

Goda utsikter för teckenspråksteknologi

Jonas Beskow och Björn Granström

KTH Centrum för Talteknologi

beskow@speech.kth.se, bjorn@speech.kth.se

Sammanfattning

I dag finns stora brister i tillgängligheten i samhället vad gäller teckentolkning. Nya tekniska landvinningar inom dator- och animationsteknologi, och det senaste decenniets forskning kring syntetisk teckentolkning har lett till att det nu finns nya förutsättningar att hitta tekniska lösningar med potential att förbättra tillgängligheten avsevärt för teckenspråkiga, för vissa typer av tjänster eller situationer. I Sverige finns idag ca 30 000 teckenspråksanvändare

Kunskapsläget har utvecklats mycket under senare år, både vad gäller förståelse/beskrivning av teckenspråk och tekniska förutsättningar för att analysera, lagra och generera teckenspråk. I kapitlet beskriver vi de olika tekniker som krävs för att utveckla teckenspråkteknologi.

Det senaste decenniet har forskningen kring teckenspråkteknologi tagit fart, och ett flertal internationella projekt har startat. Ännu har bara ett fåtal tillämpningar blivit allmänt tillgängliga. Vi ger exempel på både forskningsprojekt och tidiga tillämpningar, speciellt från Europa där utvecklingen varit mycket stark.

Utsikterna att starta en svensk utveckling inom området får anses goda. De kunskapsmässiga förutsättningarna är utmärkta; teknikkunnande inom språkteknologi, multimodal registrering och animering bl.a. vid KTH i kombination med fackkunskaper inom svenskt teckenspråk och teckenspråksanvändning vid Stockholms Universitet.

Inledning

I dag finns stora brister i tillgängligheten i samhället vad gäller teckentolkning. Nya tekniska landvinningar inom dator- och animationsteknologi, och det senaste decenniets forskning kring syntetisk teckentolkning har lett till att det nu finns nya förutsättningar att hitta tekniska lösningar med potential att förbättra tillgängligheten avsevärt för teckenspråkiga, för vissa typer av tjänster eller situationer. I Sverige finns idag ca 30 000 teckenspråksanvändare (Språkrådet 2010).

Talsyntes är ett exempel på framgångsrik tillgänglighetsförbättring med teknikens hjälp för bl.a. synskadade. Är tiden mogen för att göra något motsvarande för teckenspråkiga i Sverige?

Det är en allmän åsikt att tillgängligheten för teckenspråkiga är eftersatt i många sammanhang. Delvis beror det på att teckenspråket, ett av våra officiella språk, skiljer sig mycket från både skriven och talad svenska. Kunskapsläget, både vad gäller förståelse/beskrivning av teckenspråk och tekniska förutsättningar för att analysera, lagra och generera teckenspråk har utvecklats mycket under senare år. Internationellt pågår flera projekt med syfte att via automatiskt genererat teckenspråk förbättra tillgängligheten för teckenspråkiga.

I Sverige finns banbrytande forskningsverksamhet kring teckenspråk vid Stockholms Universitet. Vid KTH finns världsledande forskning om multimodal talsyntes och ansiktsanimering i tillgänglighetsskapande kommunikationshjälpmedel. Tekniken för analys och generering har utvecklats t ex med hjälp av motion capture-teknik och animation, som beskrivs under rubriken *Teknik och infrastruktur för teckenspråk* nedan. Uppenbart är området mycket komplext. Både användarbehov och tekniska möjligheter måste klaras ut innan ett storskaligt projekt kan startas.

I det följande ges en översikt över forskningen på området och en beskrivning av den bakomliggande tekniken. Slutligen diskuteras utsikterna att utveckla teckenteknologi i Sverige.

Internationella forskningsprojekt och system

Det senaste decenniet har forskningen kring teckenspråkteknologi tagit fart, och ett flertal internationella projekt har startat. 1999–2002 genomfördes det första EU-finansierade teckenspråkteknologiska projektet, VisiCast (Bangham et al., 2000), och till dags dato har EUs forskningsramprogram finansierat ytterligare tre större projekt, som beskrivs nedan : E-sign, DictaSign och SignSpeak. Sedan 2004 anordnas en biennial workshop kring teckenspråkteknologi och teckenspråkskorpora i samband med LREC (Language Resources and Evaluation Conference). År 2010 på Malta hade workshopen över 100 deltagare, 55 bidrag och representerade över 20 nationella och internationella forskningsprojekt. Nedan följer en genomgång av några pågående och avslutade projekt.

