

# Umstellung auf HD-Stereokameras

Bearbeitet von: Thomas Hanke

Stand: 2010-30-09

## Ist-Zustand

Das mobile Studio wurde 2009 aus Mitteln des Bund-/Länder-Akademienprogramms eingerichtet, um zwei Informanten (und einen Moderator) in gebärdensprachlicher Interaktion aufnehmen zu können. Die Studioausstattung ist darauf ausgelegt, transportabel zu sein, um Aufnahmen in einzelnen Dialektregionen durchführen zu können und bei möglichst hoher Bildqualität doch die Informanten so wenig wie möglich zu irritieren. Zur Zeit finden sieben Kameras Verwendung:

- Eine HD-Kamera nimmt die Totale auf, erfasst damit die Informanten jeweils von der Seite sowie den Moderator von vorn.
- Je eine HD-Kamera erfasst die Informanten von vorne. (Die Kamera ist über Kopfhöhe des jeweils anderen Informanten positioniert, damit die Informanten einander genau gegenüber sitzen können.)
- Je eine weitere HD-Kamera nimmt die Informanten aus der Vogelperspektive auf, um die Gebärdenraumnutzung deutlich erkennen zu können.
- Von Beginn an werden außerdem zwei Bumblebee-Stereokameras eingesetzt, die auf die Hauptkameras montiert Stereodaten für eine Bildverarbeitung mit dem Ziel halbautomatischer Annotation zu liefern, wie sie zur Zeit im EU-Projekt Dicta-Sign entwickelt wird.

In allen Fällen werden die Daten direkt per SDI, HDMI bzw. Firewire (Bumblebees) auf Rechnern aufgezeichnet (MacPro mit Final Cut Pro bzw. Software der University of Surrey unter Windows für die Bumblebees). Speicherkarten in den HD-Kameras werden als Fallback verwendet. Für die Bumblebees gibt es keine redundante Speicherung.

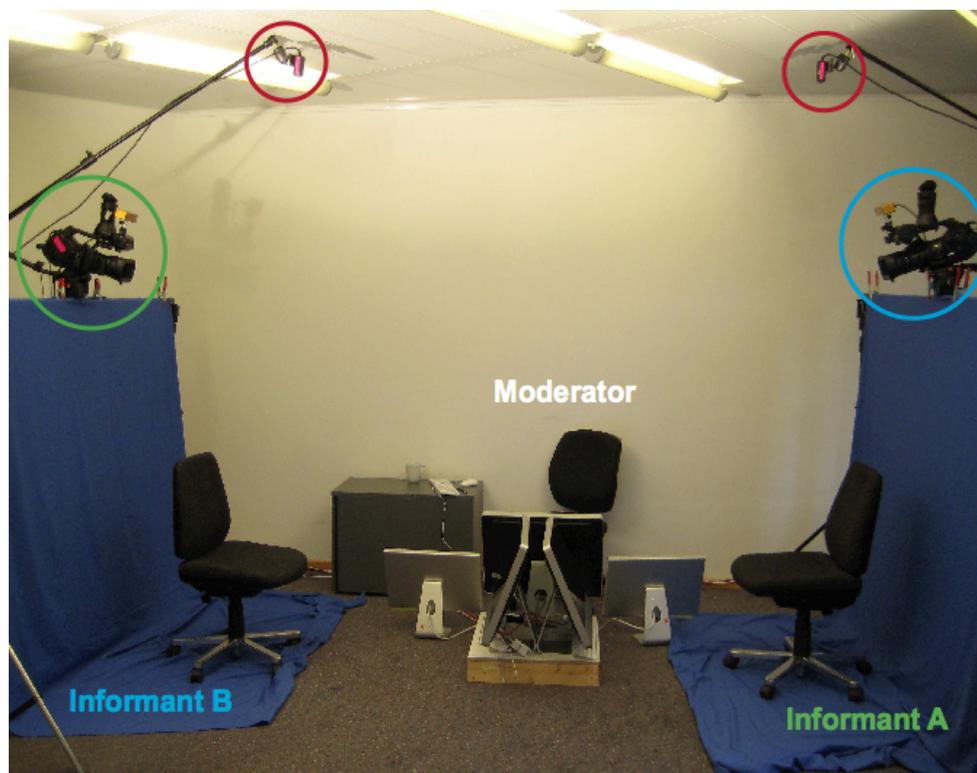


Bild 1: Aufbau des mobilen Studios: Abstand der Informanten 3m, die Bumblebee-Kameras sind die goldenen Kästchen, die auf die EX3 aufgesteckt sind.

