

3D-Gestenerkennung und Sprache - ICONIC

im Rahmen des Seminars:

Multimodale Mensch-Maschine-Kommunikation

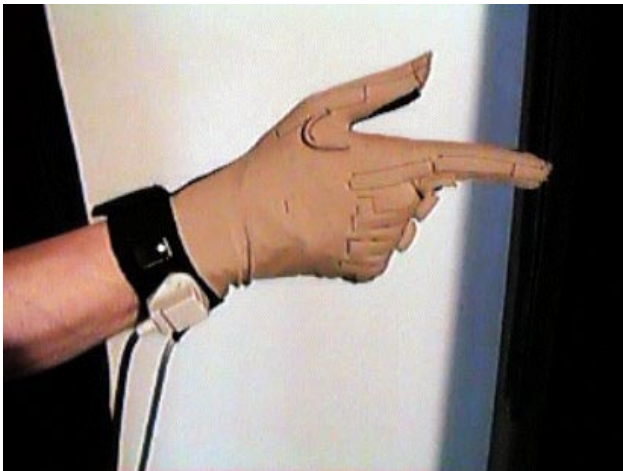
Marco Balke (mbalke@TechFak.Uni-Bielefeld.DE)

Struktur des Vortrags

- Was heißt 3D-Gestenerkennung
- Verschiedene Gestenkategorien
- Unterschied zwischen emblematischen Gesten und natürlicher koverbaler Gestik
- Das ICONIC-System
 - Ziele des Systems
 - Systembeschreibung
 - gewonnene Erkenntnisse
- Kurzvorstellung anderer Systeme am MIT
- Zusammenfassung

Was heißt 3D-Gestenerkennung

- Gesture-Space: Raum vor dem Körper, in dem gestikuliert wird
- Handschuhe: Erkennen Gelenkwinkel der Hand
- Sensoren: Erkennen Lage im Raum



- Eyetracker: Erkennt Blickrichtung
- Verarbeitung: Daten bekommen Zeitstempel und werden zueinander in Relation gesetzt

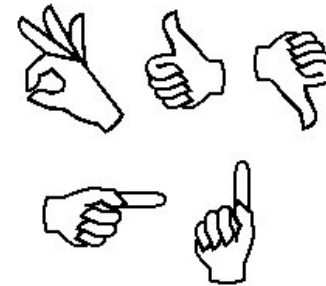
Verschiedene Gestenkategorien

„Gesten“

Körperbewegungen, die bedeutsame Information übermitteln

Interessant für Mensch-Maschine-Kommunikation:

1. **emblematische Gesten (Symbole):**
Bedeutung per Konvention festgelegt
2. **deiktische Gesten:** Handlungen des Zeigens auf ein Objekt oder eine Region
3. **mimetische Gesten:** Handlungen, die die Interaktion mit einem Objekt imitieren
4. **ikonische Gesten:** Beschreibung der Form oder Lage von Objekten mit den Händen



Koverbale Gesten: Gesten, die das Sprechen begleiten

Emblematische Gesten - Beispiel



Hand-Waving-Both



Hand-Waving-Right



Hand-Waving-Left



To-Right



To-Left



To-Top



To-Bottom



Round-Clockwise



Round-Counter-clockwise



Stop



Come



Nod-Yes

Unterschied zwischen emblematischen Gesten und natürlicher koverbaler Gestik

- emblematische Gesten:
 - Benutzer muß sich dem System anpassen
 - Unterschied Funktionstaste - Geste?
- natürliche koverbale Gesten:
 - Menschen benutzen im Gespräch begleitende Gesten
 - Gesten können in verschiedenen Kontexten unterschiedliche Bedeutung haben
 - verschiedene Gestenarten