Visicast

Målet med projektet var att utveckla en tecknande avatar som kunde användas för ökad tillgänglighet. Den tillämpning man utvecklade i projektet var en teckentolk för postkontor. Detta system, som testades på ett postkontor i England, kunde översätta ett begränsat antal standardfraser från tal till animerat teckenspråk. Systemet byggde på motion capture av hela fraser som spelas upp med en 3D-avatar, se figur 6. Fraserna kunde parametreras med vissa variabler, som penningbelopp. Avataren och tillhörande teknologi som utvecklades inom projektet har använts i ett stort antal senare projekt.

E-sign

I mångt och mycket en efterföljare till Visicast löpte E-sign som ett EU-projekt mellan 2002–2005 (Zwierslood et al., 2004). Målet var att vidareutveckla avartekniken och anpassa den till webben och informationskiosk-tillämpningar, och att relativt snabbt kunna generera innehåll, t ex från webbsidor. Ett verktyg utvecklades för att skapa animeringar från text. Motion-Capture-tekniken från Visicast övergavs eftersom den inte tillät något enkelt sätt att generera nytt innehåll utan tillgång till Mocap-studio. Istället implementerades en parametrisk syntesstyrning som är baserad på HamNoSys-transkriptioner, se figur 7.

DictaSign

Ett av de pågående EU-Projekten inom området heter DictaSign (Efthimiou et al., 2009). Projektet fokuserar på tillgänglighet över internet, speciellt för olika webbaserade tjänster såsom bloggar, chatrum och sociala nätverk. Projektet har som målsättning att ta fram teknik för både tecken-till-text och text-till-tecken. Tecken-till-text, dvs diktering, är avsett att låta användaren göra kommentarer eller chatta i textbaserade medier genom att teckna framför en vanlig webbkamera, vilket systemet sedan försöker tolka till text, som användaren sedan kan välja att redigera och posta. Text-till-tecken skulle omvänt låta användaren tillgodogöra sig texter på nätet med hjälp av en tecknande avatar.

DictaSign använder sig av en vidareutveckling av avataren från Visicast och E-sign. Både UEA, som utvecklat avataren, och universitetet i Hamburg, som skapat transkriptionssystemet HamNoSys, är parter i projektet.

SignSpeak

SignSpeak är det andra pågående EU-projektet inom teckenspråkteknologi, och det syftar till att utveckla automatisk igenkänning, förståelse och översättning av teckenspråk (Dreuw et al., 2010). Fokus ligger på utveckling av metoder för igenkänning av kontinuerligt teckenspråk baserade på datorseende, inklusive robusta metoder för särdragsextraktion, samt utveckling av statistisk maskinöversättning för översättning mellan skrivet språk och teckenspråk.

ATLAS

ATLAS är ett italienskt projekt som bl.a. involverar universitetet i Turin, RAI och Microsoft. Projektet syftar till att utveckla teknologi för översättning från skriven italienska till italienskt teckenspråk (Vendrame & Tiotto, 2010). Inom ATLAS-Projektet arbetar man bl.a. med att ta fram ett system för att översätta väderprognoser till teckenspråk, se figur 1. Denna domän valdes som en realistisk tillämpning av ett översättningssystem, och utgör samtidigt ett intressant exempel på kombination av teckenspråk och spatials referenser till väderkartan.



Figur 1: Avatar som teckentolkar väderprognoser i TV i det italienska ATLAS-projektet.

