utförs när man trycker på knappen.

Svaren på intervjufrågorna visar att intervjupersonen upplevde att interaktionen med knappatsen krävde en del uppmärksamhet. Han menade bland annat att han aldrig kunde vara helt uppmärksam på samtalet med försökspersonerna, för att en del av tiden gick åt till att leta rätt på den knapp han skulle trycka på. Han menade också att det till en viss del gick att automatisera användningen av knappatsen men att hanteringen krävde uppmärksamhet också efter en längre tids användning. Man var fortfarande tvungen att "leta efter rätt knapp".

Med *längre tid* menas den totala användningstiden av knappatsen i användarstudien plus träningstillfällen. Denna tid kan approximeras till ca. tre timmar uppdelat på minst fjorton tillfällen under drygt en veckas tid.

Intervjupersonen upplevde också att letandet efter knappar påverkade snabbheten i systemet. Vad gäller säkerheten så menade han att han mycket sällan tryckte på fel knapp.

6.2.2 Gränssnitt med sensorhandske

Intervjupersonen tyckte otvetydigt bättre om gränssnittet med sensorhandsken än knappatsen. Han poängterade tydligt att denna metod var mindre uppmärksamhetskrävande än metoden med knappatsen. Den uppmärksamhet som krävs av användaren är att komma ihåg vilken handgest som motsvarar det uttryck som skall förmedlas. Här får användaren hjälp av att de flesta handgesterna har en semantisk innebörd som motsvarar innebörden i det uttryck som förmedlas. Detta bekräftades av intervjupersonen som påpekade att det var svårare att komma ihåg de handgester som saknade semantisk koppling till avatarens uttryck, och allra lättast att komma ihåg handgesterna som översattes till samma uttryck hos avataren (t.ex. tumme upp).

En viss inlärningstid krävdes för att kunna hantera detta gränssnitt. I studien räckte det dock med några minuter för intervjupersonen för att lära sig hantera gesterna. Intervjupersonen sade sig nästan helt kunna koppla bort användningen av handsken i slutet av testet (som var ca 20 min långt). I slutet av testet tog intervjupersonen för givet att dennes förmedlade gester togs emot och förmedlades korrekt, och slutade därmed att undersöka vad som förmedlats.

I testet av gränssnittet med handsken hände det vid ett par tillfällen att fel uttryck förmedlades i förhållande till användarens avsikt. Det var endast två handgester som förväxlades, dessa var "tumme upp" och "tumme ner". Anledningen till förväxlingen beror på att användaren när denne formar gesten "tumme ner" först formar handgesten "tumme upp" för att sedan vrida tummen nedåt. Detta är ett allvarligt problem eftersom de två gesterna har motsatt semantisk innebörd. Det är inte svårt att lära sig hur gesterna skall formas så att detta problem inte uppstår, och det borde också gå att bygga bort denna bugg i programmet, men man får fortfarande betrakta detta som ett allvarligt problem.

I övrigt upplevdes gränssnittet med sensorhandsken som snabbare än knappatsen. Intervjupersonen upplevde inte handskens utformning som

särskilt klumpig, det gick bra att använda sig av musen utan att några gester utlöstes av misstag. Handsken är dock för klumpig för att det skall gå att problemfritt använda tangentbordet, vilket får ses som ett minus.

Inga utförliga studier har gjorts på hur väl gestigenkänningen fungerar på olika personer, men på det fåtal personer som använt systemet har inga större skillnader upptäckts.

6.3 Förmedla omedvetna gester medvetet

De interaktionsmetoder som beskrivits ovan bygger båda två på att användaren förmedlar uttryck *avsiktligt*. De uttryck som förmedlas med dessa metoder används dock för det mesta omedvetet och *oavsiktligt* i kommunikation öga mot öga. Hur kan man då motivera att dessa uttryck överhuvudtaget *går* att förmedla avsiktligt?

