

Prof. Dr. Rolf Schulmeister
Universität Hamburg
Institut für Deutsche Gebärdensprache (IDGS)
email: schulmeister@uni-hamburg.de

ViSiCAST: Virtual Signing, Capture, Animation, Storage and Transmission – Übersetzung in und Generierung von virtueller Gebärdensprache im Fernsehen und Internet

Das Projekt ViSiCAST (<http://www.visicast.co.uk>) wird im Rahmen des 5th Framework-Projekts im Information Societies Technology (IST)-Programm der Europäischen Kommission seit dem 1.1.2000 für zunächst drei Jahre gefördert. Unsere Partner im Projekt sind die Independent Television Commission (ITC) in Winchester, England, das Institut für Rundfunktechnik in München, die Firma Televirtual in Norwich, England, die University of East Anglia (UEA, School of Information Systems) in Norwich, das Institut National des Télécommunications (INT) in Evry, Frankreich, das Instituut voor Doven (IvD) in Sint-Michielsgestel, Niederlande, das Post Office in England und das Royal National Institute for Deaf People (RNID) in England.

• Ziel des Projekts ViSiCAST

Ziel des Projekts ViSiCAST ist die Entwicklung realistisch aussehender virtueller Menschen (Avatare), die aus gesprochener Sprache oder aus Texten Gebärdensprache generieren und auf dem Fernseh- oder dem Computer-Bildschirm darstellen sollen. Mit dieser Zielsetzung wollen wir einen Beitrag zur Verbesserung der Kommunikation zwischen Gehörlosen und Hörenden und der Teilhabe Gehörloser am sozialen und politischen Geschehen der Gesellschaft leisten.

• Motivation des Projekts ViSiCAST

Warum greifen wir diese Thematik als Projekt auf? Bevor ich das Projekt im Detail beschreibe, möchte ich auf diese Frage näher eingehen. Die Antwort auf diese Frage hat mehrere Aspekte:

1. Die kommunikative Situation der Gehörlosen und die Gebärdensprache
2. Die gesellschaftspolitische Behinderung der Gehörlosen in der Informationsgesellschaft
3. Die technologischen Fortschritte bei der Erkennung, Generierung, Übertragung und Übersetzung von Gebärdensprache.

2.1 Die kommunikative Situation der Gehörlosen und die Gebärdensprache

Gehörlose haben mit der Gebärdensprache ein eigenes faszinierendes Kommunikationsmittel entwickelt. Wenn sie in der Gebärdensprache kommunizieren, sind sie nicht behindert. Nur mit der Lautsprache und der Schriftsprache der Hörenden haben sie Schwierigkeiten. Ich werde häufig erstaunt gefragt, wieso denn die Gehörlosen nicht so gut lesen und schreiben können wie wir Hörenden. In der Tat ist die Lese- und Schreibfähigkeit der Gehörlosen trotz intensiven Trainings im Elternhaus und in der Schule nur auf einem geringen Niveau ausgebildet. Der Grund dafür ist einfach die fehlende auditive Rückmeldung: Da die Gehörlosen die Lautsprache ihres Gesprächspartners nicht direkt empfangen und zu ihrer eigenen Lautsprachproduktion nie eine direkte Rückmeldung erhalten können, fehlt ein natürlicher Weg, der zu einer besseren Beherrschung dieser Sprache führen würde. Die Laut- und Schriftsprache wird als unnatürliche Kunstsprache gelernt, so als würden wir heute noch versuchen Latein als Verkehrssprache zu erlernen, obwohl es keiner mehr spricht.

Im Vergleich dazu ist die Gebärdensprache ein lebendiges Kommunikationsmittel, das über alle linguistischen Repertoires einer natürlichen Sprache verfügt. Nur in dieser Sprache ist den Gehörlosen eine volle Teilhabe am gesellschaftlichen Geschehen möglich, was zur Konsequenz hat, daß sie in der Interaktion mit der hörenden Welt nur dann voll partizipieren können, wenn sie sich auf eine Übersetzung durch Dolmetscher verlassen können.