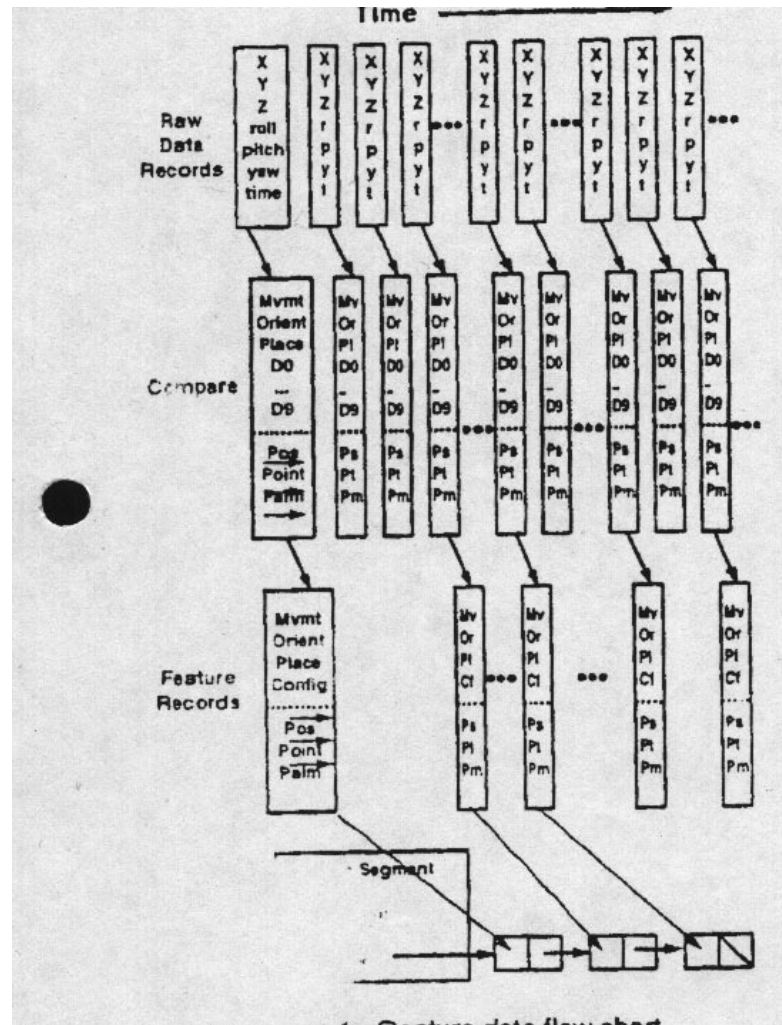
Vorteile der koverbalen Gestik

- Problem bei Befehlseingabe durch emblematische Gesten:
 - Benutzer muß *Vokabular* erlernen
 - Befehl durch Gestik hat gleiche Wirkung wie Druck auf Funktionstaste
- Lösung: Interpretation natürlicher koverbaler Gestik
 - System paßt sich Benutzer an, nicht der Benutzer dem System
 - Gesten benutzen zum Auflösen von Ambiguitäten der Sprache

ICONIC - Allgemeines

- entwickelt am MIT von Carlton J. Sparrel und David B. Koons ca.1994
- Zusammenarbeit mit den anderen hier noch vorzustellenden Projekten
- Betrachtet hauptsächlich ikonische und mimetische Gesten
- Eingabe: Sprache und Datenhandschuh
- Gestik wird zum auflösen von Ambiguitäten der Sprache benutzt
- Abstraktion von den Raw-Daten des Handschuhs auf eine merkmalsbasierte Zwischendarstellung

