

Multimodale Interfaces

Vortrag von Fabian Rutishauser und Nicolas Honegger

Ziele

- Aufzeigen des konkreten Nutzens von Multimodalität
- Zeigen des aktuellen Forschungsstandes
- Vorstellen von konkreten Beispielen
- Aufzeigen, wie heterogene Daten zusammengeführt werden können

Inhalt:

1. Einleitung

2. Multimodale Interfaces am Beispiel von:

(I) Vocabulary Recognition

(II) Affect-Sensitive Human Computer
Interaction

3. Fazit & Ausblick



Inhalt:

1. Einleitung

2. Multimodale Interfaces am Beispiel von:

(I) **Vocabulary Recognition**

(II) Affect-Sensitive Human Computer
Interaction

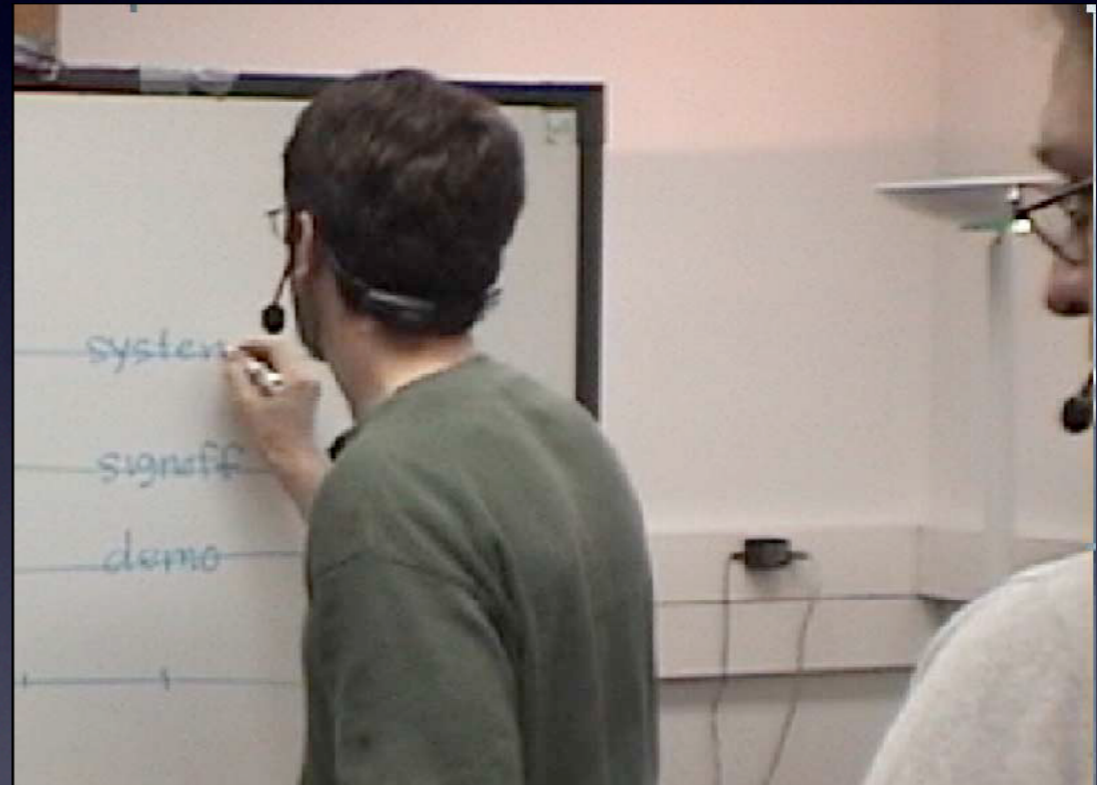
3. Fazit & Ausblick

Das Problem

- Verschiedene User - verschiedene Wörter
- Systeme lernen nicht dynamisch
- Spracheingabe und Handschriftenerkennung sind nicht eindeutig