2.2 Die gesellschaftspolitische Behinderung der Gehörlosen in der Informationsgesellschaft

Die Teilhabe der Gehörlosen an der entweder auditiv und per Schriftsprache vermittelten Informationswelt ist aus den genannten Gründen stark restringiert. Nachrichten und andere Informationssendungen in Radio und Fernsehen, aber auch in Zeitungen und Zeitschriften ermöglichen Gehörlosen keine echte Teilhabe am politischen und gesellschaftlichen Geschehen. Seit einem Beschluß des Europäischen Parlaments wurde in einigen Ländern die jeweilige nationale Gebärdensprache als Verkehrssprache anerkannt, zuletzt in Griechenland.

Das Fernsehen, insbesondere die öffentlich-rechtlichen Anstalten, haben sich lange Zeit gegen eine Übersetzung in Gebärdensprache gesperrt und stattdessen auf ihre Anstrengungen bei der Untertitelung von Sendungen verwiesen. Die Untertitel sind jedoch aus den genannten Gründen für Gehörlose wenig geeignet, weil sie die Information reduzieren und dennoch die Lesefähigkeit vieler Gehörloser überfordern. Seit einiger Zeit werden wenigstens auf Phoenix die Hauptnachrichten von ARD und ZDF gedolmetscht. Dies scheint aber als Entschuldigung dafür zu dienen, ansonsten keine weiteren Dolmetscherleistungen anbieten zu müssen.

Insgesamt ist der Prozentsatz der Dolmetschereinblendungen ausgesprochen gering. In Großbritannien hat das Parlament die terrestrischen Fernsehanstalten aufgefordert, den Prozentsatz der Dolmetschereinblendungen zu erhöhen. Dies scheint bei den Fernsehanstalten auf wenig Gegenliebe zu stoßen, da die Übertragung der Einblendung zu einer Erhöhung der zu übertra-

genden Datenmenge führt, die ansonsten profitabel verkauft werden kann. Will man also den Prozentsatz der Übersetzung im Fernsehen erhöhen, so ist es wichtig, nach einer Lösung zu suchen, die weniger Übertragungskapazität beansprucht. Dann wäre es denkbar, daß das Fernsehen den Anteil an Übersetzungen in Gebärdensprache erhöht.

In ähnlicher Weise errichtet die aktuelle Entwicklungsstufe des Internet eine zusätzliche Barriere gegen die Teilhabe der Gehörlosen an der Informationsgesellschaft. Da die Welt des Internet zur Zeit überwiegend textbasiert ist, fühlen sich viele Gehörlose trotz großen Interesses an den Neuen Medien von dieser Welt ausgeschlossen. Darstellungen von Web-Inhalten in Gebärdensprache sind bis heute selten, nur in Fachkreisen vorhanden, und wenn es mehr Netzangebote in Gebärdensprache gäbe, dann würde die Kapazität und Geschwindigkeit der heutigen Datenleitungen nicht ausreichen, um ein halbwegs vernünftiges Arbeiten und Kommunizieren in Gebärdensprache zu ermöglichen.

2.3 Fortschritte bei der Generierung, Übertragung und Übersetzung von Gebärdensprache

Das Bildtelefon hat heute einen technischen Stand erreicht, der eine annehmbare Verständigung in Gebärdensprache zu einem vernünftigen Preis erlaubt. Aber leider ist das Bildtelefon nicht zugleich für die Kommunikation zwischen Hörenden und Gehörlosen geeignet. Um dies zu ermöglichen, müßte entweder ein ständig erreichbarer Dolmetscher-Relais-Service zur Verfügung stehen oder es müßte zwischen Sender und Empfänger eine automatische maschinelle Übersetzung von Gebärden- in Lautsprache und umgekehrt dazwischen geschaltet werden. Derartige Ideen haben wir bereits seit einem Jahrzehnt verfolgt [Schulmeister 1992; Schulmeister 1994a; Schulmeister 1994b], aber erst jetzt mit dem Projekt Zu diesem Problem soll das Projekt ViSiCAST einen ersten Beitrag leisten, indem eine Übersetzung zwischen Lautsprache und Gebärdensprache entwickelt wird.