ICONIC - merkmalsbasierte Darstellung



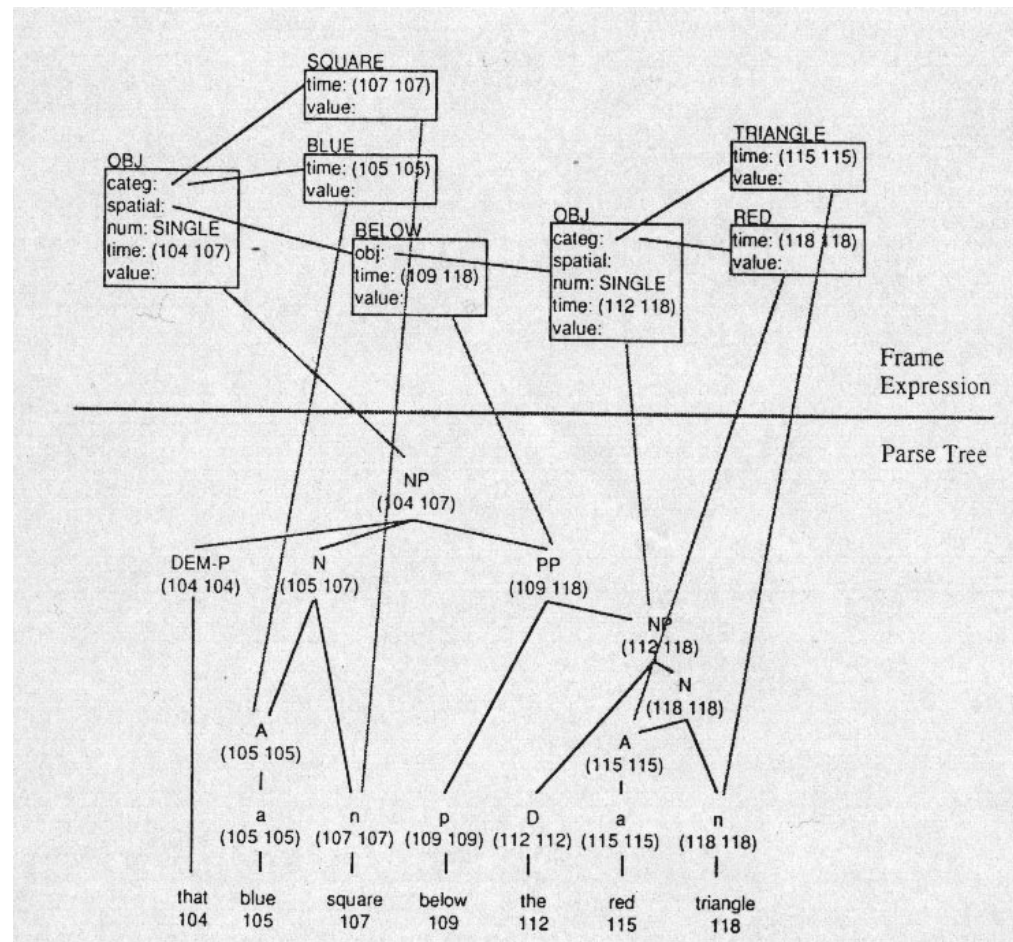
ICONIC - Ziele des Systems

- Schaffung eines multimodalen Systems, das dem Benutzer kein neues Erlernen von Befehlen abverlangt
- intuitive Bedienung
- Kommunikation mit dem System soll zwischenmenschlicher Kommunikation entsprechen

ICONIC - Systembeschreibung

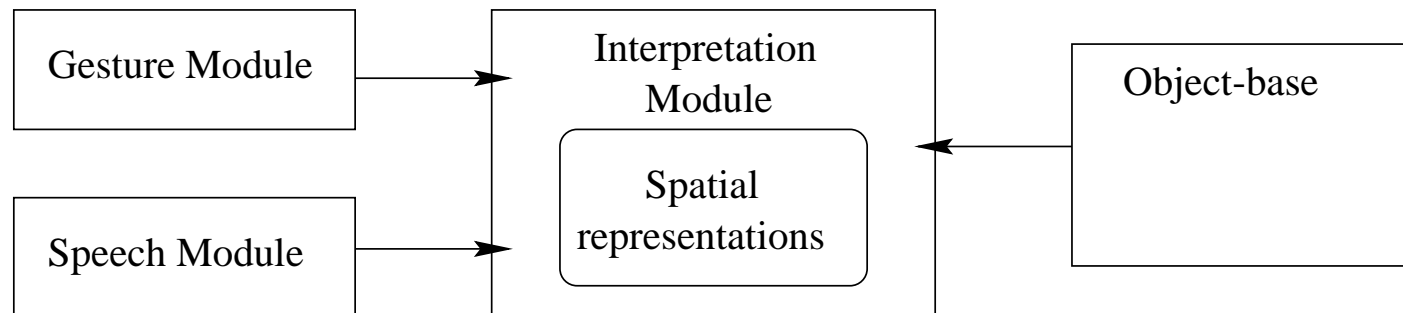
- Datenhandschuh
 - Raw-Daten werden mit Zeitstempel versehen
 - nur Daten die Änderung gegenüber Vorzustand aufweisen werden weiter betrachtet
 - Raw-Daten → merkmalsbasierte Zwischendarstellung
- Sprache
 - Syntaktische und semantische Analyse
 - Spracherkenner versieht Daten mit Zeitstempel

