

JOHAN KARLSSON, JONAS HERMELIN,  
BRITTA LEVIN, SUSANNA NILSSON



Johan Karlsson, Jonas Hermelin, Britta Levin,  
Susanna Nilsson

# Att styra med gester

Kunskapsöversikt och utvärdering av gester som gränssnitt

Titel	Att styra med gester – Kunskapsöversikt och utvärdering av gester som gränssnitt
Title	Controlling with gestures – Overview and evaluation of gesture based interaction
Rapportnr/Report no	FOI-R--5381--SE
Månad/Month	December
Utgivningsår/Year	2022
Antal sidor/Pages	39
ISSN	1650-1942
Kund/Customer	Försvarsmakten
Forskningsområde	Ledningsteknologi
FoT-område	Ledning och MSI
Projektnr/Project no	E716586
Godkänd av/Approved by	Christian Jönsson
Ansvarig avdelning	Ledningssystem

Bild/Cover: Bild genererad med AI-applikationen DALL-E 2

Detta verk är skyddat enligt lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk, vilket bl.a. innebär att citering är tillåten i enlighet med vad som anges i 22 § i nämnd lag. För att använda verket på ett sätt som inte medges direkt av svensk lag krävs särskild överenskommelse.

This work is protected by the Swedish Act on Copyright in Literary and Artistic Works (1960:729). Citation is permitted in accordance with article 22 in said act. Any form of use that goes beyond what is permitted by Swedish copyright law, requires the written permission of FOI.

## Sammanfattning

Försvarsmakten använder sedan länge gester för att kommunicera order och avsikt på såväl kort som långt håll, både inom och mellan grupper, eller team. Gestbaserad kommunikation har länge varit förbehållen kommunikation mellan människor, men detta kommer sannolikt att förändras när även autonoma system som obemannade farkoster, exempelvis UGV:er (eng. Unmanned Ground Vehicle) inkluderas i teamen. Målsättningen för arbetet i rapporten är att skapa grunden för ett *proof of concept* för geststyrning av en UGV med hjälp av kommersiellt tillgänglig teknik. Rapporten redovisar dels en teoretisk studie med fokus på gester som kommunikations- och interaktionsverktyg, dels resultatet av en inledande teknisk utvärdering, i form av två delstudier, av en motion capture-handske avsedd att användas för interaktion med en UGV. Resultatet visar att det går att nyttja motion capture-handskar för interaktion med en UGV, men att det krävs fortsatt arbete för att komma fram till vilka gester som lämpar sig för olika scenarier. Det krävs även fortsatt arbete för att justera maskininlärningsmodellen som ligger till grund för registrering och tolkning av gesterna.

Nyckelord: gränssnitt, gester, gestbaserad interaktion, obemannad farkost, UGV, maskininläring, motion capture handske, human autonomy teaming

## Summary

The Swedish Armed Forces have long used gestures to communicate orders and intent at short as well as long distances, both within and between groups, or *teams*. Gesture-based communication has primarily been used for communication between people, but this will likely change when autonomous systems and vehicles, such as UGVs (Unmanned Ground Vehicles), are also included in the teams. The objective of the work described in this report was to create the basis for a *proof of concept* demonstration of gesture control of a UGV using commercially available technology. The report presents a theoretical background with focus on gestures as a tool for communication and interaction, as well as the results of two initial technical evaluations of a motion capture glove used for interaction with a UGV. The study showed that it is possible to use the motion capture gloves for interaction with a UGV, but that further work is required to determine which gestures are suitable for different scenarios and different types of interaction. Further work is also required to adjust the underlying machine learning model for registration and interpretation of the gestures.

Keywords: human-machine interface, gesture based interaction, unmanned ground vehicle, UGV, machine learning, motion capture glove, human autonomy teaming

# Innehållsförteckning

<b>1</b>	<b>Inledning .....</b>	<b>7</b>
1.1	Bakgrund .....	7
1.2	Om projektet Framtida gränssnitt .....	8
1.3	Syfte med rapporten .....	8
1.4	Läsanvisning.....	9
<b>2</b>	<b>Gestbaserad interaktion.....</b>	<b>10</b>
2.1	Gester .....	10
2.2	Gestbaserade gränssnitt .....	12
<b>3</b>	<b>Utvärdering av motion capture-handske.....</b>	<b>15</b>
3.1	Deltagare .....	15
3.2	Utrustning .....	16
3.3	Utformning av gester .....	17
3.4	Genomförande.....	18
3.5	Identifiering av gester .....	20
3.6	Resultat.....	21
3.7	Diskussion .....	24
<b>4</b>	<b>Tekniskt test av gestbaserad manuell styrning av UGV .....</b>	<b>25</b>
4.1	Deltagare .....	25
4.2	Utrustning .....	25
4.3	Utformning av gester .....	25
4.4	Genomförande.....	27
4.5	Resultat.....	29
4.6	Diskussion .....	29
<b>5</b>	<b>Diskussion .....</b>	<b>30</b>
5.1	Återkoppling.....	30
5.2	Valet av gester.....	31
<b>6</b>	<b>Framtida arbete .....</b>	<b>32</b>
	<b>Referenser .....</b>	<b>33</b>
	<b>Appendix A: Illustrationer av gester .....</b>	<b>36</b>



# 1 Inledning

Försvarsmakten använder sedan länge gester för att kommunicera order och avsikt på såväl kort som långt håll både inom och mellan grupper, eller *team*. Team är i detta sammanhang benämningen på en grupp människor som arbetar tillsammans för att nå ett gemensamt mål eller syfte. Även om gestbaserad kommunikation varit förbehållen människa till människa kommer detta sannolikt att förändras när även obemannade farkoster inkluderas i teamen. *Human-autonomy teaming* (HAT) handlar om förmågan att använda avancerade mekaniserade enheter (såsom obemannade flygfarkoster eller markgående robotar) som en del av ett team i vilket maskinerna tar på sig roller som tidigare utförts av människor. En förutsättning för fungerande människa-maskin team är dock intuitiva gränssnitt mellan människan och maskinen (MoD, 2018). Kroppsspråk och som en del av detta, gester, är en viktig del av den kommunikation som behövs för att till exempel samarbeta effektivt i mänskliga team när dessa löser gemensamma uppgifter. Det kan utifrån detta antas att kroppsspråk då även kommer vara viktig för kommunikation inom människa-maskin team.

Denna rapport handlar om hur gester kan nyttjas med hjälp av en motion capture-handske för att styra en UGV. Målsättningen är att detta ska möjliggöra framtida studier av samverkan mellan människor och teknik, i synnerhet inom team i vilka både människor och obemannade farkoster ingår.

## 1.1 Bakgrund

All interaktion mellan teknik och människor sker via någon form av gränssnitt var genom information, kommandon, intentioner och dylikt kan kommuniceras. Vanliga gränssnitt vid människa-datorinteraktion är att presentera information via bildskärmar och att inmatning av information sker via fysiska eller virtuella tangentbord och datormus. För fall i vilka interaktionen syftar till att styra rörelsen hos tekniska enheter är också joystick ett vanligt gränssnitt.

I takt med att tekniken utvecklas skapas också nya behov att förändra och uppdatera de gränssnitt som används för interaktion mellan människa och maskin. Gester är ett exempel på gränssnitt av intresse eftersom gester både är allmänt förekommande i kommunikation mellan människor och är nära knutna till vårt tänkande (Hostetter and Alibali, 2008; McNeill, 1995). Det ter sig således att gester torde vara relevant för att möjliggöra en för oss naturlig och indirekt interaktion med teknik, exempelvis för styrning av farkoster. Termen gest används mer specifikt i denna rapport i betydelse av *en uttryckande, betydelsebärande kroppslig rörelse där fingrar, händer, armar, huvud, ansikte eller kropp används i syfte att 1) förmedla meningsfull information eller 2) interagera med omgivningen* (Mitra och Acharya, 2007).



I det arbete som beskrivs i denna rapport är avsikten att nyttja gester för kommunikation och interaktion med en obemannad farkost, en UGV (eng. Unmanned Ground Vehicle). För att möjliggöra sådan interaktion krävs att UGV:n kan uppfatta och tolka de gester som används av människan. Ett möjligt sätt att göra detta är genom visuell tolkning av gester och handrörelser, där kameror används för att detektera handens position och rörelser som sedan kan omvandlas till input för interaktion. Detta innebär dock att det krävs en kamera för detektion av handrörelsen, vilket förutsätter att handen hela tiden kan detekteras av kameran för att interaktionen ska fungera. Ett alternativ till denna lösning är att använda sig av så kallade motion capture-handskar, vars sensorer känner av handens rotation och fingrarnas relativa positioner. Denna information kan då förmedlas vidare och tolkas utan behovet av en kamera eller andra sensorer än de i handsken.

För att undersöka förutsättningar och lämplighet för gestbaserade gränssnitt vid interaktion och kontroll över olika typer har två studier genomförts i vilka en kommersiellt tillgänglig motion capture-handske har testats och utvärderats som interaktionsverktyg. I kapitel 3 redovisas ett test för hur olika typer av gester tekniskt kan identifieras och i kapitel 4 redovisas ett exempel på hur manuell styrning av en UGV kan ske med gester.