Die Technik der an digitale Fernseher angeschlossenen Settop-Boxen ermöglicht es, Signale zu mischen, die getrennt ankommen, und so dem individuellen Zuschauer die Wahl zu geben, ob er/sie eine Übersetzung einblenden möchte oder nicht. Auch zu diesem Problem soll das Projekt einen Beitrag leisten: ViSiCAST will eine Übertragung der Daten für die Gebärdensprache in niedriger Bandbreite ermöglichen und eine Settop-Box entwickeln, die es dem Empfänger ermöglicht, einen Avatar zum Fernsehbild zuzuschalten, der beispielsweise die Untertitel in Gebärdensprache darstellt. Die Übertragung der Gebärden erfolgt kompatibel zu den für Gesichts- und Körperanimationen und dreidimensionale Körper definierten Standards in MPEG-4 und MPEG-7.

Die Plug-In-Technologie der Web-Browser ermöglicht es dem Computer-Nutzer, zusätzliche Funktionen aufzurufen, z.B. ein Übersetzungsprogramm, das ausgewählten Text in andere Sprachen übersetzt. Das Projekt ViSiCAST wird ein Plug-In entwickeln, das einen Avatar aufruft, der den ausgewählten Text in Gebärdensprache übersetzt und auf dem Bildschirm darstellt.

Für die Entwicklung des Avatars kann auf die enormen technologischen Fortschritte der Animationstechnik von dreidimensionalen menschlichen Figuren zurückgegriffen werden. Das Projekt wird Avatare entwickeln, die mittels VR-Technik und Laserscan-Technik gewonnen werden, natürlich aussehen und sich im Sinne der Gebärdensprache realistisch bewegen.

Weitere Fortschritte der Forschung auf den Gebieten der Spracherkennung, der Bildererkennung, hier der Gebärdenerkennung und der Linguistik können in dem Projekt fruchtbar genutzt werden, um eine Übersetzung von Lautsprache und Schriftsprache in Gebärdensprache zu entwickeln. Die Gebärdensprache wird mit HamNoSys notiert und in GML (Gesture Markup Language) an den Avatar übertragen. Die Übersetzung von Text in natürliche Gebärdensprache stößt auf dieselben Probleme wie die maschinelle Übersetzung von einer Lautsprache in eine andere, obwohl die linguistischen Merkmale der Gebärdensprachen andere sind als die der Lautsprachen.

• Die Idee des Projekts ViSiCAST

Um gleich auf einen Vorbehalt einzugehen, der bei der Darstellung dieses Projektes in der Öffentlichkeit entstehen könnte: Es ist nicht das Ziel von ViSiCAST, die Leistungen menschliche Dolmetscher zu substituieren, sozusagen die Dolmetscher brotlos zu machen. Das wäre eine unrealistische Zielsetzung und vor allem eine unsoziale.

Aber es ist das Ziel von ViSiCAST, Übersetzungsleistungen **dort und dann** zur Verfügung zu stellen, wo sie unabhängig von Ort und Zeit benötigt werden und deshalb nicht von menschlichen Dolmetschern erbracht werden können, z.B.

- im Fernsehen bei Warn- und Notmeldungen, wenn ein Dolmetscher nicht schnell genug zur Stelle sein kann,
- im World Wide Web im Internet, in dem nicht vorhergesehen werden kann, welche Information ein Gehörloser anfordert und in Gebärdensprache übersetzt haben möchte,
- in Face-to-Face-Transaktionen, die nur gelegentlich und unregelmäßig auftreten und zeitlich zu kurz sind, um dafür dauerhaft einen Dolmetscher vorhalten zu können, z.B. an den Schaltern der Post oder der Banken oder öffentliche Einrichtungen, an denen nur selten Gehörlose auftreten.

Diese technischen und technologischen Fortschritte ermöglichen es heute, die Problematik der Entwicklung eines Übersetzungssystems mit künstlichen Menschen in Angriff zu nehmen. Anwendungen des Systems im Fernsehen, im Internet, in Multimedia-Lernprogrammen und in Transaktionen zwischen Hörenden und Gehörlosen können dazu beitragen, die Teilhabe der Gehörlosen an sozialer Kommunikation, an Information und an öffentlichen und kommerziellen Diensten der hörenden Gesellschaft zu verbessern.

4. Beschreibung des ViSiCAST-Projekts

Kern des ViSiCAST-Projekts ist die Entwicklung eines maschinen-lesbaren Systems zur Beschreibung der Gebärdensprache. Dieses System setzt auf der HamNoSys-Notation für Gebärdensprachen auf, die vom Institut für Deutsche Gebärdensprache entwickelt wurde [Prillwitz et al 1989; Prillwitz & Zienert 1990] und die im Projekt um einige Aspekte (Mimik, Syntax) ergänzt und erweitert werden soll. ViSiCAST nutzt diese deskriptive Beschreibungssprache, um darauf Übersetzungswerkzeuge für die Übersetzung von Lautsprache und Text aufzusetzen.