SNCF Informationssystem

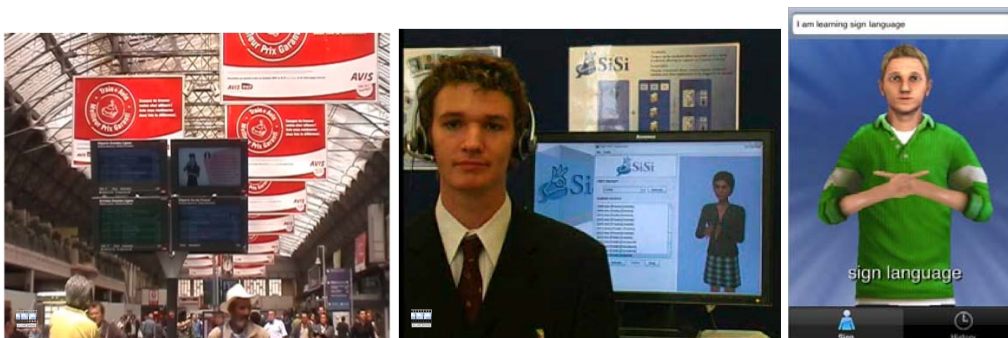
Som ett exempel på en existerande tillämpning kan nämnas SNCFs satsning på trafikinformation på teckenspråk. Detta system använder animerade tecknande avатарer på skärmar och är installerat på 60 järnvägsstationer i Frankrike (figur 2, vänster). Dock är avatarna animerade ”för hand”, dvs man använder sig av samma sk key-frame teknik som används vid produktion av animerad film för att skapa rörelser efter mänsklig förlaga. Systemet skulle alltså lika gärna kunna använda verkliga videobilder, och valet att använda avатарer är bara estetiskt. Dock är det ett intressant exempel på en tillämpning där det tycks finnas ett stort behov som skulle kunna vidareutvecklas med hjälp av teckenspråkteknologi, och som kommit upp som förslag i diskussioner med intressenter i Sverige.

IBM SiSi

SiSi (Say it sign it) var ett studentprojekt vid en sommarskola hos IBM i England 2007. Projektet gick ut på att utnyttja IBMs teknik för taligenkänning, i kombination med den avatartechnologin från UEA, som utvecklats i projekten Visicast och E-sign. Det kombinerade systemet kunde översätta engelska fraser till animation med British Sign Language (BSL). IBM gick ut med en pressrelease men projektet har i övrigt inte tillgängliggjorts officiellt eller kommersialiserats. (Se figur 2, mitten).

Sign4me

En av få existerande kommersiella applikationer av teckenspråkteknologi som finns på marknaden kommer från det Florida-baserade företaget Vcom3D som är specialiserat på utbildnings- och träningsystem med avатарer. De säljer programmet Sign4me för Apple Iphone (Vcom3D, 2010), en tillämpning som marknadsförs som ett träningsprogram för American Sign Language (ASL), och som låter användaren skriva in valfri text och få den uppspelad som en animation av en 3D-avатар (Figur 2, höger). Det är möjligt att se avataren ur olika vinklar och zooma in och ut som man vill. Ord som inte finns i programmets lexikon bokstaveras med handalfabet.



Figur 2: Tre exempel på teckenspråkteknologi i produkter eller företagsprototyper. Från vänster: Tecknande avатар på tågstation i Frankrike; Say-it-sign-it, prototyp framtagen genom ett studentprojekt vid IBM; Sign4me – teckenspråksapplikation för iPhone (Vcom3D).

Teknik och infrastruktur för teckenspråk

I detta avsnitt sammanfattas status för ett antal teknologier som kan komma att ingå i, eller användas vid utvecklingen av ett system för teckenspråkteknologi. Detta gäller teknik för registrering av hand- och ansiktsrörelser i 3D, teknik för analys av teckenspråk från videobilder samt teknik för generering av teckenspråk från olika representationer med hjälp av avатарer. Dessutom avhandlas system för transkribering av teckenspråk.

Igenkänning av teckenspråk

I många tillämpningar av teckenspråkteknologi krävs att en maskin kan känna igen tecken utförda av en verklig person, baserat på analys av videosekvenser.