## 1.2 Om projektet Framtida gränssnitt

Projektet Framtida gränssnitt är ett FoT<sup>1</sup>-projekt inom ramen för Försvarsmaktens samlingsbeställning till FOI och löper mellan 2021 och 2023. Syftet med projektet är att ta ett framåtblickande och utforskande grepp om gränssnittsområdet i syfte att kartlägga och bedöma vilka typer av idag kända gränssnitt som i framtiden kan bli relevanta för militära tillämpningar och på vilket sätt dessa kan användas inom Försvarsmaktens verksamhet.

Under 2021 gjordes en bred genomgång av potentiella gränssnitt av intresse att studera djupare (Levin m.fl, 2021). Genomgången visade att det finns ett antal olika sätt att registrera kroppsrörelser. Bland de bärbara varianterna finns handskar försedda med olika sensorer och i vissa fall även taktorer för haptisk återkoppling. Handsken kan registrera gester oberoende av omgivningens förhållanden, dock krävs viss infrastruktur för att förmedla signalerna.

## 1.3 Syfte med rapporten

Denna rapport beskriver det inledande arbetet med att utvärdera möjliga sätt att använda gestbaserad interaktion. Målsättningen för arbetet i rapporten var att skapa ett *proof of concept* för geststyrning av en UGV med hjälp av kommersiellt tillgänglig teknik. Rapporten redovisar dels en teoretisk bakgrund med fokus på

---

<sup>1</sup> Forskning- och teknikutveckling

gester som kommunikations- och interaktionsverktyg, dels resultatet av en inledande teknisk utvärdering, fördelat på två delstudier, av en motion capture-handske avsedd för interaktion med en UGV.

Syftet med den teoretiska översikten är att ge en förståelse för gester som kommunikationsmedel samt möjligheten att använda gester i interaktion, inte bara med andra människor utan också som en del i interaktionen med (och styrningen av) teknik, företrädesvis i formen av obemannade farkoster som en del av framtida team.

Syftet med den tekniska utvärderingen är att undersöka möjligheterna att med hjälp av en kommersiellt tillgänglig motion capture-handske möjliggöra styrning (och annan interaktion) med en UGV, dels genom att undersöka möjligheter och begränsningar gällande handskens förmåga att automatiskt uppfatta och särskilja handrörelser och gester, dels genom att undersöka möjligheten att nyttja handsken för direkt styrning av en UGV.

## 1.4 Läsanvisning

Kapitel 1 beskriver bakgrunden till och syftet med studien i rapporten och kontexten, det vill säga projektet i vilket denna studie ingår.

Kapitel 2 beskriver grunderna för gestbaserad interaktion, det vill säga uppbyggnaden och förståelsen av gester samt hur de kan registreras och tolkas.

Kapitel 3 beskriver utvärderingen av ett verktyg för att registrera och tolka gester, en så kallad motion capture-handske.

Kapitel 4 beskriver ett tekniskt test och demonstration av hur den utvärderade handsken kan användas för att styra en UGV.

Kapitel 5 utgörs av en diskussion av studien och det tekniska testet.

Kapitel 6 beskriver möjligt framtida arbete inom området gestbaserad styrning inom ramen för projektet.

## 2 Gestbaserad interaktion

För att kunna använda gester för interaktion, inte bara med andra människor utan även med teknik, behövs en förståelse för vilka typer av gester som används av människor, hur dessa gester är uppbyggda och kan tolkas maskinellt. Utvecklingen av gestbaserade gränssnitt bygger på ett antal huvudsakliga teoretiska och tekniska förutsättningar. Först och främst behövs en förståelse för vad och hur gester är uppbyggda och hur de kan förmedla information vilket beskrivs i kapitel 2.1. Därefter behövs tekniska förutsättningar för att dels registrera relevanta rörelser och dels tolka rörelser som en betydelsebärande gest vilket beskrivs i kapitel 2.2.

### 2.1 Gester

En övergripande målsättning inom området interaktionsdesign är att skapa en interaktion som användaren upplever som naturlig och intuitiv. Inom gränssnittsdesign betyder det bland annat en strävan efter att nyttja de sätt människor naturligt använder för att kommunicera med varandra även vid kommunikation och interaktion med teknik. Gester är en integrerad del av hur vi som varelser naturligt kommunicerar och hur vi tänker (McNeill, 1995). Att människor spontant använder kroppslig rörelse för att kommunicera kan bland annat ses i hur spädbarn kommunicerar innan de utvecklat verbalt språklig kommunikation (Rowe och Goldin-Meadow, 2009) eller i studier om hur gester används spontant i samband med talat språk (Hostetter and Alibali, 2008).

Människor använder en lång rad olika typer av gester som sträcker sig från kontextuellt beroende gestikulering till tydligt definierade gester i teckenspråk (McNeill, 2005). Gester kan definieras som *en uttryckande och betydelsebärande kroppslig rörelse* (Mitra och Acharya, 2007) och kan delas in i tre kategorier, (1) *oberoende kommunikativa gester*, vilka har ett kommunikativt syfte och kan användas oberoende av verbal kommunikation, (2) *språkrelaterade gester*, vilka har ett kommunikativt syfte och används tillsammans med och för att kompletterar verbal kommunikation och (3) *manipulerande gester*, som interagerar med eller modifierar den fysiska världen eller komponenter i ett gränssnitt (Vuletic m.fl., 2019).

Nedan diskuteras i mer detalj gester utifrån deras symbol eller form och de koncept eller betydelser som de uttrycker.

#### 2.1.1 Formen på en gest

En enskild gest kan beskrivas utifrån ett antal aspekter. Svenskt teckenspråk använder tre aspekter för att beskriva utförandet av ett handtecken (Institutet för språk och folkminne, 2021): (1) *artikulator* är formen som handen/händerna antar, (2) *artikulationsställe* är handen/händernas position och (3) *artikulation* är den

rörelse som handen/händerna utför. Motsvarande sätt att beskriva en enskild gest görs av Mitra och Acharya (2007) genom att dela upp den i följande informationsbärande element: (1) *spatial information* om var gesten görs, (2) *riktningsinformation* om vilken rörelse/riktning gester har, (3) *symbolisk information* som är vilket tecken gesten består av och de tillför även elementet (4) *affektiv information* som är gestens känslomässiga kvalitet. En gest är också statisk, dynamisk eller en kombination av dessa två. En statisk gest innebär att en viss position antas och är sedan stilla, medan en dynamisk gest består av en sekvens av flera olika positioner, vilka delas in i tre eller fem faser (McNeill, 1995): *preparation*, *prestrike hold* (gestspecifikt), *stroke*, *post stroke hold* (gestspecifikt) och *retraction*. Registrering av gester kan således behöva kunna identifiera en komplex rad av rörelser för att kunna tolka vilken gest som utförts.

Sett till handgester kan dessa också göras antingen i form av enhandsgester eller tvåhandsgester, där tvåhandsgester antingen är symmetriska eller asymmetriska. En symmetrisk tvåhandsgest innebär att båda händerna tar samma form eller rörelse men spegelvänt, medan en asymmetrisk tvåhandsgest innebär att händerna gör olika former eller rörelser som en del av gesten (McNeill, 1995; Mitra och Acharya, 2007).

### 2.1.2 Betydelsen av en gest

Det koncept eller betydelse som gesten förmedlar delas av Vuletic m.fl. (2019) in i tre kategorier:

1. *oberoende kommunikativa gester* – gester som har ett kommunikativt syfte och som användas fristående eller tillsammans med verbal kommunikation.
2. *språkrelaterade gester* – gester som har ett kommunikativt syfte och kompletterar det som kommuniceras verbalt.
3. *manipulerande gester* – gester som fysiskt interagerar med omvärlden genom att till exempel flytta på ting eller trycka på knappar.

De två första kategorierna delas upp i ett flertal mer specifika typer av gester, vilka listas nedan.

#### Oberoende kommunikativa gester delas in i följande typer:

- *Symboliska gester* representerar ett symboliskt objekt eller koncept. Betydelse hos gesterna är allmänt känt inom viss kultur och kan direkt översättas i ord.
- *Semaföriska gester* används för att initiera en förutbestämd händelse. Betydelsen hos gesterna är definierade i formaliserade lexikon och utvecklade för ett specifikt syfte, t.ex. inom en viss profession.

- Pantomimiska gester representerar familjära koncept och imiterar vad som åsyftas. Gesten utformas för att efterlikna det som refereras, till exempel att forma handen som en telefonlur och hålla mot kinden.