ICONIC - Sprachverarbeitung



ICONIC - Systembeschreibung

- Interpreter
 - verarbeitet Sprache und Gesten parallel
 - bei Ambiguitäten der Sprache wird geprüft, ob Gesten diese auflösen können



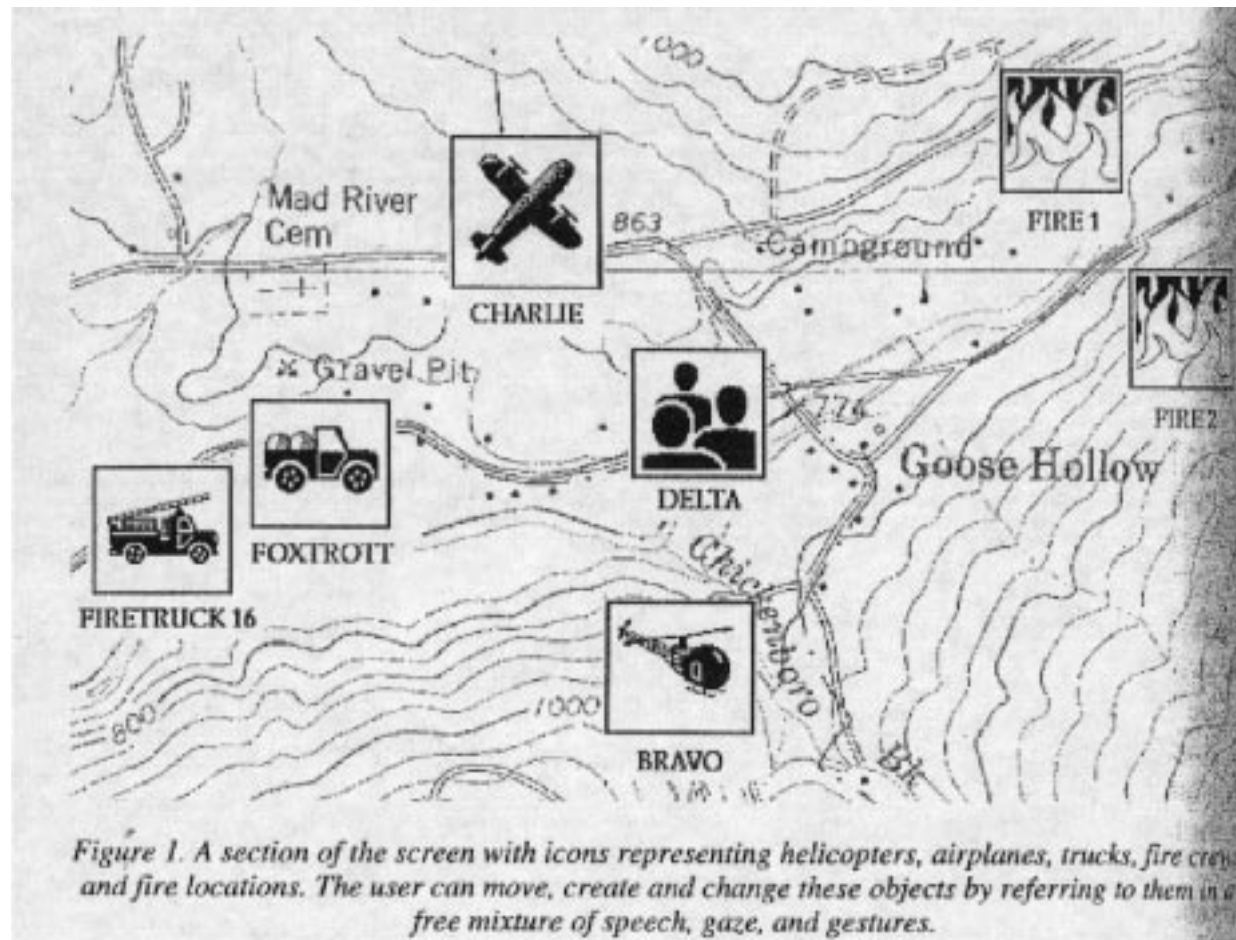
ICONIC - Gewonnene Erkenntnisse

- ICONIC erlaubt dem Benutzer sich wie in einer *normalen* Kommunikation zu verhalten
- parallele Verarbeitung von Sprache und Gestik erlaubt Interpretation der Gesten im Kontext der aktuellen Situation
- Dadurch wesentlich mächtiger als Befehlseingabe über emblematische Gesten

Andere Systeme am MIT - System 1

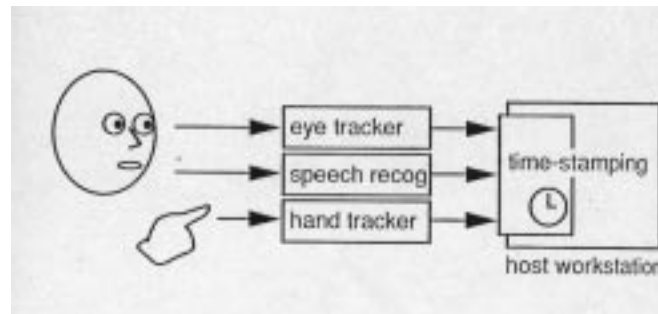
- entwickelt von Koons / Sparrell / Thorisson ca.1994