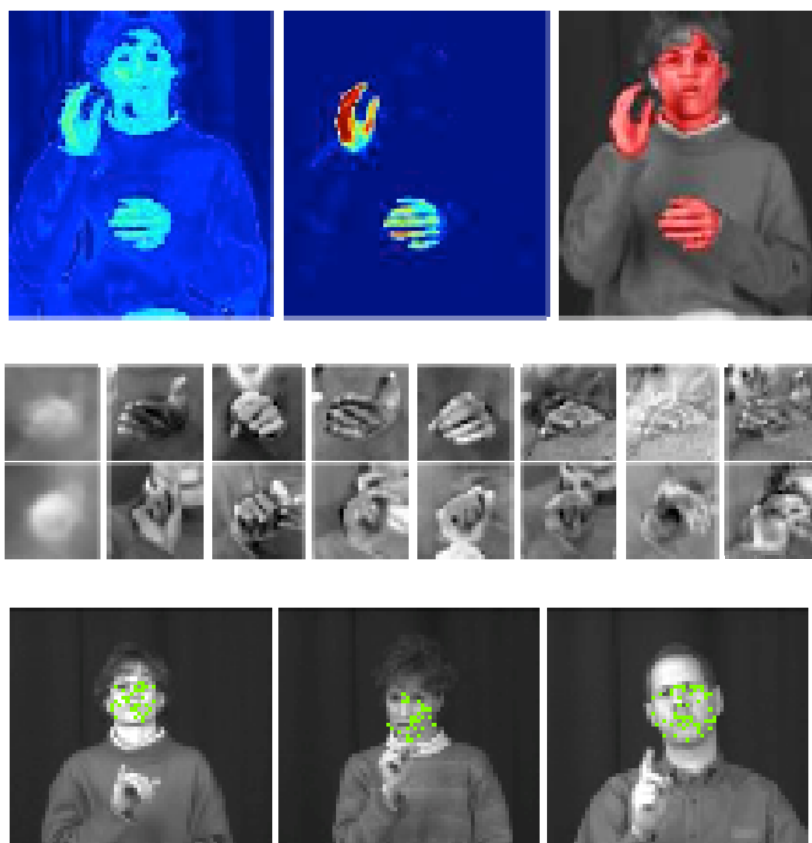
Det första steget i ett sådant system är videoanalys baserad på algoritmer för datorseende. Bilden måste segmenteras och händer, ansikte och kropp måste identifieras och analyseras i bilden (se figur 3). För händer finns olika tekniker som baseras på klassificering i en av ett antal prototypiska handformer, alternativt återskapande av handformen utifrån fingerledernas individuella lägen. Ett exempel på angreppssätt ges av Piater et al. (2010), som utnyttjar s k principalkomponentanalys (PCA) för att koda handformer. PCA är en statistisk metod att beskriva samvariation och reducera informationsmängden i en stor mängd datapunkter, i detta fall videopixlar. För ansiktet måste typiskt huvudpose, käke, läppar och ögonbryn följas för att ge komplett information. Det finns mycket forskning kring ansiktsanalys, och även kommersiella system, som kan extrahera mun-, ögonbryn- och huvudrörelser ur en videoström. Hur väl dessa tekniker fungerar är dock starkt beroende på en mängd faktorer såsom ljusförhållanden, bakgrund, bildfrekvens mm. Ocklusioner, t ex att en hand skymmer munnen, utgör ett vanligt problem och kan leda till att ansiktsföljningen ”spårar ur”.

Steget efter bildbehandlingen är typiskt att tolka rörelserna i termer av enskilda tecken. Paralleller kan dras till problemet med automatisk taligenkänning: i båda fallen handlar det om att omvandla en kontinuerlig ström (ljud/video) till betydelsebärande symboler (ord/glosor). I båda fallen finns problem med koartikulation, dvs att realiseringen är beroende av kontexten, vilket försvårar tolkningen. Detta anses dock vara mer omfattande i teckenspråket, eftersom samma tecken kan utföras i olika spatiala lägen. En annan viktig skillnad mellan tal och teckenspråk är att teckenspråket kan innehålla många parallella skeenden, uttryckta med händer, ansikte respektive kroppspose medan tal och text är sekventiellt till sin natur. Slutligen kan taligenkänning ofta dra nytta av tystnad för att segmentera yttranden, något som inte är möjligt i teckenspråket.

De flesta system som presenterats har fokuserat på igenkänning av enskilda glosor men några har även angripit problemet med kontinuerlig teckenspråksigenkänning. De flesta systemen bygger på maskininlärningsmetoder som behöver tränas på mänskligt annoterade träningsdata, och annotering är en tidskrävande process. Buehler et al. (2010) beskriver en metod som bygger på att utnyttja TV-utsändningar som är både textade och teckentolkade, och låta ett självlärande system bygga upp sin kunskap automatiskt från detta utan mänsklig annotering utöver den vanliga textningen.