### **Språkrelaterade gester kan delas in i följande typer:**

- Ikoniska gester – representerar en betydelse som är nära relaterat till det semantiska innehållet i det talade språket. Gesterna illustrerar det som sägs, som former eller rörelse.
- Metaforiska gester – är som ikoniska gester, men för abstrakta koncept.
- Modalt symboliska gester – kompletterar talat språk genom att förstärka eller förtydliga någon del av ett uttalande.
- Sammanhållande gester – är upprepande gester som används för att binda ihop ett avbrutet samtal.
- Adaptor gester – är rörelser som utförs för att avlasta/släppa spänningar i kroppen.
- Deiktiska gester – är rörelser som är utpekande, t.ex. genom att peka på faktiska objekt eller abstrakta koncept.

Vad som eventuellt saknas i den ovanstående kategoriseringen är de gester som Hostetter and Alibali (2008) benämner som *interaktiva gester*, vilka används för att interagera med andra talare, till exempel för att reglera talordning.

## **2.2 Gestbaserade gränssnitt**

Även om gester kan göras med flera olika kroppsdelar har tidigare forskning om gester som gränssnitt mot teknik primärt studerat handgester (Vuletic m.fl., 2019). Termen gestbaserade gränssnitt används för både system som är *touch-based* (svepa med fingrar på en skärm) och system för *in-air* gester (Graichen, Graichen, Krems, 2017). Oavsett typ behöver systemet för att tolka gester: (1) kunna avgöra start och stopp för meningsbärande gestmönster, (2) segmentera rörelsen hos en gest, t.ex. bortse från transaktionen mellan en gest till en annan, samt (3) hantera en naturlig variation hos gester mellan olika tillfällen (Zimmerman m.fl., 1987).

### **2.2.1 Gester för interaktion med teknik**

Det har föreslagits ett flertal olika metoder för hur gestbaserade gränssnitt tekniskt kan fungera, antingen med kroppsmonterade sensor eller videobaserade lösningar (Vuletic m.fl., 2019). Omfånget för registrering sträcker sig från att fånga hela kroppens rörelser till att enbart fokusera på någon del, som händer eller ansiktet. Forskning gällande gester har primärt handlat om handgester från tidiga exempel

som DataGlove som tolkade enhandsgester (Zimmerman, m.fl., 1987) till nutida forskning (Vuletic m.fl., 2019).

Den teoretiska basen avseende kommunikativa gester gäller främst gester som utförs i samband med, eller istället för talat språk. Enligt Vuletic m.fl. (2019) skiljer sig denna användning av gester åt från de gester som använts i studier gällande gränssnitt. De menar därmed att det saknas forskning och en teoretisk bas för att avgöra vilka gester och typer av gester som är lämpliga i olika tillämpningar av gestbaserade gränssnitt. Vad som är känt är dock en speciell ergonomisk utmaning som utformning av gester behöver ta hänsyn till. Tidiga exempel på geststyrning har påvisat att långvarig interaktion med gester kan orsaka fysiska besvär, vilket kallas *gorilla arm syndrome*. Dessa besvär kan minskas genom att undvika gester som utförs ovanför hjärtat och långt bort från kroppen (Hansberger, 2017).

Medan människor i sin kommunikation använder sig av både gester med tydligt definierade betydelser och spontana gester som enbart kontextuellt ges en betydelse har forskningen avseende gestbaserade gränssnitt främst gällt det förstnämnda. I geststudier är det vanligast att använda förutbestämda och förinövade gester som är bestämda av en forskare/designer. I de studier i vilka gester med fri form används handlar det om att styra robotar, skapa/rita linjer/ytor och modifiera virtuella 3D objekt (Vuletic m.fl., 2019).

Exempel på system i vilka gester använts för att styra UGV:er innefattar ett kamerabaserat system med tio olika gester för att ge kommandon som *Initiera gesttolkning, Gå hem, Gå till mål, Snurra runt, Gå och utforska, Stanna* (Cicirelli m.fl, 2015). Ytterligare ett exempel på ett kamerabaserat system med fyra gester (*öppen hand, knuten hand, utsträckt pekfinger, utsträckt pekfinger och tumme*) för att styra en autonom robot finns i Van den Bergh, m. fl. (2011) medan (Voronchuk, Klym och Dunets (2019) presenterar ett handskbaserat system med tröghetsensorer och böjningssensor för att med pekfingret kunna styra en UGV. I en Youtube-film från 2018 demonstrerar ett företag för militära UGV:er en kamerabaserad geststyrning i vilka användare med stora armrörelser kan ge fyra olika kommandon: starta motor, stanna motor, kör framåt, kör bakåt (Trotsenko, 2018).

## 2.2.2 Registrering och tolkning av gester

Rörelser kan beröringsfritt registreras på ett flertal tekniska sätt, vilket redovisas i Levin m.fl. (2022). Teknikerna kan delas in i metoder som antingen registrerar rörelse på distans, t.ex. visuellt, IR och radar, och metoder som använder kroppsmonterade sensorer, t.ex. tröghetsmätare (Vuletic m.fl., 2019). Vuletic m.fl. noterar också att det saknas ett standardramverk för hur tolkning av kroppsrorelserna till gester ska ske och att många olika metoder (statistisk analys, maskininlärning osv) används i den akademiska forskningen, vilket också påverkas av att de olika tekniska sensorerna har olika förmågor i hur de registrerar

rörelser. Detta kan bland annat gälla huruvida registreringen kan ge absolut positionering eller enbart relativa förändringar i rörelse.

Två exempel på metoder för att tolka rörelser till gester är regeldrivna och datadrivna metoder.

- **Ett regeldrivet system** baseras på att olika rörelser tydligt har definierats till att tolkas som specifika gester, t.ex. genom att identifiera en knuten hand baserat på medelvärdet av fingrarnas böjning. En utmaning med att robust definiera mer än enstaka gester finns i den naturliga variationen i hur en person utför samma gest vid olika tillfällen samt i variationen mellan hur olika personer utför gester.
- **Datadrivna metoder** använder istället matematiska modeller och algoritmer som baserat på insamlad data av kroppsrörelser statistiskt kan beräkna sannolikheten för att vissa mätvärden motsvarar en viss gest. Ett exempel på detta är det som benämns som maskininlärning. Dessa metoder kan automatiskt identifiera vilka mätpunkter som är viktiga för att särskilja olika gester och generalisera denna tolkning för att hantera naturlig variation mellan tillfällen och personer.

## 3 Utvärdering av motion capture-handske

Detta kapitel redovisar en studie som utforskar de tekniska förutsättningarna med beröringsfri registrering av handrörelser med hjälp av burna sensorer monterade i en kommersiellt tillgänglig motion capture-handske. Studien använder två olika typer av gester: ett urval av så kallade fasta handformer från svenskt teckenspråk och ett urval av militära handtecken.

De frågeställningar som studien utforskar är:

- Vad är de tekniska förutsättningarna med föreslagen lösning att identifiera och särskilja mellan fasta handformer från svenskt teckenspråk?
- Vad är de tekniska förutsättningarna med föreslagen lösning att identifiera och särskilja olika handtecken som används militärt för gruppkommunikation?

Syftet med studien är att skapa en plattform för vidare arbete med gestbaserade gränssnitt för att konceptuellt kunna demonstrera och utvärdera denna typ av gränssnitt i militära användningsfall, med speciellt fokus på interaktion med autonoma enheter, och att stödja interaktion mellan människor.

Studien är avgränsad vad det gäller teknisk implementation och omfattning i utvärdering. Urvalet av gester som systemet ska registrera är styrt baserat på bedömd förmåga hos den tekniska lösningen. Det som beaktas är bland annat att handsken inte kan registrera alla typer av rörelser eller positioner med den aktuella uppsättningen sensorer (t.ex. rakt vertikala rörelser). Gesterna är begränsade till enhandsgester och primärt statiska gester. Hantering av dynamiska gester bedöms kräva en egen fokuserad delstudie.

### 3.1 Deltagare

Deltagarna i studien rekryterades bland anställda och studenter knutna till myndigheten FOI. Totalt deltog tio personer, fördelat på sex kvinnor och fyra män. Deltagarnas åldersspann hos deltagarna var 24 till 56 år (medel: 32.7 år, sd: 11.8). En deltagare angav en preferens att använda vänsterhand, medan övriga hade en preferens för högerhand. Ingen deltagare angav nedsatt rörlighet i handen. I Tabell 1 redovisas storleken på varje deltagares hand, uppmätt i enlighet med ISO (2017).



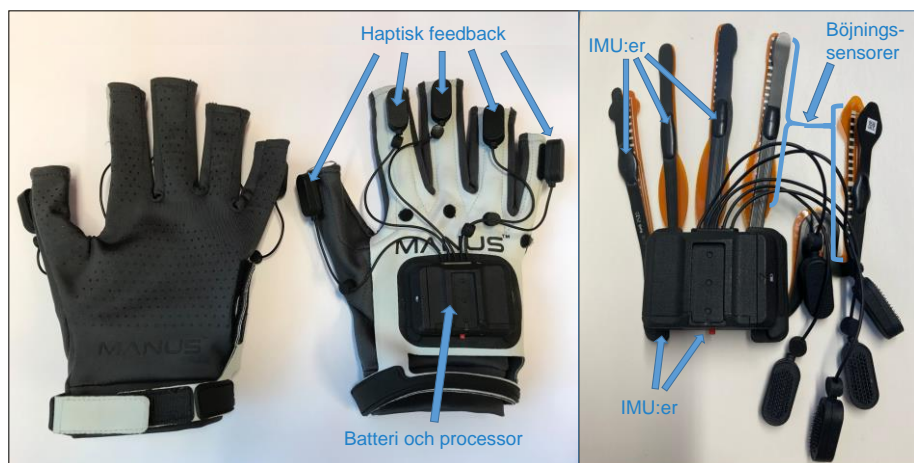
Tabell 1 Mått för deltagarnas händer redovisat i cm.