In der Transkription von Gebärdensprachdaten bedeutet das Fehlen eines Schriftsystems und damit verbunden einer etablierten Orthographie, dass viele Schritte, die für Lautsprachen grobenteils automatisiert werden können, zeitaufwändig von Hand geleistet werden müssen, insbesondere das Type-Token-Matching (Lemmatisierung). Im Zuge der weiteren linguistischen Analyse wird unter anderem auch die Gebärdenraumnutzung untersucht, was in gewisser Weise eine Rekonstruktion der 3D-Szenerie darstellt. Dreidimensionale Darstellungen sind daher für Gebärdensprachwissenschaftler immer von hohem Interesse.

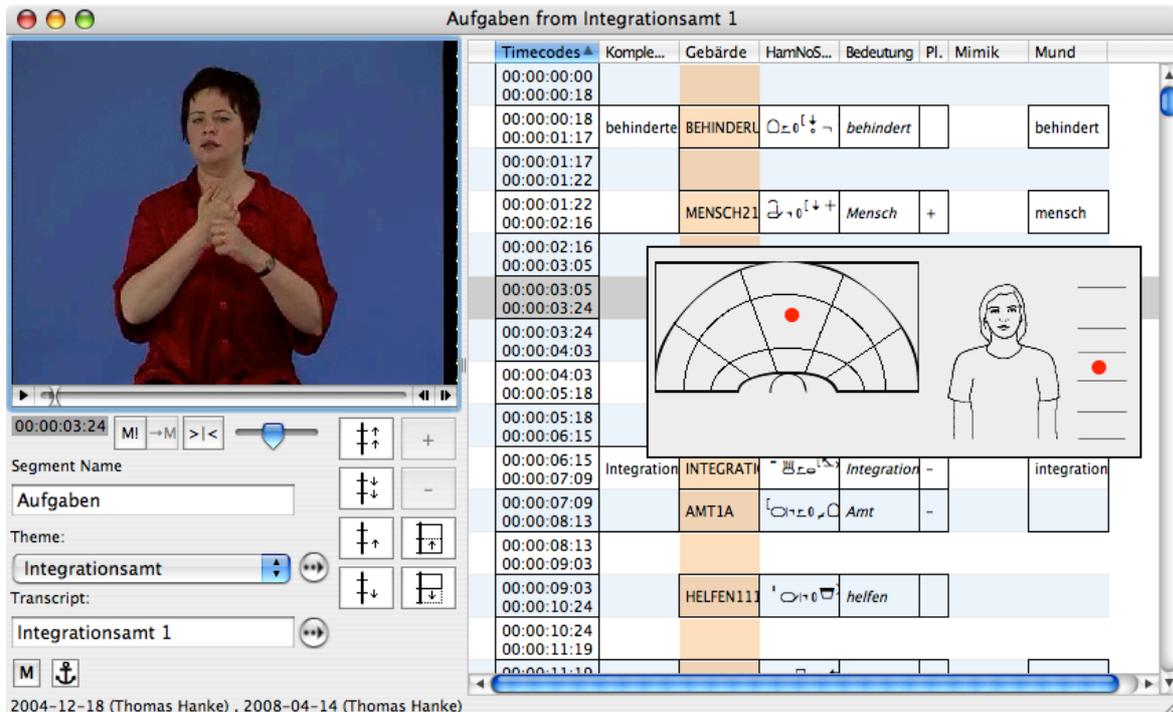


Bild 2: Screenshot aus der Transkriptionsumgebung iLex mit 3D-Verortung

Zum Zeitpunkt der Planung des Studios im Jahr 2007 war eine Ausstattung mit 3D-Technologie jedoch weder notwendig noch in vernünftigem Rahmen realisierbar. Dafür gab es neben budgetären vor allem organisatorische Gründe: Stereo-HD-Kameras waren nicht verfügbar, es hätten also jeweils zwei Monokameras auf einem Rigg verwendet werden müssen. Die Empfindlichkeit der Rigg-Mechanik sowie die Schwierigkeit der Synchronisation der Kamerapositionen erfordern einen Personalaufwand vor Ort, der von den nutzenden Projekten unmöglich hätte geleistet werden können.

Im Laufe der letzten zwei Jahre gab es in dieser Hinsicht allerdings eine rasante Entwicklung:

- Das Projekt Dicta-Sign war in der Lage, erste praxistaugliche Ergebnisse der Stereobildfolgenverarbeitung für Gebärdensprache zu erzielen.
- Die Firma Panasonic hat eine 3D-HD-Kamera vorgestellt (April 2010), die aus Sicht von 3D-Filmern viel zu wünschen übrig lässt, unsere Studioanwendung jedoch ideal bedient.
- Ein Markttrend zur Einführung von 3D-Technologie im Consumer-Bereich, verstärkt durch Kinoerfolge wie „Avatar“, könnte die Sehgewohnheiten der zukünftigen Nutzer unserer Korpusdaten nachhaltig ändern.