Das ViSiCAST-System soll

1. Gebärdensprache mit niedriger Bandbreite im Fernsehen übertragen können,
2. In Echtzeit Gebärdensprache generieren können
3. Mit Standards wie MPEG/DVB und XML kompatibel sein.

Gebärdensprachen zeigen besondere morphologische, phonologische und syntaktische Merkmale. Phonologische Komponenten der Gebärdensprache sind beispielsweise die Handform, die Handstellung, die Ausführungsstelle im dreidimensionalen Raum vor dem Körper oder am Körper und schließlich die Bewegung selbst. Gebärden, Mimik und Körperbewegung werden synchron ausgeführt, bedeutungsvolle grammatische Funktionen (z.B. Verb plus Pronomen, Subjekt und Verb, Zahlen und Nomen) werden häufig in einer einzigen Gebärde inkorporiert und nicht, wie in der Lautsprache, sequentiell ausgeführt.

ViSiCAST strebt in der erste Phase des Forschungsprojekts eine interaktive semi-automatische Übersetzung in zwei Stufen an: Der zu übersetzende Text wird in Form einer Diskurs-Repräsentations-Struktur (DRS) dargestellt, die dann in eine HamNoSys-Notation für Gebärdensprache überführt und in GML transformiert wird. Das Ergebnis ist eine halbautomatische-maschinelle Form der Übersetzung, die Möglichkeiten des Eingriffs offen läßt und eine Verbesserung der Übersetzung erlaubt.

ViSiCAST wird Werkzeuge für die semi-automatische Übersetzung von Text in Gebärdensprache zur Verfügung stellen. Die Übersetzung erfolgt zunächst in eine Diskurs-Repräsentations-Struktur und von dort in mehrere Europäische Gebärdensprachen (BSL, DGS und NGT) über eine computerlesbare Notation. Zu diesem Zweck wird die von uns entwickelte HamNoSys-Notation weiterentwickelt und um neue Funktionen ergänzt. HamNoSys wird darüber hinaus in eine Gesture Markup Language (GML) überführt, eine zum XML-Standard kompatible Notation.

Das Ergebnis wird dann an eine Animationsmaschine, den Avatar, übergeben. ViSiCAST verfügt bereits über hochaufgelöste mittels dreidimensionaler Scannertechnik gewonnene menschliche Avatare, die in Echtzeit Gebärden aus einer Datenbank generieren können. Das System ist in der Lage, Gesichtsausdrücke optisch zu erfassen, Körperposition und Körper-

haltung mittels eines Datenanzugs zu ermitteln und über Datenhandschuhe Hand- und Fingerstellungen detailliert zu erfassen.



Diese Animationsmaschine soll in dreierlei Form existieren: Als Software in einer SetTop-Box, die auf Fernsehübertragungen draufgeschaltet werden kann, als Browser-PlugIn für Anwendungen im World Wide Web und als Software in einem lokalen Computer für die Übersetzung von Face-to-Face-Transaktionen. Diese beispielhaften Anwendungsbereiche sollen im folgenden beschrieben werden.

5. Anwendungsbeispiele für die ViSiCAST-Technologie

ViSiCAST hat drei Anwendungsbereiche als Prüfsteine für die zu entwickelnde Technologie ausgewählt:

- Die Übersetzung von Untertiteln oder anderem Begleittext im Fernsehen.
- Die Übersetzung von Texten im Internet.
- Die Übersetzung von Transaktionen am Schalter der Post.

Wie hat man sich diese Anwendungsfälle vorzustellen?