I interaktionen med andra människor använder vi oss av ett *beteende* som inbegriper de konventioner som bestämmer hur vi kommunicerar ickeverbalt med varandra. Bara för att detta beteende består av signaler som vi sänder ut omedvetet betyder inte det att vi inte kan lära in ett nytt beteende där vi använder oss av avsiktliga signaler. Att detta är möjligt kan bland annat ses hos skådespelare som har lärt sig att kontrollera en stor del av sitt omedvetna beteende. Vid interaktionen med de verktyg som beskrivits får vi helt enkelt skådespela litet. Detta gör att de signaler vi sänder ut inte blir "äkta" på samma sätt som om de är omedvetna, och samspelet med andra personer kanske inte flyter på lika friktionsfritt som i verkligheten, på samma sätt som interaktionen mellan dåliga skådespelare inte alltid går friktionsfritt.

En av slutsatserna i Bowers (1997) studie var att människor använder sig av de medel som står till buds i de virtuella miljöerna för att försöka kommunicera på samma sätt som de gör i verkligheten. Detta kan ses som ett stöd för att användarna kan lära sig att använda gesterna så att de fyller samma funktion som i verklig kommunikation.

6.4 Slutsatser

Att förmedla återkoppling genom att trycka på knappar kräver så pass mycket uppmärksamhet av användaren att det upplevs som störande för kommunikationen även efter en tids träning. Det finns dock sätt att förmedla återkoppling som kräver så pass lite uppmärksamhet att de ej upplevs som störande. En sådan metod bygger på att användaren interagerar med systemet med hjälp av naturliga gester som överförs till systemet via en sensorhandske. Denna metod upplevdes som snabb och endast i liten grad störa konversationen.

7 Summering och fortsatt arbete

Detta arbete har visat att möjligheten att ge återkoppling med gester i virtuella miljöer påverkar hur användare av sådana miljöer upplever andra användare. Störst effekt identifierades vad gäller upplevd uppmärksamhet hos andra användare. Arbetet har också redovisat en metod att förmedla gester till systemet som är snabb och kräver liten uppmärksamhet av användaren.

Ytterligare arbete har påbörjats inom ramarna för detta examensarbete. En del av detta arbete och förslag på hur man kan gå vidare redovisas nedan:

7.1 Artificiella uttryck

Den virtuella miljön ger möjlighet att använda sig av uttryck som ligger utanför den naturliga uppsättningen gester. Det är också möjligt att förstärka naturliga uttryck, och göra dem tydligare. Problemet med att införa nya artificiella uttryck är att användarna måste lära sig och enas om vad de betyder. En jämförelse kan göras med de *smileys* (":-)") som används för att uttrycka t.ex. ironi i e-post och chat-meddelanden. Dessa teckenkombinationer är inte omedelbart begripliga för någon som inte stött på dem förut. Efter ett tag lär man sig dock och accepterar konventionerna för användningen av dessa uttryck.



Figur 7-1. Artificiella objekt kan användas för att förstärka uttryck.

I testmiljön gjordes två försök med artificiella gester. Det ena är ett frågetecken som dyker upp i anslutning till ett frågande ansiktsuttryck, som en förstärkare av uttrycket. Denna gest förstärkte på ett bra sätt det förmedlade uttrycket. I det andra försöket användes en pratbubbla, i vilken olika typer av bilder eller text kan visas. Det var dock svårt att hitta lämpligt innehåll att visa i pratbubblan. Denna typ av artificiella uttryck är intressanta men ytterligare studier krävs vad gäller utformning och användningsområden.

7.2 Visa fokus genom blickriktningsbestämning

Inom ramarna för detta arbete implementerades också en metod för att förmedla var i den virtuella miljön användaren har sin uppmärksamhet. Denna metod bygger på att man känner till var på skärmen användaren tittar. Ett sådant system håller på att utvecklas på Telia Research AB, av Per Nystedt. I skrivande stund kan detta system med hjälp av en vanlig videokamera, bestämma var på skärmen användaren tittar, med ca 1,5 graders noggrannhet. Med hjälp detta system implementerades en applikation som i den virtuella miljön vrider avatarens huvud mot den avatar som är närmast användarens blickriktning. Användaren märker ingenting av det som sker, då dennes vy inte förändras. De andra avatarerna i miljön kan dock avgöra om den berörda avataren tittar på dem eller ej. I figur 7-2 visas en exempelsituation där tre avatarer sitter runt ett bord.