Registrering

De bildbehandlingstekniker som beskrivs ovan kan även vara användbara för korpusarbete, men för inspelningar som görs i studio finns möjlighet att använda registreringsteknik för att få mera tillförlitliga data. Optisk motion capture är en teknik som bygger på att reflekterande markörer fästs på det objekt man vill mäta. Ett system med IR-känsliga kameror kan sedan följa dessa reflexer ur olika vinklar och därmed beräkna exakta 3D-positioner för markörerna (se figur 4). Denna teknik har använts flitigt för inspelning av både ansikts- och kroppsrorelser bl.a. i spel- och filmsammanhang och även inom forskningen, bl.a. på KTH inom projekten SynFace (Beskow et al., 2003) med fokus på ansiktet, och Spontal (Edlund & Beskow, 2010) med fokus på gestik i dialog.

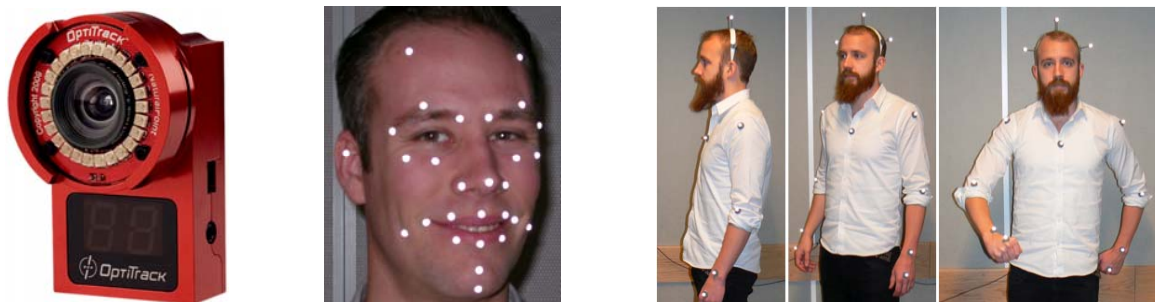


Figur 3: Bildanalys för teckenspråksigenkänning. Överst: Segmentering av bilden; mitten: analys av handformer; nederst: ansiktsanalys. Ur Piater et al. (2010).

Optiska motion capture system har använts för registrering av teckenspråk, dock behöver dessa ofta kompletteras med annan teknik eftersom t ex slutna handformer är svåra att fånga på grund av oklusioner, dvs att markörerna skymms av handen. Så kallade datahandskar (figur 5) kan vara en lösning på detta problem. Ett alternativ till optiska system är motion capture baserat på accelerometrar och/eller elektromagnetiska sensorer. Motion capture-teknik kan heller inte registrera ögonrörelser.

Ett system som kan fånga allt som är kommunikativt relevant för teckenspråk kräver därför en kombination av flera registreringstekniker. Huenerfauth & Lu (2010) beskriver t ex en konfiguration för teckenspråksinspelning som består av fyra delar: datahandskar, en huvudmonterad eye-tracker, en huvudtracker samt en accelerometer/gyroskop-baserad motion capture-dräkt.

Bildbehandlingsteknik kan också göras betydligt mer robust om man har möjlighet att hjälpa analysen på traven. Wang & Popovic (2009) beskriver en teknik med färgmönstrade latexhandskar (se figur 5, höger) som gör det möjligt att följa handrörelser med god noggrannhet i 3D med hjälp av en vanlig webbkamera. Ansiktsanalys och eye-tracking går att göra med hög noggrannhet från videobilder om man har kontroll över inspelningssituationen. Genom att använda en huvudmonterad kamera erhåller man en betydligt mer stabil och lättanalyserad bild. Denna teknik har använts i system för automatisk läppavläsning och utnyttjades även (i kombination med optisk motion capture) under inspelningen av filmen Avatar.



Figur 4: Med motion capture –teknik kan man registrera rörelser i 3D med hög precision. Till vänster: Mocap-kamera; mitten: markörplacering för ansiktsregistrering; till höger: markörplacering för registrering av överkropp/huvudrörelser.



Figur 5: Vänster: Datahandske (VR realities Cyber Glove II); höger: mönstrad latexhandske för optisk hand-tracking (Wang & Popovic, 2009).