- Systembeschreibung
 - Dateneingabe über Sprache, deiktische Gesten und Eyetracker
 - Planen von Feuerbekämpfung auf 2-dim. Leinwand, verschieben von Einheiten (ähnelt CUBRICON)
 - Interpreter kennt Kontext der Anwendung

Andere Systeme am MIT - System 1



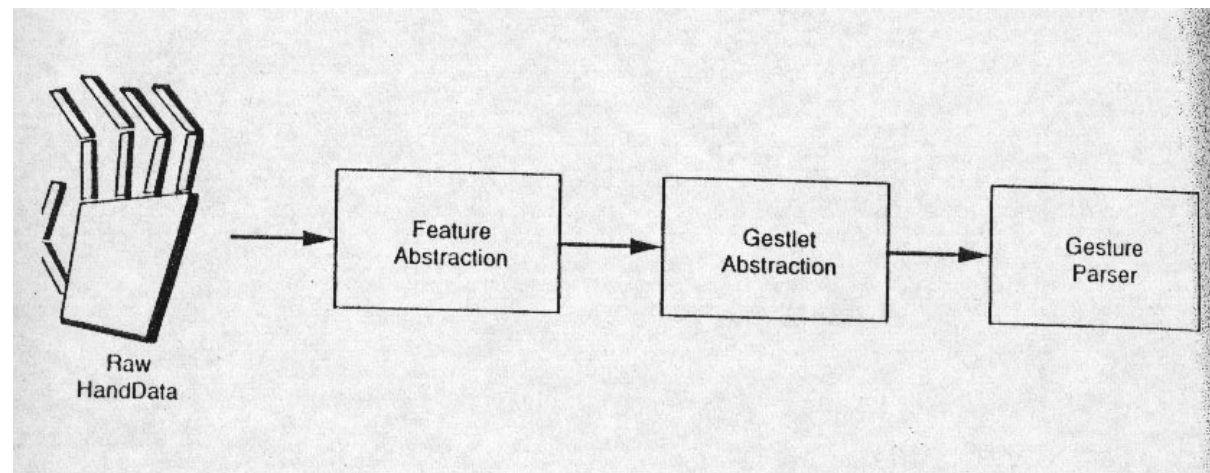
Andere Systeme am MIT - System 1

- Besonderheiten:
 - Freie Kombination aller Modalitäten
 - Benutzung der Eyetrackerdaten wie Zeigegeste, um genauere Zielerfassung zu ermöglichen
 - Daten werden durch unabhängige Parser verarbeitet, mit Zeitstempeln versehen und gemeinsam interpretiert
 - Ausgabe in einer gemeinsamen Framebasierten Zwischendarstellung