Figur 7-2. Testuppsättning.

Figur 7-3, visar den ena avataren genom ögonen på var och en av de övriga två avatarerna. Man kan i denna bild tydligt avgöra vem avataren tittar på. Systemet med blickriktningsdetektion belastar inte användarens överhuvudtaget och förmedlar likväl intressant information om användaren till andra användare.



Figur 7-3. Det är lätt att avgöra på vem denna avatar har sin uppmärksamhet.

Möjligheten att visa fokus är intressant främst i virtuella miljöer med flera användare. Ytterligare studier med flera användare krävs för att kunna utforma och utvärdera denna funktion på ett bra sätt.

7.3 Användarnas förväntningar

Utifrån resultat från användarstudien kan en ny hypotes formas angående påverkan av gesternas och avatarens utformning på användarnas förväntningar. Resultaten från användarstudien visar att acceptansen för VR-miljön med stöd för gester bland många personer är lägre än miljön utan stöd för gester. Den hypotes som ställs är följande: *En miljö med människolika avatarer som uttrycker sig med mänskliga gester ger upphov till högre förväntningar på systemet hos användarna än ett system med avatarer som mindre påminner om människor. Om avatarerna i högre utsträckning kan identifieras som människor förväntas de i högre utsträckning bete sig som människor.* Detta arbete ger inga belägg som styrker denna hypotes, så denna fråga är öppen för vidare undersökningar. Om hypotesen skulle stämma betyder det att system med mindre människolika representationer av användarna (seriefigurer, boxar, eller dylikt,) har större möjligheter att accepteras av användarna. Med sådana system kommer dock nya problem, som hur visuell återkoppling skall förmedlas på ett sätt som förstås av användarna.

7.4 Slutord

Ännu har inte virtuella miljöer av typen som diskuterats i detta arbete slagit igenom på bred front. Utvecklingen av kraftfulla personatorer med stöd för tredimensionell grafik, tillsammans med den snabba utvecklingen av Internet, gör dock att man kan räkna med att goda tekniska förutsättningar snart kommer att finnas för utvecklingen av applikationer med tredimensionella fleranvändarmiljöer. Om denna teknologi kommer slå igenom beror därför mer på applikationernas utformning, och om man kommer hitta lämpliga användningsområden för dessa, än på tekniska svårigheter. I alla typer av sådana applikationer kommer dock kommunikationen mellan användarna av systemet vara viktig.

Detta arbete har visat att visuella uttryck och framförallt visuell återkoppling påverkar kvaliteten på kommunikationen i virtuella miljöer. Det återstår dock mycket att göra för att hitta lämpliga former för kommunikationen, både vad gäller utformningen av uttrycken och hur användaren skall interagera med systemet för att förmedla dem.

8 Referenser

Om kommunikation med gester, kroppsspråk m.m.

- ANDERSON, J. R. 1995. *Cognitive psychology and its implications*. W.H. Freeman and Company. ISBN 0-7167-2385-9.
- ARMSTRONG, F. STOKOE, W. C. WILCOX S.E. 1995. *Gesture and the nature of language*. Cambridge University Press. ISBN 0-521-46213-4.
- ARGYLE, M. 1988. *Bodily communication*. –2nd ed. Routledge. ISBN 0-415-05114-2.
- DUNCAN, S. FISKE, D. W., 1977. *Face to face interaction*. Lawrence Erlbaum Associates, Inc. ISBN 0-470-99113-5.
- KENDON, A. 1981. *Nonverbal communication, interaction and gesture*. Mouton, the Hague. ISBN 90-279-3489-4.
- LINELL, P. 1982. *Människans språk*. Gleerups Förlag. ISBN 91-40-60035-8.
- MCNEILL, D. 1992. *Hand and Mind. What gestures reveal about thought..* The University of Chicago Press. ISBN 0-226-56132-1
- POYATOS, F. 1983. *New perspectives in nonverbal communication*. Pergamon Press. ISBN 0-08-030203-3.