Syntes av teckenspråk

Syntes av teckenspråk har varit ett växande forskningsområde under det senaste decenniet. Huvudsakligen finns två angreppssätt, som har sin motsvarighet i akustisk talsyntes: konkatenering av förinspelade enheter, samt parametrisk syntes. I det första fallet kan enheterna antingen bestå av videoklipp av en verklig tecknare, eller animerade sekvenser med en virtuell avatar. Denna teknik kräver en uppsättning förinspelade enheter, som spelats in med video eller motion capture-teknik enligt ovan. Förutsatt att rätt enheter finns tillgängliga så leder denna teknik typiskt till den bästa kvaliteten. Dock leder teckenspråkets höga komplexitet snabbt till ett väldigt stort antal nödvändiga enheter, vilket gör detta till en metod som lämpar sig bäst för små och väl avgränsade domäner.

Större flexibilitet erhålls genom parametrisk syntes. Denna teknik bygger på abstrakta beskrivningar av de enskilda rörelser som bygger upp tecknen. Syntesapparaten använder dessa definitioner och kombinerar dem till godtyckliga tecken utifrån en detaljerad transkription, som animeras med en avatar. Denna typ av syntes är robust och flexibel och gör det möjligt att hantera övergångar och koartikulationseffekter på ett bättre sätt, dock inte med samma naturlighet i rörelserna som motion-capture baserade system.

Exempel på några avatarer för teckenspråk kan ses i figur 6.

De två mest använda verktygen för teckenspråkstranskription är ELAN, utvecklat vid Max Planck Institute, Nederländerna och ILEX, utvecklat vid universitetet i Hamburg. I praktiken sker transkriptionen enligt ett tvåstegsförfarande: först skriver man in de ingående glosorna, varpå systemet plockar fram kanoniska HamNoSys-transkriptioner för dessa. Nästa steg är att justera den kanoniska transkriptionen så att den stämmer med den specifika realiseringen. Här behöver man alltså ta hänsyn till reduktioner, koartikulationseffekter och övergångar mellan de ingående kanoniska glosorna. Skulle en glosa saknas, eller skulle det inte finnas något lexikon alls för det aktuella språket, blir det en väldigt tidsödande process.

Exempeltranskriptionen till höger i figur 7 beskriver en mening på svenskt teckenspråk, för vilket det inte finns något HamNoSys-transkriberat lexikon. Den tog drygt 2 timmar att färdigställa för en experttranskriberare.

Sammanfattande diskussion

Det är uppenbart att det finns en stor potential i teckenspråkteknologiska tillämpningar, men också stora svårigheter. Det står även klart att området är på stark frammarsch, vilket det stora antalet projekt och bidrag på internationella konferenser vittnar om. För att möjliggöra teckenspråkstillämpningar på svenskt teckenspråk finns ett stort behov av en språkteknologisk infrastruktur. Detta avser främst korpora och lexika som omfattar transkriptioner på tillräcklig detaljnivå för att möjliggöra animeringar, exempelvis HamNoSys-transkriptioner, alternativt inspelningar med kompletta motion capture-system för teckenspråk. Därför är det troligt att ett första projekt inom teckenspråkteknologi för svenskt teckenspråk skulle innefatta en satsning på att bygga upp erforderlig infrastruktur.

En generell svårighet är att definiera tillämpningar med otvetydig nytta för teckenspråksanvändaren, speciellt om det rör sig om att tillhandahålla information som redan finns tillgänglig i skrift, t ex på webben.

En tänkbar tillämpning som inte utgår från skriven information har att göra med anonymisering av tecknaren – något som kom fram som ett behov flera gånger under förstudien. För denna typ av system ligger kraven snarare på tekniken än på infrastrukturen, vilket har fördelen att det är lättare att utnyttja icke-språksspecifika komponenter, eventuellt i samarbete med andra projekt inom EU eller övriga världen. En möjlig tjänst skulle t ex kunna översätta video från en webbkamera till teckenspråksanimering med en avatar. Dock är det svårt att bedöma omfattningen av den tekniska utveckling som krävs innan tekniken är tillräckligt robust för att möjliggöra en sådan tjänst.

Sammanfattningsvis kan dock sägas att utsikterna att starta och genomdriva ett implementeringsprojekt inom området i Sverige får anses goda om finansiering för ett sådant projekt går att uppbringa. De kunskapsmässiga förutsättningarna är utmärkta; teknikkunnande inom språkteknologi, multimodal registrering och animering bl.a. vid KTH i kombination med fackkunskaper inom svenskt teckenspråk och teckenspråksanvändning vid Stockholms Universitet.