	Omkrets över knogar	Handens längd	Handflatans längd	Pekfingrets längd	Tummens längd
<b>Min / max</b>	18,5 / 21,0	17,0 / 12,0	9,5 / 12,0	6,5 / 8,0	5,5 / 7,5
<b>Genomsnitt (sd)</b>	19,3 (0,82)	18,4 (0,94)	10,4 (0,69)	7,1 (0,57)	6,6 (0,62)

Data knuten till enskilda deltagare pseudonymiserades genom att tilldela varje deltagare en unik identifieringskod som använts i efterföljande databehandling.

## 3.2 Utrustning

Registrering av handrörelser skedde med hjälp av motion capture-handsken Prime II Haptic, tillverkad av företaget Manus<sup>2</sup> (se Figur 1 samt Tabell 2 för ytterligare tekniska specifikationer). Handskarna har fem flexsensorer som mäter fingrarnas individuella böjning (ett värde för knogled och ett för mellanled) och sex tröghetsmätare (eng. Inertial Measurement Unit, IMU). Tillgång till mätvärden från sensorer sker trådlöst via en proprietär lösning till en dator med programvara från Manus. Programmet erbjuder inte direkt tillgång till sensorernas rådata, utan enbart ett antal sammanvägda beräknade mätvärden. Dessa mätvärden var i studien: två värden för graden av böjning hos respektive finger och fyra värden för handryggens rotation i tre dimensioner.



Figur 1 Bild till vänster: Placering av taktorer och dess kablage för haptisk feedback. Bild till höger: Handskens innanmäte med IMUer och flexsensorer. Den haptiska återkopplingen nyttjades inte för att registrera träningsdata.

<sup>2</sup> <https://www.manus-meta.com/>

Tabell 2 Teknisk specifikation för Prime II Haptic

Egenskap	Specifikation
Signalfördröjning	< 5ms
Avläsning av sensor	90 Hz
Vikt	60 g
Flexsensor	10 st (två per finger), 2 DoF <sup>3</sup>
IMU	6 st (fingrar och handrygg), 9 DoF
Noggrannhet	+/- 2.5 grader

Typen av sensorer och urvalet av tillgängliga mätvärden innebar en rad begränsningar i vad som kunde registreras, vilket inneburit att det saknats tekniska förutsättningar för att mäta handens rörelser i rak vertikal eller horisontell riktning och handens position i förhållande till kroppen.

### 3.3 Utformning av gester

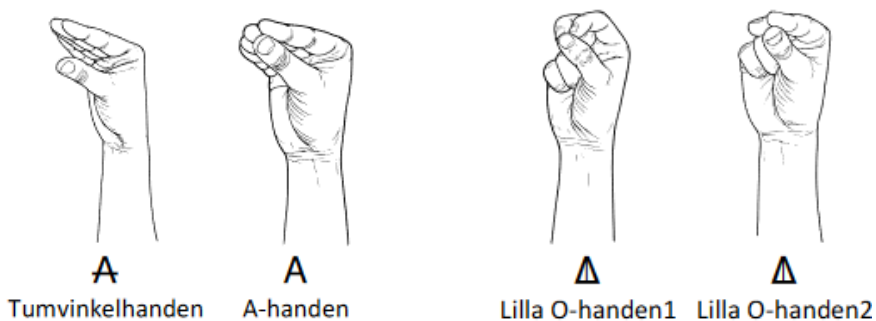
I studien användes två uppsättningar av gester hämtade från (1) svenskt teckenspråk och från (2) Försvarmaktens definierade arm/handtecken. Bilder på samtliga gester återfinns i Appendix A. Skillnaden mellan gester var primärt baserad på deras form (artikulator), men även i några fall för de militära handtecknen en kombination av form, position (artikulationsställe) och rörelse (artikulation). Detta gäller till exempel tecknen för *Allt klart*, *Samband sökes*, *Färdiga till strid* och *Framåt*, som alla görs med samma form (flat hand), men med olika position och rörelse.

Svenskt teckenspråk innehåller 42 fasta handformer. Flera av dessa har en snarlik form, men med till exempel små skillnader i fingrarnas vinkel, som den använda tekniska utrustningen har speciellt svårt att skilja på (Figur 2). Av detta skäl gjordes ett urval av de fasta handformerna och totalt 29 handformer inkluderades. I Tabell 3 ges en namnöversikt över gesterna som användes.

Försvarmakten har 38 definierade tecken för användning *vid all verksamhet*<sup>4</sup> (SoldF, s. 441-445) och studien använder elva av dessa. Urvalet exkluderar tecken som kräver två händer eller innehåller en rörelse som enbart är vertikal eller horisontal, då dessa rörelser inte kan registreras med den aktuella utrustningen.






























<sup>3</sup> Frihetsgrader (eng. Degrees of Freedom)

<sup>4</sup> Utöver dessa beskrivs även larmtecken, tecken för kolonnkörning, körning med ledare och tecken för helikopter.



Figur 2. Illustration av två fall med snarlika handformer där en form från respektive fall exkluderades, i dessa fall Tumvinkelhanden och Lilla O-handen 2. (Publiceras med tillstånd av Svenskt teckenspråkslexikon vid avd. för teckenspråk på Stockholms universitet).

Tabell 3. Lista på de handformer som användes i studien (se Appendix A för större bilder)

Namn på handform			
4 Handen 	L handen 	O handen 	Stora nyphanden 
A Handen 	Lilla O handen 1 	Pekfingret 	T handen 
Dubbelkroken 	Lillfingret 	Raka mätthanden 	Tumhanden 
Flata handen 	Långfingret 	Runda långfingret 	Tupphanden 
Flyghanden 	Mätthanden 	Sprethanden 	U handen 
Klohanden 	N handen 	Stora hållhanden 	V handen 
Knutna handen 	Nyphanden 1 	Stora långfingret 	Vinkelhanden 
Krokfingret 			

### 3.4 Genomförande

Datainsamlingen med handsken genomfördes med en deltagare per tillfälle. Varje tillfälle inleddes med att deltagaren delgavs skriftlig information om försöket, varefter denne godkände nyttjandet av insamlad data. Därefter samlades uppgifter in om deltagaren samt kroppsmått för dennes högerhand.

Innan datainsamlingen påbörjades presenterades de uppsättningar av gester som försöksdeltagaren skulle utföra och deltagaren fick öva på att utföra respektive gest. Försöksledaren kontrollerade att deltagaren utförde gesten på korrekt sätt och korrigerade vid behov.

Datainsamlingen var uppdelad i två omgångar, men genomfördes på samma sätt. Den första omgången innehöll de fasta handformerna och genomfördes sittande medan den andra omgången som innehöll de militära handtecknen, genomfördes stående.

För att styra insamlingen av data vid genomförandet användes en programvara för att dels instruera deltagarna i vilka gester som skulle göras, dels lagra alla mätvärden från handsken tillsammans med nödvändig metadata. Programmet presenterade löpande de gester som deltagarna skulle utföra i en följd och samlade in ett förutbestämt antal datapunkter per gest.

Inför varje omgång ombads deltagaren att positionera sig framför en datorskärm med motion capture-handsken på höger hand. På datorskärmen visades löpande den gest som deltagaren skulle utföra. Inför varje gest fick deltagaren sju sekunder på sig att anta korrekt gest innan handens position började registreras. Två hundra registreringar samlades sedan in under ungefär två sekunder. När tiden för en gest gått ut visades omedelbart nästa gest. Varje omgång bestod av tre genomföranden, en testiteration och därefter två iterationer där data samlades in (Tabell 4). Testiterationen användes endast för att försökspersonerna skulle bekanta sig med systemet. Alla gester visades en gång per iteration i slumpmässig ordning.

Tabell 4 Förklaring av genomförandet

<b>Dataset</b>	<b>Position</b>	<b>Ordning</b>
<b>Omgång 1: fasta handformer, sittandes.</b>		
Teckenspråk: Testiteration	Sittande	Fast
Teckenspråk Iteration A	Sittande	Slumpvis
Teckenspråk Iteration B	Sittande	Slumpvis
<b>Omgång 2: militära handtecken, ståendes.</b>		
SoldF: Testiteration	Stående	Fast
SoldF Iteration A	Stående	Slumpvis
SoldF Iteration B	Stående	Slumpvis

Under datainsamlingen kontrollerade försöksledaren att deltagarna utförde gesterna på ett korrekt sätt. Vid avvikelser påtalade försöksledaren detta för deltagarna för att de skulle utföra gesten korrekt. Försöksledaren noterade också problem att utföra gesten korrekt i ett försöksprotokoll.