Forskning om kroppsspråk i VR-miljöer

- BERS, J. 1995. *Directing Animated Creatures through Gesture and Speech*. In partial fulfillment of the requirements for the degree of master of science. Massachusetts institute of technology
- BOWERS, J. 1997. *Collaborative Virtual Environments: Grounding development and evaluation in social analysis*. CID-rapport nr 9. TRITA-NA-D9702, KTH, Stockholm.
- CASSEL, J. STEEDMAN, M. BADLER, N. PELACHAUD, C. STONE, M. DOUVILLE, B. PREVOST, S AND ACHORN, B. 1994. *Modelling the interaction between speech and gesture*. University of Pennsylvania.
- HORPRASERT, T. HARITAOGLU, I. HARFWOOD, D. DAVIS, L. WREN, C. PENTLAND, A. 1998. *Real-time 3D motion capture*. Proceedings of the 1998 Workshop on Perceptual User Interfaces.
- STURMAN J, D. 1992. *Whole hand input*. Massachusetts Institute of Technology. Del av doktorsavhandling.
- THÓRISSON, K. R. 1992 *Multi-Modal Natural Dialogue*. Massachusetts Institute of Technology. Del av doktorsavhandling.
- THÓRISSON, K. R. 1996. *Communicative Humanoids*. Massachusetts Institute of Technology. Del av doktorsavhandling.
- VERTEGAAL, R. VELICHKOVSKY, B. VEER, G. *The GAZE Groupware system*. <http://reddwarf.wmwutwente.nl/gaze/home.html>.

VILHJALMSSON, H. H. *Autonomous Communicative Behaviours in Avatars*. Massachusetts Institute of Technology. Master's Thesis in Media Arts and Sciences.

Användarstudier

DIX, A. FINLAY, J. ABOWD, G. Beale, R. (1993). *Human-computer interaction*. Prentice Hall International. ISBN 0-13-458266-7

Gestigenkänning

FELS, S. HINTON, E. (1998) *GloveTalkII. A neural network interface which maps gestures to parallel formant speech synthesizer controls*. University of Toronto Press.

HAYKIN, S. (1994). *Neural networks*. Macmillan College Publishing Company Inc. ISBN 0-02-352761-7.

KADOUS M.W. 1996. *Machine Recognition of Auslan Signs Using Powergloves: Towards Large-lexicon Recognition of Sign Languages*. Workshop on the Integration of Gestures in Language and Speech, Wilmington Delaware, October 7-8, 1996.

MASE, K. 1991. *Recognition of facial expression from optical flow*. IEICE Transactions, Vol. E 74, No. 10 October 1991.

STARNER, T. 1995. *Visual recognition of american sign language using hidden Markov models*. Massachusetts Institute of Technology. Master's Thesis in Media Arts and Sciences.

VR-system

GREENHALGH, C. BENFORD, S. 1995. *MASSIVE: a Distributed Virtual Reality System Incorporating Spatial Trading*. Proc. IEEE 15th International Conference on Distributed Computing Systems (DCS'95).

ActiveWorlds. 1998. www.activeworlds.com. Länk till Active Worlds hemsida (giltig 981201).

Appendix

A. Frågeformulär

B. En översikt av gester och ansiktsuttryck

Appendix A: Frågeformulär

Försäkran

Härmed försäkras att de individuella uppgifter du lämnar i denna undersökning inte kommer att föras vidare till tredje person. I redovisningar av resultaten av undersökningen kommer inga individuella uppgifter redovisas.