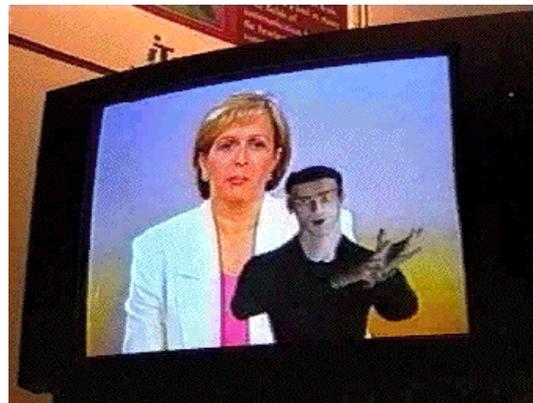
5.1 Übersetzung im Fernsehen

Der Anteil an Übersetzung in Gebärdensprache im Fernsehen ist sehr gering (unter 1%). Dabei sind es nicht die Kosten für das Dolmetschen, die zu diesem geringen Anteil führen, sondern zum einen die ideologischen Vorstellungen der Programmdirektoren, nach deren Ansicht

die Einblendung einer Dolmetscherin das Sehvermögen der Zuschauermehrheit beeinträchtigt und zum anderen die Limitierung der Übertragungskapazität im terrestrischen Fernsehen, das jedes freie Kilobyte profitabel vermarkten kann. Will man den Anteil an gedolmetschten Sendungen substantziell erhöhen, muß demnach eine Methode gewählt werden, die es ermöglicht, den/die Dolmetscher/in gezielt ein- und auszublenden und die zu übertragende Datenmenge gering zu halten. Zwei Methoden sollen in ViSiCAST entwickelt werden:

Die Übertragung der Daten, die ein mit einem Datenanzug ausgestatteter Dolmetscher erzeugt, an eine SetTop-Box beim Empfänger. Die SetTop-Box enthält einen Avatar, der die empfangenen Daten in eine photorealistische menschliche Figur umsetzen und über das empfangene Fernsehbild einblenden kann. Die so als VBI oder MPEG-2 (MPEG-4)-Datenstrom übertragene Datenmenge ist erheblich geringer als ein originales Fernsehbild [Mozelle & Preteux 1998]. Diese Methode modifiziert nur die Datenübertragung. Sie ist unintelligent, indem sie keine Übersetzung benötigt. Untersuchungsgegenstand in diesem Fall sind die Datenkompression, die Methode der Enkodierung und Dekodierung und die Qualität des Bildsignals.

Eine zweite Methode arbeitet mit derselben SetTop-Box beim Empfänger, setzt aber ein System voraus, das in der Lage ist, Lautsprache und Text in Gebärdensprache zu übersetzen. Solch ein System kommt vor allem für Anwendungen in Betracht, in denen keine Dolmetscher zur Verfügung stehen, also z.B. gelegentliche kurze und rasch zu sendende Meldungen. Die Übersetzung wird ViSiCAST-GML produzieren (s.o.), und dieser Code wird den Avatar der SetTop-Box aktivieren und kontrollieren.



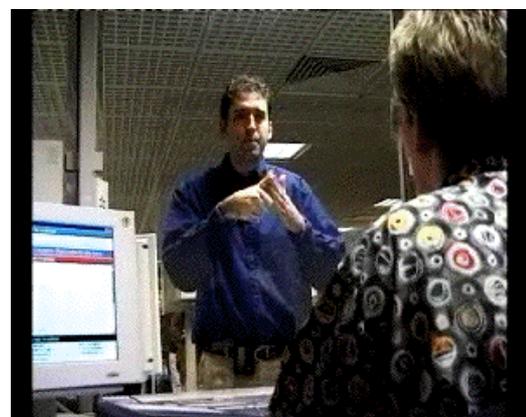
5.2 Übersetzung im Internet

Gehörlose haben nicht nur Probleme, Lautsprache verständlich zu artikulieren, sozusagen „eine Stimme“ zu entwickeln, sondern aufgrund der fehlenden Rückmeldung aus der sprachlichen Interaktion (Gehör) auch erhebliche Schwierigkeiten die Schriftsprache zu erlernen. Lesen und Schreiben ist für sie mühselig und fehlerbehaftet. Trotz der visuellen Benutzeroberflächen der Computer stoßen sie deshalb auf große Probleme bei der effektiven Nutzung von Informationen im Internet. Eines der Ziele des ViSiCAST-Projekts ist daher die Entwicklung eines „Viewer“ für Gebärdensprache im Internet. Der Viewer wird als PlugIn für WWW-

Browser entwickelt und soll einen Avatar generieren, der die aus Text per Übersetzung erzeugten GML-Sequenzen in Gebärden konvertieren kann. Die Software wird zunächst nur eine relativ begrenzte Fähigkeit zum Übersetzen besitzen, wird aber mit dem Fortschritt der maschinellen Übersetzung ausbaubar und erweiterbar sein. Diese Technik kann später auch für andere Multimedia-Anwendungen genutzt werden, etwa für Lernprogramme, die einen Gebärden-Avatar als Tutor oder Guide einsetzen wollen, oder für Sprachlernprogramme, die es den Lernenden ermöglichen wollen, Sätze in Textform einzugeben und zu probieren, wie sie der Avatar in Gebärden übersetzt.