## 3.5 Identifiering av gester

Tre olika maskininlärningsmodeller användes i delstudien för att klassificera gester baserat på data från handsken: logistisk regression, ett enkelt så kallat *fully-connected artificiellt neuronnät* och ett *Long Term Short Term Memory*-nätverk. Dataanalysen skedde i Python i vilken den logistiska regressionen implementerades med scikit-learn<sup>5</sup>, medan ANN-modellerna implementerades med PyTorch<sup>6</sup>. Resultatet från respektive modell var en beräknad sannolikhet för varje gest i uppsättningen av gester givet ett visst indata.

*Modell 1: Logistisk regression (LR)*. Denna modell användes som baseline och använde standardvärden för hyperparametrar<sup>7</sup> enligt version 1.0.2 av scikit-learn.

*Modell 2: Enkelt fully-connected neuralt nätverk (FC)*. Denna modell använde ett artificiellt neuralt nätverk bestående av två dolda lager i form av fully-connected lager med 64 noder. Fully connected innebär att alla noder i ett lager är kopplade till alla noder i föregående lager. Efter varje dolt lager användes även ett dropout-lager<sup>8</sup> med sannolikhet satt till 0,4. Outputlagret bestod av en nod per gest i aktuell gestuppsättning.

*Modell 3: Neuralt nätverk med minnesfunktion (LSTM)*. Denna modell använde ett artificiellt neuralt nätverk bestående av dolda lager i form av två LSTM-lager med 64 noder vardera och två fully-connected lager med 64 noder vardera. LSTM är en teknik för hur artificiella nät kan ackumulera information över sekvenser av data igenom återkopplade lager (eng. *recurrent layers*) (Goodfellow m.fl., 2016, s. 404). Efter varje dolt lager fanns ett dropout-lager med en definierad sannolikhet satt till 0,2. Outputlagret bestod av en nod per gest i aktuell gestuppsättning.

### 3.5.1 Dataförsörjning

Modellerna tränades separat för respektive uppsättning av gester, det vill säga tre modeller för handformerna från teckenspråk och tre modeller för militära handtecken. Data från en iteration vid datainsamlingen användes som träningsdata och data från den andra iterationen som testdata, för att utvärdera modellerna mot.

---

<sup>5</sup> <https://scikit-learn.org/stable/index.html>

<sup>6</sup> <https://pytorch.org/>

<sup>7</sup> Inställningar för hur anpassning av modellen sker

<sup>8</sup> Dropout-lager används för att motverka att modellen blir överspecialiserad (Goodfellow m.fl., 2016, s. 255), vilket skulle försämra förmågan att hantera naturlig variation i indata.

### 3.5.2 Indata för modellerna

För att identifiera gester som inkluderar en rörelse krävs det att modellen kan hantera temporal data, vilket innebär att modellen kan hantera en serie av registreringar där gester kan och får variera över tid. Ett sätt att lösa detta är att föra ett fönster över de registrerade datapunkterna och associera sekvensen inom fönstret med en gest. I detta fall valdes en fönsterlängd om 100 registreringar vilket betyder att indata för modellen består av 100 stycken registreringar à 24 värden. Denna metod användes för samtliga gester, oavsett om gester innehöll rörelse eller ej.

LR- och FC-modellerna använde sig av ett indatalager som bestod av lika många noder som antalet indatapunkter, i det här fallet alltså 2400 stycken. Dessa modeller fick ingen explicit fingervisning om att det handlade om sekvenser av data. LSTM-modellen hade ett indatalager av endast 24 noder, då hantering av sekventiell data är inbyggt i denna typ av modeller. Nyttjandet av LSTM-lager innebär att modellen har ett minne som gör att tidigare data påverkar bearbetning av ny stimuli.

### 3.5.3 Träningmetodik

Samtliga tre maskininlärningsmodeller tränades tre gånger varav den bästa valdes ut med avseende på slutgiltig *training loss* som är ett mått på hur väl modellen lärt sig prediktera träningsmängden.

För FC- och LSTM-modellerna användes standardvärdena för hyperparametrar i PyTorch med undantag för följande: antal epoker sattes till 150, learning rate sattes till  $0,01 * 0,98^{\#epok}$ , batch-storleken till 200 och som optimeringsalgoritm användes *Adam* (Goodfellow m.fl, 2016, s. 305).

## 3.6 Resultat

Prestationen hos respektive modell utvärderades baserade på deras precision och sensitivitet och med så kallade *Precision Recall Characteristics*-kurvor (Saito och Rehmsmeier, 2015). För att jämföra modellerna med varandra användes måttet AUC (Area under curve). Resultatet visar att modellen FC-A hade högst AUC-värde för omgången med handformer från svenskt teckenspråk, medan modellen LSTM-A hade högst AUC-värde för de militära handtecknen (Tabell 5).

Gränsvärdet för vid vilken sannolikhet modellens klassificering ska accepteras sattes till 0,9 för att prioritera att minska mängden felaktiga klassificeringar. Klassificeringar med lägre sannolikhet än 0,9 räknas som ingen klassificering.

Med modellen FC-A klassificeras hälften av (12 st) de fasta handformerna korrekt (Tabell 6) med lägre än 15% falska prediktioner (kombination av felaktig och avsaknad av klassificering). Modellen LSTM-B klarar av att identifiera ca en tredjedel (4 st) av militära handtecknen (Tabell 7).

Tabell 5 Beräknat AUC-värde för respektive modell och träning/test-set. Det högsta värdet per gestuppsättning är fetmarkerat.

<i>Modell</i>	<i>LR-A</i>	<i>LR-B</i>	<i>FC-A</i>	<i>FC-B</i>	<i>LSTM-A</i>	<i>LSTM-B</i>
<b>AUC</b>						
<i>Handformer (Svenskt teckenspråk)</i>	0,824	0,835	<b>0,855</b>	0,844	0,815	0,776
<i>Militära handtecken</i>	0.831	0,833	0,831	0,833	0,923	<b>0,933</b>

Tabell 6. Gester där modellen (FC-A) korrekt klassificerar gester med över 0,9 sannolikhet och med lägre än 15% falska prediktioner.

Gest	Felaktig klassificering	Ingen klassificering	Falsk prediktion
<i>4-handen</i>	0,0%	0,0%	0,00%
<i>Dubbelkroken</i>	0,0%	11,1%	11,10%
<i>Flyghanden</i>	0,0%	0,0%	0,00%
<i>Klohanden</i>	0,0%	11,1%	11,10%
<i>Knutna handen</i>	0,0%	13,3%	13,30%
<i>L-handen</i>	0,0%	10,9%	10,90%
<i>Lillfingret</i>	0,0%	0,0%	0,00%
<i>Långfingret</i>	0,0%	0,0%	0,00%
<i>Pekfingret</i>	0,0%	0,0%	0,00%
<i>Runda långfingret</i>	0,0%	0,0%	0,00%
<i>Tupphanden</i>	0,0%	10,9%	10,90%
<i>U-handen</i>	4,0% <sup>1</sup>	6,1%	10,10%








































<sup>1</sup> Felaktig klassificering som *Lillfingret*

Tabell 7 Gester där modellen (LTSM-B) korrekt klassificerar gester med över 0,9 sannolikhet och med lägre än 15% falska prediktioner.

Gest	Felaktig klassificering	Ingen klassificering	Falsk prediktion
<i>Egen trupp</i>	0,0%	0,0%	0,00%
<i>Fientlig fotrupp</i>	0,0%	0,0%	0,00%
<i>Fientligt pansarskyttefordon</i>	0,0%	0,0%	0,00%
<i>Mina i spåret</i>	0,0%	0,0%	0,00%

Modellerna har svårigheter att särskilja ett flertal gester från varandra, då skillnaden beror på små skillnader i vinkel hos fingrarna eller vinkelskillnader som inte mättes under insamlingen (fingrarnas spretning). I Tabell 8 redovisas felaktiga klassificeringar på handformer som modellen FC-A gjorde för att illustrera vad för gester som är svårare att särskilja.

Tabell 8 Översikt av handformer i vilka det förekommer förväxlingar.