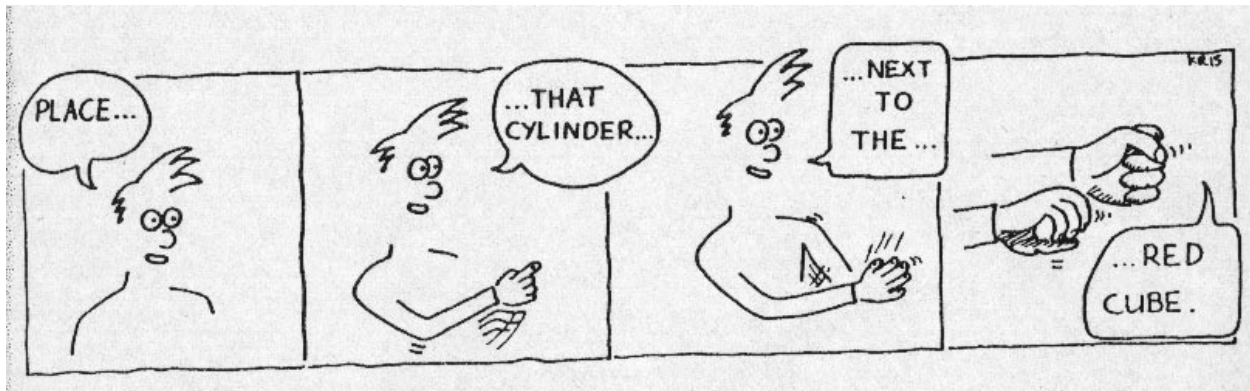
Andere Systeme am MIT - System 2

- entwickelt von Koons / Sparrell / Thorisson ca.1994
- Systembeschreibung
 - Dateneingabe über Sprache, deiktische / mimetische / ikonische Gesten
 - Manipulation einer *Blockwelt*