5.3 Übersetzung in Face-to-Face-Transaktionen

Das ViSiCAST-System ist aber ebenso für Face-to-Face-Kommunikation geeignet. Eine Installation wird bereits im UK Post Office getestet [Pezeshkpour et al 1999]. Sobald ein Gehörloser an den Schalter tritt, kann die Postbeamtin ihre Fragen und Antworten in ein Mikrofon sprechen. Die auf einem PC installierte Software erkennt die lautsprachlichen Eingaben und übergibt sie als Text an den Avatar „Tessa“. Tessa generiert dann Gebärdensprache auf einem dem gehörlosen Kunden zugewandten Bildschirm. Das System ist zur Zeit halbautomatisch. Ein Großteil der Gebärden wird aus einer Datenbank geholt. Aber Tessa kann bereits Gebärden für variable Daten, z.B. für Zahlen, die mit den referenzierten Nomen wie Uhrzeit, Geld etc. inkorporiert werden, adhoc generieren und in den Datenstrom einbetten. Zwei wesentlichen Erweiterungen des Systems sind geplant: Die Lautspracherkennung soll noch deutlich verbessert werden, so daß die Postbeamtin nicht an bestimmte Sätze oder Begriffe gebunden ist, die Synthese der Gebärdensprache wird zukünftig auf dem zu entwickelnden Übersetzungssystem (s.o.) aufbauen, und die Äußerungen des gehörlosen Kunden sollen durch eine Kamera erkannt werden, so daß auch die reziproke Richtung der Face-to-Face-Transaktion abgedeckt werden kann. Die restringierte Domäne der Transaktionen in einem Post Office macht den Erfolg eines solchen bidirektionalen Übersetzungssystems eher wahrscheinlich.





Referenzen

Prillwitz, S. et al (1989): Hamburg Notation System for Sign Languages — An Introductory Guide. -In: International Studies on Sign Language and the Communication of the Deaf, Vol. 5. Institute of German Sign Language and Communication of the Deaf: University of Hamburg 1989.

Prillwitz, S./Zienert, H. (1990): Hamburger Notation System for Sign Language: Development of a sign writing computer application. In: Prillwitz, S./Vollhaber, T. (eds): Current Trends in European Sign Language Research. Proceedings of the 3rd European Congress on Sign Language Research. Hamburg July 26-29, 1989. (International Studies on Sign Language and the Communication of the Deaf; 9) Hamburg: Signum (1990) - S. 355-380

Schulmeister, R. (1992): Generierung und Erkennung der Gebärdensprache. -In: ISI 92. Proceedings des 3. Internationalen Symposiums für Informationswissenschaft, Reden zur Eröffnung (= Bericht 22), Universitätsverlag Konstanz 1993 (ISSN 0942-2625).

Schulmeister, R. (1994a): Computer Assistance in Learning Sign Language. In: Brunstein, K./Raubold, E. (eds.): Applications and Impacts. Information Processing '94. IFIP-Transactions A-52. Volume II. 1994. North-Holland, pp. 702-707.

Schulmeister, R. (1994b): Evaluation des Bildtelefons für Gehörlose. In: Das Zeichen 28 (1994), S. 204 - 216.

Mozelle, G./Preteux, F. (1998): „Tele-sign: A compression framework for sign language distant communication“, Proceedings SPIE Conference on Mathematical Modeling and Estimation Techniques in Computer Vision, San Diego, CA, Vol. 3457, July 1998.

Pezeshkpour, F./Marshall, I./Eliott, R./Bangham, A.J. (1999): Development of a legible deaf-signing virtual human. In Proc. IEEE Conf. Multi-Media, Florence, 1999.