

# Robustheit durch Multimodalität

Hendrik Hasenbein\*

11.1.2001

Seminar  
„Multimodale Mensch-Maschine-Kommunikation“  
WS'00/'01  
Universität Bielefeld

---

\*hhasenbe@techfak.uni-bielefeld.de

# Inhaltsverzeichnis

<b>1 Motivation</b>	<b>2</b>
<b>2 Methoden</b>	<b>2</b>
<b>3 Interact</b>	<b>2</b>
3.1 Spracherkennung mit Lippenlesen . . . . .	2
3.2 Handschrifterkennung . . . . .	3
3.3 Gestenerkennung . . . . .	3
<b>4 Fazit</b>	<b>3</b>

## 1 Motivation

Multimodalität soll die Robustheit der Mensch-Maschine-Kommunikation durch Nutzung aller Modalitäten, welche der Mensch bei seiner eigenen Kommunikation verwendet, erhöhen. Durch die Verwendung weiterer Modalitäten sollen so Uneindeutigkeiten einer einzelnen Modalität überbrückt werden.

## 2 Methoden

Es gibt verschiedene Möglichkeiten mehrere Modalitäten zu integrieren. Die Modalitäten können bereits auf Eingabeebene oder nach Bearbeitung und Auswertung kombiniert werden. Je später die Modalitäten verknüpft werden, desto eher gehen Zusatzinformationen verloren, die zur Robustheit beitragen können. Allerdings kann durch spätere Kombination die Erkennung parallelisiert werden. Zusätzlich sollten Modalitäten gewählt werden, die möglichst wenig korrelieren, um ein Maximum an Zusatzinformationen zu bekommen.

## 3 Interact

Interact ist ein Projekt zur Kombination der Modalitäten von Alltagskommunikation. Das Projekt umfasst mehrere Einzelprojekte an der Carnegie Mellon University in Pittsburgh und an der Universität von Karlsruhe, die sich mit konkreten Anwendungsfällen beschäftigen. Diese Projekte beinhalten meistens eine bereits gebräuchliche Mensch-Maschine-Schnittstelle, die um eine weitere Modalität ergänzt wurde, die Menschen auch bei dieser Kommunikationsform mitbenutzen.

### 3.1 Spracherkennung mit Lippenlesen

Dieses Projekt befaßt sich mit der Spracherkennung, die um Lippenlesen erweitert wurde. Akustische und visuelle Daten werden parallel bearbeitet. Die

Ausgaben beider Verfahren werden dann analysiert, um die Ausgabe zu erzeugen. Beide Stufen, benutzen Multi-State Delay Neural Networks. Auf der aufgespaltenen Stufe ermitteln die Netze die kleinsten noch unterscheidbaren Merkmale (viseme/phoneme) der Eingaben. Diese Ausgabe benutzt die zweite Stufe zur Bestimmung der Buchstaben.

Die Probleme bei dieser Kombination liegen zum größten Teil beim Lippenlesen, wo zuerst der Mundausschnitt erkannt werden muß und unterschiedliche Mundhaltung die Erkennungsleistung stark reduzieren. Durch aufwendigeres Training der Netze auf verschiedene Mundhaltungen beim Sprechen ist diese Fehlerquelle annähernd ausgemerzt worden.

### **3.2 Handschrifterkennung**

Ein weiteres Projekt befaßt sich mit der Handschrifterkennung, welche allerdings im Gegensatz zum normalen OCR schon beim Schreiben des Textes geschieht (OLCR). Dadurch erfährt das System zusätzliche Informationen, wie Richtungsvektor und Andruckstärke, die sonst nur schwer aus der statischen Aufnahme des Textes zu bestimmen sind. Zu diesem Zweck wird in Intervallen Stiftandruck und Bewegungsrichtung aufgezeichnet. Zusammen mit einer Bitmap des aktuellen Zustandes werden die Daten von einem Multi-State Time-Delay Neural Network ausgewertet. Aus den gelieferten Zuständen des Netzes werden mittels Dynamic Time Warping die Wörter aus dem Wörterbuch herausgesucht.

### **3.3 Gestenerkennung**

Dieses Projekt befaßt sich mit der Erweiterung von Handschrift- oder Spracherkennung um Gestenerkennung. Das Ziel bei diesem Projekt ist es dem jeweiligen Experten weitere Hinweise auf die erkannten Texte zu geben, um z.B. Korrekturen während des Sprechens besser erkennen zu können. Hierzu werden die Gesten getrennt erkannt und dann dem Datenstrom zum Kommandointerpreter hinzugefügt.

## **4 Fazit**

Multimodalität erhöht die Robustheit der Kommunikation, da durch die weiteren Kommunikationsmodalitäten ähnliche Fälle in einer Modalität unterschieden werden können. Zusätzlich läßt sich durch die Multimodalität die Erkennung parallelisieren, so daß es nicht zu starken Geschwindigkeitseinbußen kommt.

## Literatur

- [1] Alex Waibel, Minh Tue Vo, Paul Duchnowski, Stefan Manke:  
Multimodal Interfaces, 1994
- [2] Interactive Systems Labs  
<http://werner.ira.uka.de/ISL.multimodal.html>