Utförd gest	Felaktiga klassificering	Utförd gest	Felaktiga klassificering
 A-handen	 Nyvohanden1  Vinkelhanden	 Sprethanden	 J Flata handen
 Knutna handen	 O-handen	 Stora långfingret	 Runda långfingret
 Lilla O-handen1	 O-handen  Mättanden	 Stora nyvohanden	 Stora hållhanden
 Mättanden	 Krokfingret  Lilla O-handen2	 T-handen	 Raka mättanden  Tumhanden
 N-handen	 V-handen	 B Tumhanden	 Mättanden
 Nyvohanden1	 Raka mättanden	 U-handen	 I lillfingret
 O-handen	 Klophanden  Knutna handen  Lilla O-handen1  I lillfingret  Vinkelhanden	 Vinkelhanden	 A-handen  Tumhanden
 Raka mättanden	 T-handen		



### 3.7 Diskussion

Studien demonstrerar en teknisk lösning där det är möjligt att med använd utrustning särskilja olika gester, militära handtecken och handformer från svenskt teckenspråk. Medan de militära gesterna redan har en definierad betydelse motsvarar handformerna snarare den uppsättning av *möjliga* handformer som kan användas för att skapa specifika gester med en angiven betydelse. De fasta handformerna representerar således endast aspekten artikulatorn, som tillsammans med aspekterna artikulationsposition och artikulation bildar en unik gest som kan tillskrivas en specifik betydelse.

Resultaten bekräftar de utmaningar som finns när det gäller gester som använder samma eller snarlik artikulator och endast särskiljs genom artikulation eller artikulationsställe. Detta är något som gäller ett flertal militära handtecken – *Allt klart*, *Samband sökes*, *Färdiga till strid* och *Framåt* – som använder samma artikulator (*Flata handen*), men med olika artikulation och artikulationsställe. För att särskilja dessa gester från varandra krävs ett system som kan mäta relativa rörelser och handens positionering i relation till kroppen. I den utrustning som användes var det vid detta läge inte möjligt att extrahera data gällande handens acceleration, utan mätvärdena gav enbart handens rotation vid ett visst tillfälle. De fasta handformerna var alla statiska gester, medan ett par av de militära handtecknen inkluderade rörelse. Det är dessa gester med rörelse som avgörande särskiljande drag som innebär att de båda LSTM-modellerna presterade bättre än de andra modellerna för denna uppsättning gester. Modellen klarar dock inte av att med hög konfidens särskilja dessa gester med samma artikulator, vilket är orsaken till att enbart ett fåtal militära handtecken kunde särskiljas.

Resultaten ger en indikation på vad aktuell utrustning klarar av, även om det troligen skulle gå att förbättra resultatet med vidare justering av maskininlärningsmodellen och ytterligare träningsdata. Den uppsättning av gester som denna utrustning och modell klarar av att särskilja kan dock användas för fortsatta försök under tiden som modellen kan förfinas. Ytterligare förbättring för att särskilja dessa typer av gester skulle uppnås genom att få ut mätvärden gällande fingrarnas separation samt värden gällande handens acceleration.

Vid insamlingen höll försökspersonen gesten antingen statiskt eller gjorde en upprepande rörelse under en period. Det kan därför förväntas att datapunkterna för en statisk gest vid samma tillfälle är lika varandra och därför inte kan användas både för träning och för test. Datasetet delades därför upp enligt insamlingsiterationerna, men med nackdelen att detta enbart gav ett insamlings-tillfälle per gest och försöksperson att träna modellen på. För att säkerställa att fler datapunkter kan användas för träning respektive test skulle mer data behöva samlas in, men med flera och kortare iterationer för att fånga många *olika* tillfällen då en gest formas.

## 4 Tekniskt test av gestbaserad manuell styrning av UGV

En tillämpning för gestbaserad interaktion är styrning av enheter som UGV:er. Detta kan sedan i förlängningen kombineras med andra interaktionssätt för att nå närmare en intuitiv interaktion inom ramen för ett antal olika scenarier där en UGV utgör en del av ett team.

För att testa de tekniska förutsättningarna för att manuellt styra en UGV med hjälp av en motion-capture handske (som beskrivs i Kapitel 3) implementerades en variant av manuell geststyrning som bygger på fingrarnas utsträckning för att kontrollera fart och handens rotation för att kontrollera rattutslag. Denna implementation testades i ett mindre *proof of concept*-försök för att samla erfarenhet inför framtida utvärderingar med Försvarmakten.

### 4.1 Deltagare

Då försöket enbart var ett tekniskt test av den implementerade styrningen med handskarna utgjordes försöksdeltagarna av projektgruppen samt ytterligare två FOI-anställda.

### 4.2 Utrustning

Utöver handskarna (som beskrivs utförligt i Kapitel 3.1) användes i proof of concept-testet en fyrhjuling UGV av modell Husky från Clearpath Robotics (Figur 5). UGV:n används vid FOI som forskningsplattform för bland annat autonom navigering med hjälp av olika sensorer och utgör även en plattform för tester av olika användargränssnitt (Rantakokko m.fl., 2021). Huskyn är 100 cm lång, 67cm bred, väger 50kg och kan bära 20 – 70kg i nyttolast beroende på terrängen. Manuell styrning av Huskyn sker som standard med en vanlig handkontroll. UGV:n använder ROS (eng. Robotic Operating System) som ramverk för intern kommunikation med sensorer och styrdon. ROS är ett vanligt förekommande ramverk i forskningssammanhang för att styra samt interagera med bland annat UGV:er.

Huskyns styrning sker med så kallad sladdstyrning (eng. skid-steer) vilket innebär att fordonet svänger runt sin axel genom att hjulen på varsin sida snurrar i motsatt riktning.

### 4.3 Utformning av gester

Utformning av lämpliga gester är utmanande när det gäller gestbaserad interaktion. Gesternas utformning behöver ta hänsyn till vilka typer av gester som det tekniskt

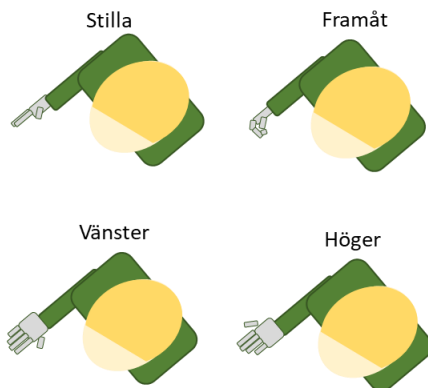
är möjligt att registrera samt vilken typ av kommandon (t.ex. diskreta eller kontinuerliga) som gesten ska signalera och alltså vara ergonomiskt lämpliga. I detta test var fokus på manuell styrning och till viss del förprogrammerad styrning (Tabell 9) då UGV:n vid tillfället för testet saknade förmåga för styrning över kategori nivå 2.

Tabell 9. Kategorisering av sätt att styra en UGV beroende på förmåga till autonom styrning.

#	Kategori	Beskrivning
1	Manuell styrning	Direkt och kontinuerlig kontroll via styrdon.
2	Förprogrammerad styrning	Automatiserad manuell styrning då diskreta kommandon aktiverar förutbestämda regelbaserade beteende, t.ex. kör 10 m framåt och stanna.
3	Reaktiv styrning	Farkosten tar viss hänsyn till omgivningen som påverkar dess framfart, t. ex. undviker hinder genom att köra runt eller följer efter en viss operatör.
4	Uppgiftsstyrning	Farkosten hanterar självständigt styrning för att nå ett visst mål, t.ex. att själv välja väg för att nå en viss målpunkt.

I detta test implementerades därför följande preliminära styrsystem för att manuellt styra en UGV (Figur 3):





- Gaspåslag genom hur mycket handen är knuten. Öppen hand ger ingen gas, medan stängd hand ger full gas.
- Handledens rotation styr rattutslaget. Handen med tummen riktat uppåt motsvarar inget rattutslag medan handen riktad ca 90° åt respektive vänster och höger betyder fullt rattutslag.
- Böjd tumme motsvarar gas framåt medan uppsträckt tumme betyder bakåt.



Figur 3. Illustration över de gester som användes för att manuellt styra UGV:n.

Utöver manuell kontroll var även fyra förprogrammerade beteenden definierade, vilka aktiverades med fyra olika fasta handformer baserade på svenskt teckenspråk (Tabell 10). Syftet med dessa var inte att utvärdera gesterna som sådana, utan att börja resonera om det är ett rimligt sätt att utföra enklare kommandon hos UGV.

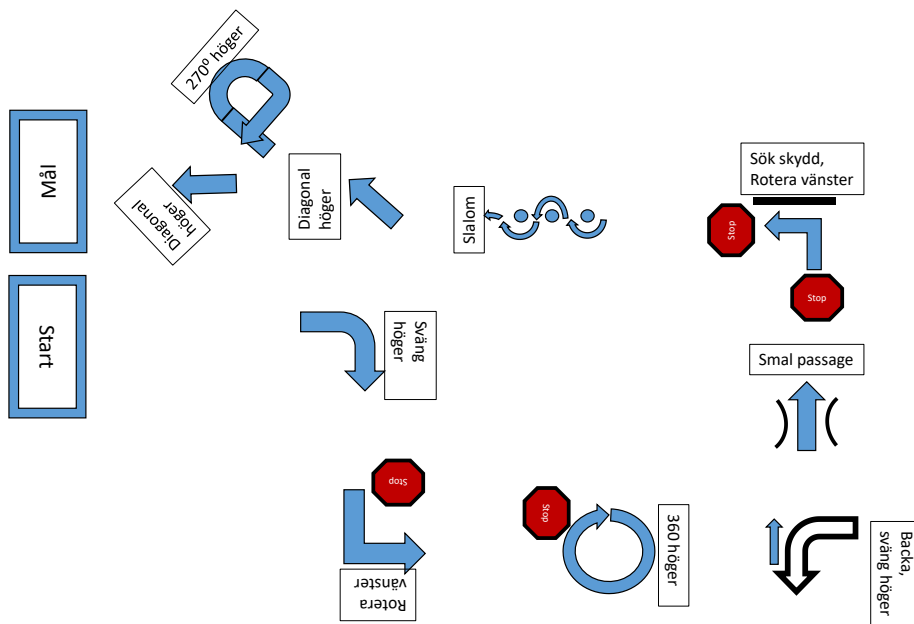
Tabell 10 Gester som används för att aktivera förprogrammerade beteenden

Gestnamn	Gest	Beteende
O-handen		Köra runt i en ring genom att hålla konstant fart framåt och konstant rattutslag.
Flyghanden		Köra fram och tillbaka genom att skifta farten framåt eller bakåt i jämna intervall
Måtthanden		Köra i en fyrkant genom att hålla konstant rattutslag under ett kort tag med jämna intervall.
Flata handen		Stanna genom att nollställa fart och rattutslag.