Referenser

- Bangham, J.A., Cox, S. Elliott, R., Glauert, J.R.W., Marshall, I., Rankov, S., Wells, M. (2000). Virtual Signing: Capture, Animation, Storage and Transmission – an Overview of the ViSiCAST Project. *Proceedings of the fourth international ACM conference on Assistive technologies*. ACM.
- Beskow, J., Engwall, O., & Granström, B. (2003). Resynthesis of Facial and Intraoral Articulation from Simultaneous Measurements. In Solé, M., Recasens, D., & Romero, J. (Eds.), *Proceedings of the 15th ICPhS* (pp. 431–434). Barcelona, Spain.

- Buehler, P., Everingham, M., Zisserman, A. (2010). Employing signed TV broadcasts for automated learning of British Sign Language. In *Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valetta, Malta, May 22-23, 2010.
- Dreuw, P., Forster, J., Gweth, Y., Stein, D., Ney, H., Martinez, G., Llahi, J.V., Crasborn, O., Ormel, E., Du, W., Hoyoux, T., Piater, J., Miguel Moya, J.M., and Wheatley, M. (2010): SignSpeak – Understanding, Recognition, and Translation of Sign Languages. In *Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valetta, Malta, May 22–23, 2010.
- Edlund, J., & Beskow, J. (2010). Capturing massively multimodal dialogues: affordable synchronization and visualization. In Kipp, M., Martin, J-C., Paggio, P., & Heylen, D. (Eds.), *Proc. of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (MMC 2010) (pp. 160–161). Valetta, Malta.
- Efthimiou, E., Fotinea, S., Vogler, C., Hanke, T., Glauert, J., Bowden, R., Braffort, A., Collet, C., Maragos, P. and Jérémie Segouat, J. (2009). Sign Language Recognition, Generation, and Modelling: A Research Effort with Applications in Deaf Communication. *Proceedings of UAHCI 2009/HCI 2009*, Springer Lecture Notes in Computer Science 5614, pp 21–30, 2009.
- Huenerfauth, M. (2009). A Linguistically Motivated Model for Speed and Pausing in Animations of American Sign Language. *ACM Transactions on Accessible Computing*, 2, pp. 1–31.
- Piater, J., Hoyoux, T., Du, W. (2010). Video Analysis for Continuous Sign Language Recognition. In *Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valetta, Malta, May 22–23, 2010.
- Schmaling, C. & Hanke, T. (2004). *HamNoSys 4.0*. <http://www.sign-lang.uni-hamburg.de/projekte/hamnosys/hns4.0/englisch/hns4.pdf>
- Schnepf, J, Wolfe, R. & McDonald, J. (2010). Synthetic Corpora: A Synergy of Linguistics and Computer Animation. In proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Valetta, Malta, May 22-23, 2010.
- Språkrådet (2010). *Om Svenskt Teckenspråk*. http://www.sprakradet.se/servlet/GetDoc?meta_id=2345.
- Vendrame, M & Tiotto, G. (2010). ATLAS Project: Forecast in Italian Sign Language and Annotation of Corpora. In *Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valetta, Malta, May 22-23, 2010.
- Wang, R., Popović, J. (2009). Real-Time Hand-Tracking with a Color Glove. *ACM Trans. Graph.* 28, 3, Article 63 (August 2009).
- Zwitterslood, I., Verlinden, M., Ros, J., van der Schoot, S. (2004). Synthetic Signing for the Deaf: eSIGN. In *Proceedings of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment*, CVHI 2004, 29 June–2 July 2004, Granada, Spain.

Slutnot

Hjälpmedelsinstitutet tog under våren 2009 initiativ till två möten, för en diskussion av förutsättningar och behov av teckenspråkteknologi. Deltagare var representanter för brukare och forskare, samt organisationer med ansvar för tillgänglighet. Mötena ledde fram till en förstudie som samfinansierats av Hjälpmedelsinstitutet och Kommunikationsmyndigheten PTS, och genomfördes vid KTH, Centrum för Talteknologi. Denna artikel är baserad på förstudien. En rapport med en fullständig referenslista finns tillgänglig på: www.speech.kth.se/publications/files/3497.pdf.