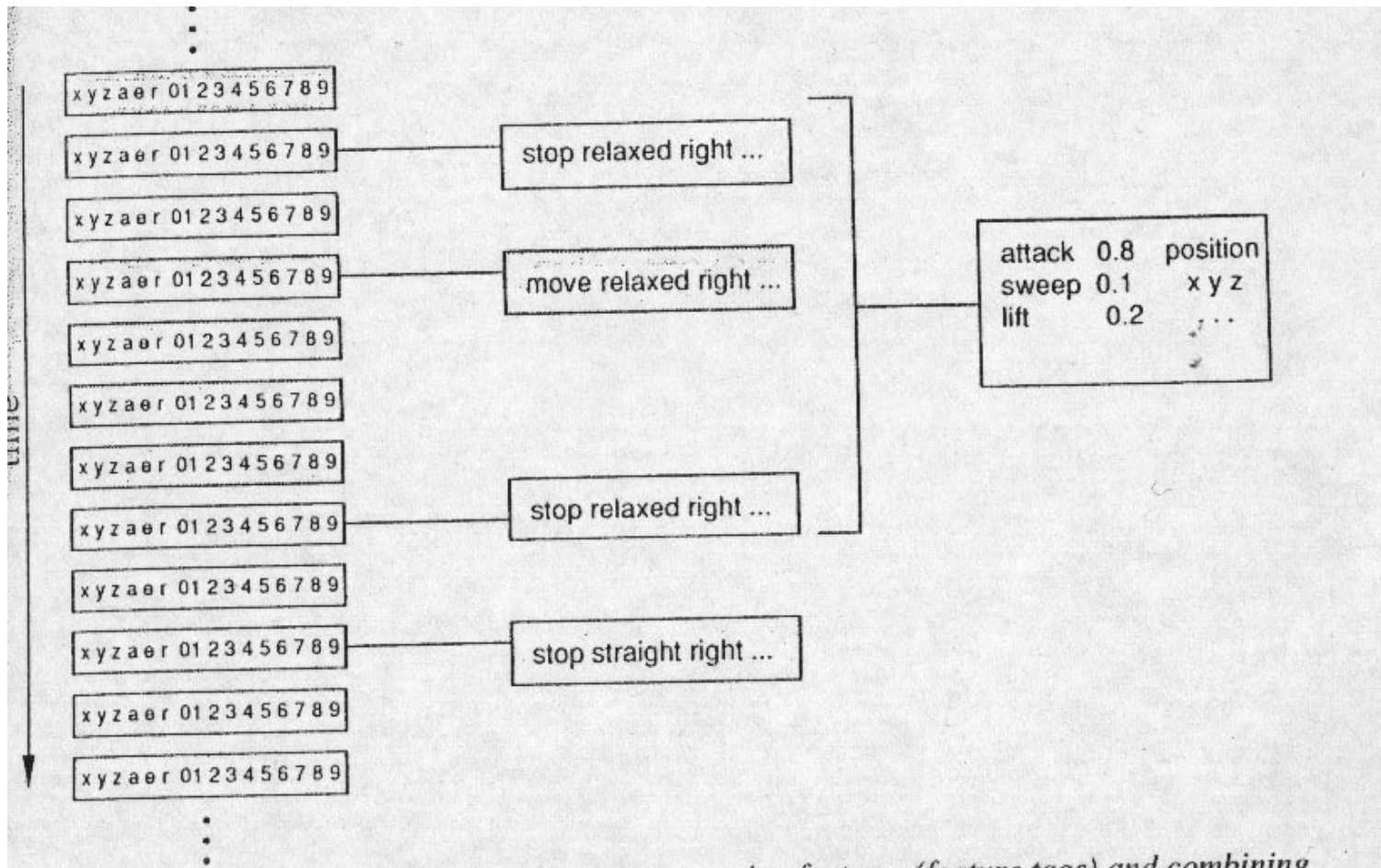


Andere Systeme am MIT - System 2

- Besonderheiten:
 - Interpreter bestimmt den bedeutungsgehalt einer Geste im Kontext der *Welt* und der Spracheingabe
 - Verwendung sog. Gestlets
 - * Gestlet: Zwischendarstellung zwischen Raw-daten des Handschuhs und der ganzen Geste
 - * Gestik als Unterstützung der Spracheingabe



Andere Systeme am MIT - System 2



Andere Systeme am MIT - System 3

- entwickelt von Alan Wexelblatt ca. 1994
- Systembeschreibung
 - Versuchspersonen sahen Ausschnitte aus *Casablanca* oder *Terminator*
 - danach sollte Versuchsperson Handlung der Szene wiedergeben
 - Beobachtung der Gestik der Versuchspersonen
 - Eingabe durch einen sog. Data Suit:
 - * Eyetracker
 - * Datenhandschuhen
 - * 4 Sensoren(Hände, Schulter, Kopf)

Andere Systeme am MIT - System 3

- Ziele des Systems: Überprüfung folgender Hypothesen
 - Actionszenen werden mit mehr Gesten beschrieben als nicht-Action-Szenen
 - Versuchspersonen, die nicht in ihrer Muttersprache beschreiben, benutzen mehr Gesten
 - Männer benutzen wesentlich mehr Gesten als Frauen (Männer bevorzugen eine offensivere Kommunikation)
 - Ergebnisse:
 - keine der Hypothesen trifft zu
- ⇒ Es lassen sich verallgemeinernden Aussagen über die Benutzung von Gesten machen

Zusammenfassung

- 3D-Gestenerkennung ist ein hat das Potential zu einem mächtigen Werkzeug in der Mensch-Maschine-Kommunikation.
- Durch Interpretation der natürlichen koverbalen Gestik ist der Mensch nicht gezwungen, neue Befehle zu erlernen.
- Die Interpretation der verschiedenen Modalitäten muß parallel erfolgen, da die Gestik ihre Bedeutung aus dem Kontext erhält.
- Gerade mimetische und ikonische Gesten enthalten viele Informationen, die verbal nur umständlich dargestellt werden können.