Martin Jonsson

Försöksledare

Bakgrundsfrågor

1. Kön: Kvinna ☐
 Man ☐

2. Ålder: _____ år

3. Huvudsaklig arbetsuppgift (teknik, administration, ekonomi m.m.) _____

4. Teknikintresse Mycket stort ☐
 Stort ☐
 Medelstort ☐
 Litet ☐
 Mycket litet ☐

5. Datorvana Mycket stor ☐
 Stor ☐
 Medelstor ☐
 Liten ☐
 Mycket liten ☐

6. Videokonferensvana Mycket stor ☐
 Stor ☐
 Medelstor ☐
 Liten ☐
 Mycket liten ☐

7. Vana vid 3D-fleranvändarsystem Mycket stor ☐
 Stor ☐
 Medelstor ☐
 Liten ☐
 Mycket liten ☐

8. Har du vid testtillfället märkbart nedsatt syn? Nej: ☐ Ja: ☐

9. Har du vid testtillfället märkbart nedsatt hörsel? Nej: ☐ Ja: ☐

Delmomentsfrågor 1

Nedan följer ett antal frågor som berör den fråga du just besvarat. För att svara på frågorna sätter du ett kryss på den plats på den streckade linjen som bäst överensstämmer med din inställning.

Fråga 1

1. Hur uppmärksam upplevde du din samtalspartner?

Inte alls Mycket
(0-----100%)

2. Hur väl förstod han vad du menade?

Inte alls Allt
(0-----100%)

3. I hur stor utsträckning höll han med dig?

Ingenting Allt
(0-----100%)

Fråga 2

1. Hur uppmärksam upplevde du din samtalspartner?

Inte alls Mycket
(0-----100%)

2. Hur väl förstod han vad du menade?

Inte alls Allt
(0-----100%)

3. I hur stor utsträckning höll han med dig?

Ingenting Allt
(0-----100%)

Fråga 3

1. Hur uppmärksam upplevde du din samtalspartner?

Inte alls Mycket
(0-----100%)

2. Hur väl förstod han vad du menade?

Inte alls Allt
(0-----100%)

3. I hur stor utsträckning höll han med dig?

Ingenting	Allt
(0-----	100%)

Delmomentsfrågor 2

Nedan följer ett antal frågor som berör den uppgift du just genomfört. För att svara på frågorna sätter du ett kryss på den plats på den streckade linjen som bäst överensstämmer med din inställning.

1. Hur kändes det att kommunicera genom detta system?

Onaturligt	Naturligt
(0-----	100%)

Tråkigt	Roligt
(0-----	100%)

Otrevligt	Trevligt
(0-----	100%)

2. Vilken är din inställning till att använda detta system?

Negativ	Positiv
(0-----	100%)

3. Hur upplevde du din samtalspartners visuella beteende?

Störande	Inte störande
(0-----	100%)

Onaturligt	Naturligt
(0-----	100%)

4. Hur medveten var du om gränssnittet som skilde dig och samtalspartnern åt?

Tänkte på det hela tiden	Tänkte inte på det
(0-----	100%)

5. Upplevde du gränssnittet som störande? Ja ☐ Nej ☐

Hur? _____

6. Hur fokuserad var du på händelserna på skärmen?

Tänkte inte på det	Tänkte på det hela tiden
(0-----	100%)

7. Anser du att ert samtal skulle varit lättare om ni samtalat direkt med varandra i samma rum?

Ja ☐ Nej ☐

Varför? _____

Avslutande formulär

Hur naturlig upplevde du den kommunikation nu du haft jämfört med kommunikation i andra medier?

Placera ut siffrorna som motsvarar nedanstående system på linjen. Systemen placeras ut efter *graden av naturlighet* vid kommunikation genom dem.

1. Videokonferens (med i testet)
2. VR-möte utan gester (med i testet)
3. VR-möte med gester (med i testet)
4. Telefonsamtal
5. E-post
6. Samtal öga mot öga

Grad av naturlighet

(0-----100%)

Appendix B: En översikt av gester och ansiktsuttryck som användes i studien.



Normal



Acceptera



Betona



Intresserad



Avslå



Stopp



Hälsning



Tveka

Artificiella gester



Frågande



Utrop

Ansiktsuttryck



Normal



Intresserad



Blinkar



Frågande



Tveksam



Leende



Stort leende