## 4.4 Genomförande

Testet genomfördes på en gräsmatta utomhus där deltagarna med manuell styrning fick navigera UGV:n genom en markerad hinderbana. Hinderbanan var utformad för att undersöka olika aspekter av manuell styrning (Figur 4). Finmotoriska aspekter som slalomelement och smala passager ställer relativt höga krav på att handskarna är exakta och att det är enkelt att styra oavsett hur operatören får feedback. Deltagarna styrde Huskyn genom hinderbanan, dels med visuell kontakt, dels genom en kamera fastmonterad i fronten på Huskyn (Figur 5). Videobilden från kameran presenterades på en datorplatta som operatören hade monterad på bröstet (Figur 6).

Efter genomförandet av hinderbanan fick även deltagarna testa att aktivera de olika förprogrammerade beteendena.



Figur 4. Översiktsbild över hinderbanan som användes.



Figur 5. UGV med en framåtriktad kamera. Högst upp sitter en lidarsensor monterad, men denna användes inte i det aktuella testet.



Figur 6. Operatör som styr UGV och som har en datorplatta på bröstet som visar en bild från UGV:ns kamera.

## 4.5 Resultat

Resultatet från testet visar att det var möjligt för deltagarna att navigera sig genom banan med den implementerade styrningen. Vissa personer upplevde dock problem med att handsken var för känslig, vilket fick resultatet att UGV:n reagerade för snabbt, alternativt för okänslig, vilket resulterade i att UGV:n reagerade för långsamt för att erbjuda full kontroll.

De fall då deltagarna även nyttjade en datorplatta där en filmström kunde ses från den kamera som var monterad på UGV:n kunde operatören se hinderbanan från UGV:ns perspektiv. Denna möjlighet uppskattades av några deltagare, medan det resulterade i viss rörelsesjuka för en deltagare.

## 4.6 Diskussion

Erfarenheter från det tekniska testet visar att det finns ett utvecklingsbehov för att kunna göra personliga justeringar avseende olika former av tröskelvärden för vad som ska registreras som styrkommandon.

I framtida studier är ambitionen att kombinera gestbaserad interaktion med andra former av hjälpmedel/interaktionssätt för att nå närmare en intuitiv interaktion med en UGV för olika scenarier. I takt med att förmågan hos UGV:n ökar avseende autonomi är det också av intresse att undersöka hur gestbaserad interaktion kan användas inte bara för manuell styrning, utan mer för att inrikta autonoma beteenden.

Relevanta användningsfall som kan användas för framtida studier är:

- Spaning med UGV utomhus (avsökning av en yta/område, rekognosering) inför exempelvis en framryckning
- Genomsök med UGV inomhus (avsökning av en yta/rum, rekognosering) för att leta efter objekt och människor

Syftet med att integrera ett system för gestbaserad interaktion med styrningen av en UGV var att sätta in tekniken i en kontext som är specifik och realistisk för Försvarsmakten och ge användare möjlighet att själv prova i en fysisk miljö. Målet är således att utforska möjligheter, begränsningar och tekniska utmaningar med denna typ av interaktionssätt genom prototyper, inte att utveckla ett styrsystem med en hög HRL-nivå (eng. Human Readiness Level, Salazar, m.fl., 2020). .

## 5 Diskussion

Gester är en del av naturlig mänsklig kommunikation och gester kan användas för att interagera med teknik. Hur gester ska kunna användas för att kommunicera och interagera inom team som inkluderar autonoma enheter som UGV:er är dock inte lika väl utrett. Sett till forskningen om gestbaserade gränssnitt framstår det som att fokus under 2010-talet primärt har varit att utforska hur själva registreringen av gester kan ske samt att presentera olika prototyplösningar för hur gester kan användas i relation till robotar. Mindre fokus tycks ha legat på vilka gester som ska användas och hur dessa utvärderas. Studierna i denna rapport utgör en inledande del i det arbetet – att undersöka hur en motion capture-handske kan fånga upp gester och omvandla dem till ett ingångsvärde för interaktion med en UGV.

Studierna visar att det går att nyttja motion capture-handskar för interaktion med en UGV, men det krävs fortsatt arbete för att komma fram till vilka gester som lämpar sig för olika scenarier. Det krävs även fortsatt arbete för att justera maskininlärningsmodellen som ligger till grund för registrering och tolkning av gesterna.

### 5.1 Återkoppling

I riktlinjer för god användbarhet och interaktionsdesign finns vanligen *återkoppling* på en användares interaktion. I fall med mus- och tangentbord kan den återkopplingen utöver vad som sker på skärmen eller motsvarande vara ljud eller känsla av att klicka på en fysisk knapp eller skärm. För gestbaserad interaktion saknas det en vedertagen best practice för hur återkopplingen ska utformas - visuellt, auditivt eller haptiskt.

I fallet med att styra en UGV med handgester kommer återkopplingen kanske främst bestå av att användare ser (eller hör) UGV:ns rörelser. Detta är dock inte fallet då UGV:n rör sig utom synhåll för användare, då andra typer av återkoppling måste användas. I det tekniska testet som redovisas i denna rapport nyttjades en kamera för att användarna skulle kunna följa UGV:ns färd via en videoström. Dock ses det då ur UGV:ns perspektiv, vilket kan göra det svårare att få en överblick då det blir svårt att se UGV:ns position i relation till omvärlden via enbart kameraströmmen. Genom att nyttja ytterligare stöd i form av en realtidsuppdaterad karta skulle användarna få en bättre överblick.

Om den enda återkoppling som ges är de rörelser UGV:n gör kan det vara svårt för operatörer att uppfatta om styrkommandon tolkats korrekt. Detta gäller speciellt kommandon som inte exekveras momentant, till exempel initiering av förprogrammerade beteenden. Den handsken som använts i studien har taktorer inbyggda vilket i framöver skulle kunna ge möjlighet att ge haptisk återkoppling

med exempelvis tryck/vibrationer. Mer precist hur en sådan haptisk återkoppling ska utformas krävs dock vidare studier och tester för att komma fram till.

## 5.2 Valet av gester

I studien utgick urvalet av gester dels från svenskt teckenspråk, dels från Försvarsmaktens definierade arm/handtecken. Dessa uppsättningar av gester är utvecklade i andra syften än att användas för att kommunicera med teknik. I team bestående av människor och autonoma enheter finns det i dagsläget inte någon vedertagen gest- eller teckenuppsättning. En fara med att exempelvis använda gester som redan har en vedertagen betydelse i ett nytt sammanhang är att det finns en risk för sammanblandning av de mänskliga teammedlemmarna. Fördelen med att till viss del nyttja de vedertagna gesterna är att de redan är kända och kan användas på samma sätt i kommunikationen med UGV:n som i kommunikationen mellan övriga gruppmedlemmar. En stoppsignal för hela gruppen kan då med motion capture-handskens hjälp även förstås av UGV:n och operatören behöver således inte använda någon annan teknik (styrdon) för att förmedla detta.

Inom ramarna för studien som beskrivs i denna rapport finns ett antal kriterier för geststyrning i team i vilka både människor och autonoma enheter ingår:

- Gesten ska kunna uppfattas och förstås av en maskin.
- Gesten ska uppfattas och förmedlas till människa.
- Gesten återges på distans till människa.
- Gesten ska kunna uppfattas och förstås av en maskin och någon specifik människa som är direkt föremål för gesten.
- Gesten ska kunna uppfattas och förstås av en maskin och olika människor, alltså inte bara den som är direkt föremål för gesten.

Dessa kriterier kan utgöra en grund för att framöver utvärdera de gester som tas fram som interaktions- och kommunikationsverktyg. I dagsläget klarar den lösning som utvärderats i den här rapporten den första och fjärde punkten. Det vill säga att gesten kan uppfattas och förstås av specifik människa och maskin. Målsättningen framöver inkluderar att möjliggöra ytterligare aspekter av gestbaserad kommunikation i team, där olika scenarier kan utvärderas för att svara på projektets mer övergripande frågeställningar om hur och när denna typ av kommunikation (och interaktion) är möjlig och lämplig.