## Planung

Vor diesem Hintergrund haben wir uns, vorbehaltlich des positiven Ausgangs eines Praxistests am 10.11.2010 mit einem Panasonic-Serienmodell – wir hatten bisher nur Zugang zu einem Vorseriengerät –, dazu entschlossen, das Studio auf 3D-Technologie umzustellen. Finanzielle Unterstützung hierfür erfahren wir dankenswerterweise seitens der Hamburgischen Akademie der Wissenschaften.

Konkret werden die drei Hauptkameras durch HD-Stereokameras ersetzt. Die dadurch frei werdenden Kameras werden mit leichter Verschiebung der Perspektive weiter verwendet, um bei geringerer räumlicher eine höhere zeitliche Auflösung aufzuzeichnen (720p50 statt 1080i25). Dafür entfallen die beiden Bumblebees.

Für die wissenschaftliche Nutzung der Studiodaten versprechen wir uns erhebliche Vorteile:

- Während wir noch weit von automatischer Gebärdenerkennung aus dem Video entfernt sind, bringen halbautomatische Annotation oder Überprüfung manueller Annotation schon einen merklichen praktischen Gewinn. Ein einfaches Beispiel hierzu ist die Entscheidung, ob eine Person mit einer oder zwei Händen gebärdet, was vergleichsweise einfach aus dem Video zu ermitteln ist und mit der manuellen Annotation abgeglichen werden kann: Die Transkriptionsumgebung kann automatisch mögliche Konflikte erkennen und signalisieren. Hat etwa der Transkribent eine zweihändige Gebärde notiert, die Bildererkennung jedoch nur eine Hand in Bewegung ermittelt, muss der Transkribent überprüfen, ob er einen Fehler in der Type- Zuordnung gemacht oder einfach das relativ häufig vorkommende Weak Drop nicht notiert hat.
- Die automatische Verfolgung von Lokationen im Gebärdenraum kann präziser arbeiten als bisher und damit Modellierungen metrisch genauer visualisieren.

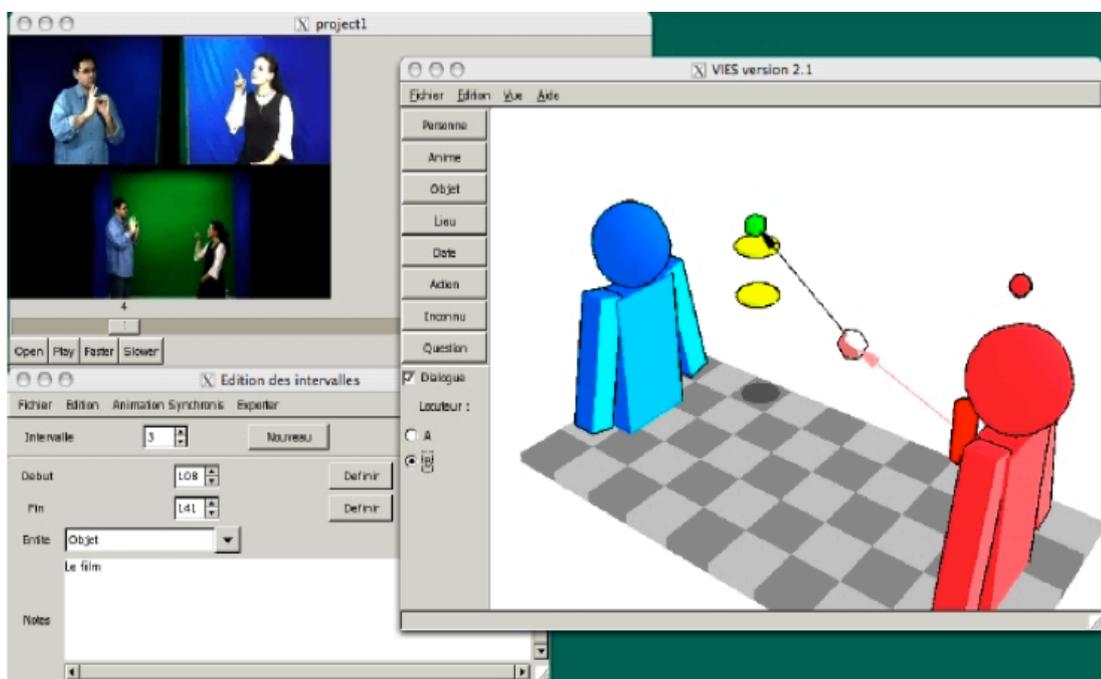


Bild 3: Prototyp eines 3D-Modells des Gebärdenraums mit den zuvor eingeführten Referenzobjekten, Dicta-Sign/IRIT-Toulouse

- Die aus phonetischer Notation gesteuerte Avatar-Animation von Gebärden kann ebenfalls mit metrischen und zeitlichen Daten abgeglichen und damit realistischer gemacht werden.