## 6 Framtida arbete

Grunden för ett fungerande team är god kommunikation (att veta hur och när man ska kommunicera), teamkoordinering (gemensam lägesuppfattning, ömsesidig prestationsövervakning, hjälpbeteende, anpassningsförmåga och ledarskap) och samarbete (teamorientering och ömsesidigt förtroende).

Kommunikation baserat på gester har föreslagits för kontroll av delvis autonoma robotar på slagfältet. Teaming mellan människa och maskin kommer att ha samma beroende som mänskliga team, inklusive samarbete, integration av förmågor samt kunskap och kompletterande ansvarsområden.

Denna rapport redovisar en teknisk implementation för registrering av gester med en motion capture-handske, som också påvisar ett flertal utmaningar i användandet av motion capture-handskar för interaktion med till exempel en UGV. Genom att anpassa användningen av handsken till dess begränsningar och använda gester som den har förmågan att konsekvent särskilja kan dock en del av begränsningarna hanteras.

I framtida studier behöver fler aspekter av gestbaserad interaktion utvärderas, i synnerhet i relation till den militära kontexten och i sammanhanget human – autonomy teaming, det vill säga hur gester kan användas inom ett team både för kommunikation och interaktion mellan människorna i teamet och mellan människorna och den autonoma enheten. Även hur gestbaserad interaktion kan och bör kombineras med andra former av gränssnitt för exempelvis återkoppling bör utredas vidare.

## Referenser

- Chandler, D. (2006). *Semiotics: the basics*. Routledge.
- Cicirelli, G., Attolico, C., Guaragnella, C., & D'Orazio, T. (2015). A kinect-based gesture recognition approach for a natural human robot interface. *International Journal of Advanced Robotic Systems*, 12(June). <https://doi.org/10.5772/59974>
- Försvarsmakten (2001). Signaler och tecken. I *SoldF Soldaten i fält*, M7742-100002. Stockholm: Försvarsmakten.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <https://www.deeplearningbook.org/>
- Graichen L., Graichen M., & Krems J. F. (2017b). Gestures in human-computer interaction: Which shape and how many? In Blashki K. (Ed.), *Proceedings of the International Conference Interfaces and Human Computer Interaction 2017* (part of the Multi Conference on Computer Science and Information Systems (MCCSIS) (pp. 69–76). Lisbon, Portugal, July 21–23, 2017. IADIS Press.
- Hansberger, J.T. *m.fl.* (2017). Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions. In: Lackey, S., Chen, J. (eds) *Virtual, Augmented and Mixed Reality. VAMR 2017. Lecture Notes in Computer Science()*, vol 10280. Springer, Cham. [https://doi.org/10.1007/978-3-319-57987-0\\_41](https://doi.org/10.1007/978-3-319-57987-0_41)
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, 15(3), 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Institutet för språk och folkminne (2021). *Den manuella delen: händerna*. <https://www.isof.se/stod-och-sprakrad/utbildningsmaterial/det-svenska-teckensprakets-grammatik/teckenbildning/lexikala-tecken/teckenuppbyggnad/handerna> Hämtat 2022-01-13
- ISO (2017). Size designation of clothes — Part 1: Anthropometric definitions for body measurement. (ISO standard 8559-1:2017).
- Levin, B., Nilsson, S., Hermelin, J., Svensson, E., Stenius, C. (2021). *Teknisk utveckling inom människa-maskininteraktion – Applikationer för framtida gränssnitt*. FOI-R--5245--SE.

- McNeill, D. (1995). *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press. <https://doi.org/10.2307/1576015>
- McNeill, D. (2005) *Gesture & Thought*. University of Chicago Press.
- Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 37(3), 311–324. <https://doi.org/10.1109/TSMCC.2007.893280>
- MoD (2018). *Joint Concept Note: Human Machine Teaming*. DCDC.
- Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, 12(1), 182–187. <https://doi.org/10.1111/j.1467-7687.2008.00764.x>
- Saito T., & Rehmsmeier M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* 10(3): e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Salazar, G., See, J. E., Handley, H. A. H., & Craft, R. (2020). Understanding Human Readiness Levels. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 64(1), 1765–1769. <https://doi.org/10.1177/1071181320641427>
- Trotsenko, E. (2018). *Gestures for the robotic platform UGV*. Besökt 2022-11-18. YouTube. <https://www.youtube.com/watch?v=E9ZcloH6Q7w>
- Van Den Bergh, M., Carton, D., De Nijs, R., Mitsou, N., Landsiedel, C., Kuehnlentz, K., Wollherr, D., Van Gool, L., & Buss, M. (2011). Real-time 3D hand gesture interaction with a robot for understanding directions from humans. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, May 2014*, 357–362. <https://doi.org/10.1109/ROMAN.2011.6005195>
- Voronchuk, O., Klym, H., & Dunets, R. (2019). Specialized Device to Control Work of Executive Mechanism Using Operator’s Hand Gestures. *2019 11th International Scientific and Practical Conference on Electronics and Information Technologies, ELIT 2019 - Proceedings*, 218–221. <https://doi.org/10.1109/ELIT.2019.8892324>
- Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., & Grealy, M. (2019). Systematic literature review of hand gestures used in human

computer interaction interfaces. *International Journal of Human Computer Studies*, 129, 74–94. <https://doi.org/10.1016/j.ijhcs.2019.03.011>

Zimmerman, T., Lanier, J., Blanchard, C., Bryson, S., & Harvill, Y. (1986). A hand gesture interface device. *ACM Sigchi Bulletin*, 17, 189-192. 10.1145/30851.275628

## Appendix A: Illustrationer av gester

Illustrationer av handformer från svenskt teckenspråk (Institutet för språk och folkminne, 2021).



4

4-handen



L

L-handen



O

O-handen



H

Stora  
nyphanden



A

A-handen



Δ

Lilla O-handen1



∟

Raka måtthanden



T

T-handen



∩

Dubbelkroken



I

Lillfingret



L

Pekfingret



B

Tumhanden



**J**

Flata handen



**R**

Långfingret



**N**

N-handen



**V**

V-handen



**†**

Flyghanden



**O**

Måtthanden



**♀**

Sprethanden



**U**

U-handen



**Σ**

Klohanden



**Z**

Runda  
långfingret



**F**

Stora  
hållhanden



**∇**

Tupphanden



**G**

Knutna handen



**Δ**

Nyphanden1



**Ψ**

Stora  
långfingret



**Λ**

Vinkelhanden



**∩**

Krokfingret

---

Bilderna publiceras med tillstånd av Svenskt teckenspråkslexikon vid avd. för teckenspråk på Stockholms universitet.

Illustrationer av militära handtecken (Försvarsmakten, 2001)



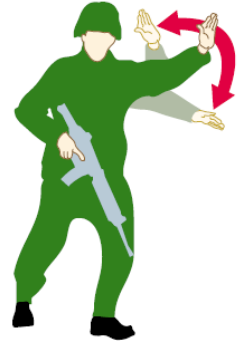
**Allt klart – Uppfattat**  
 Handen hålls i höjd med huvudet



**Samband sökes**



**Färdiga till strid**



**Framåt – Öka hastigheten**



**Runtomkringsförvar –  
 "Igelkottsförvar"**  
 En knuten hand mitt  
 uppe på huvudet



**Egen trupp (inget fientligt)**  
 Används vid målangivning.  
 Tummen upp



**Fientlig fottrupp**  
 Används vid målangivning.  
 Två fingrar riktas nedåt och  
 förs i en "gående" rörelse



**Fientligt pansarskyttefordon**  
 Används vid målangivning.  
 Knuten hand vrids fram och tillba-  
 ka upprepade gånger



**Mina i spåret**  
 Hand med utspärrade  
 fingrar mot ansiktet



**Plutonchef hitåt**  
 Handen på axeln



**Gruppchef hitåt**  
 Handen på överarmen

Bilderna publiceras med tillstånd av Försvarsmakten.



FOI är en huvudsakligen uppdragsfinansierad myndighet under Försvarsdepartementet. Kärnverksamheten är forskning, metod- och teknikutveckling till nytta för försvar och säkerhet. Organisationen har cirka 1000 anställda varav ungefär 800 är forskare. Detta gör organisationen till Sveriges största forskningsinstitut. FOI ger kunderna tillgång till ledande expertis inom ett stort antal tillämpningsområden såsom säkerhetspolitiska studier och analyser inom försvar och säkerhet, bedömning av olika typer av hot, system för ledning och hantering av kriser, skydd mot och hantering av farliga ämnen, IT-säkerhet och nya sensorers möjligheter.



FOI  
Totalförsvarets forskningsinstitut  
164 90 Stockholm

Tel: 08-55 50 30 00  
Fax: 08-55 50 31 00

[www.foi.se](http://www.foi.se)