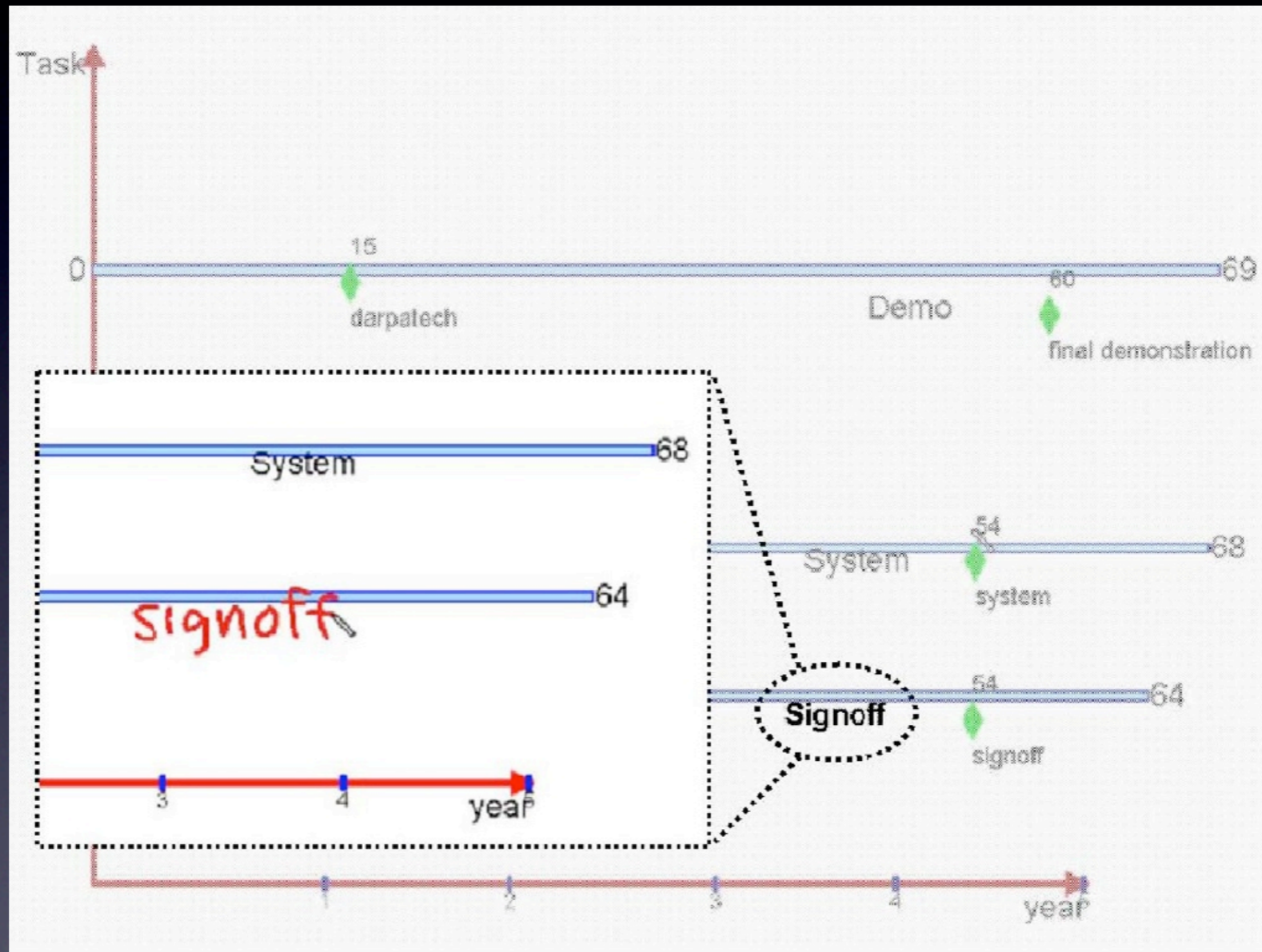
Das System

- Whiteboard, Microfon, Recheneinheit
- Sprach und Handschrift
- Gantt-Chart



Edward C. Kaiser. "Multimodal New Vocabulary Recognition through Speech and Handwriting in a Whiteboard Scheduling Application,"
Proceedings of the International Conference on Intelligent User Interfaces, San Diego, CA., January 9-12, 2005

Der Chart



Edward C. Kaiser. "Multimodal New Vocabulary Recognition through Speech and Handwriting in a Whiteboard Scheduling Application,"
Proceedings of the International Conference on Intelligent User Interfaces, San Diego, CA., January 9-12, 2005

OPS-Tupel

- Sprach und Handschrifteneingabe nicht eindeutig
- Jeweils verschiedene Werte mit Wahrscheinlichkeit
- Kombination von Orthografie, Phonetik, Semantik → OPS - Tupel
- Bsp: (hand, hh ae n d, Nomen)

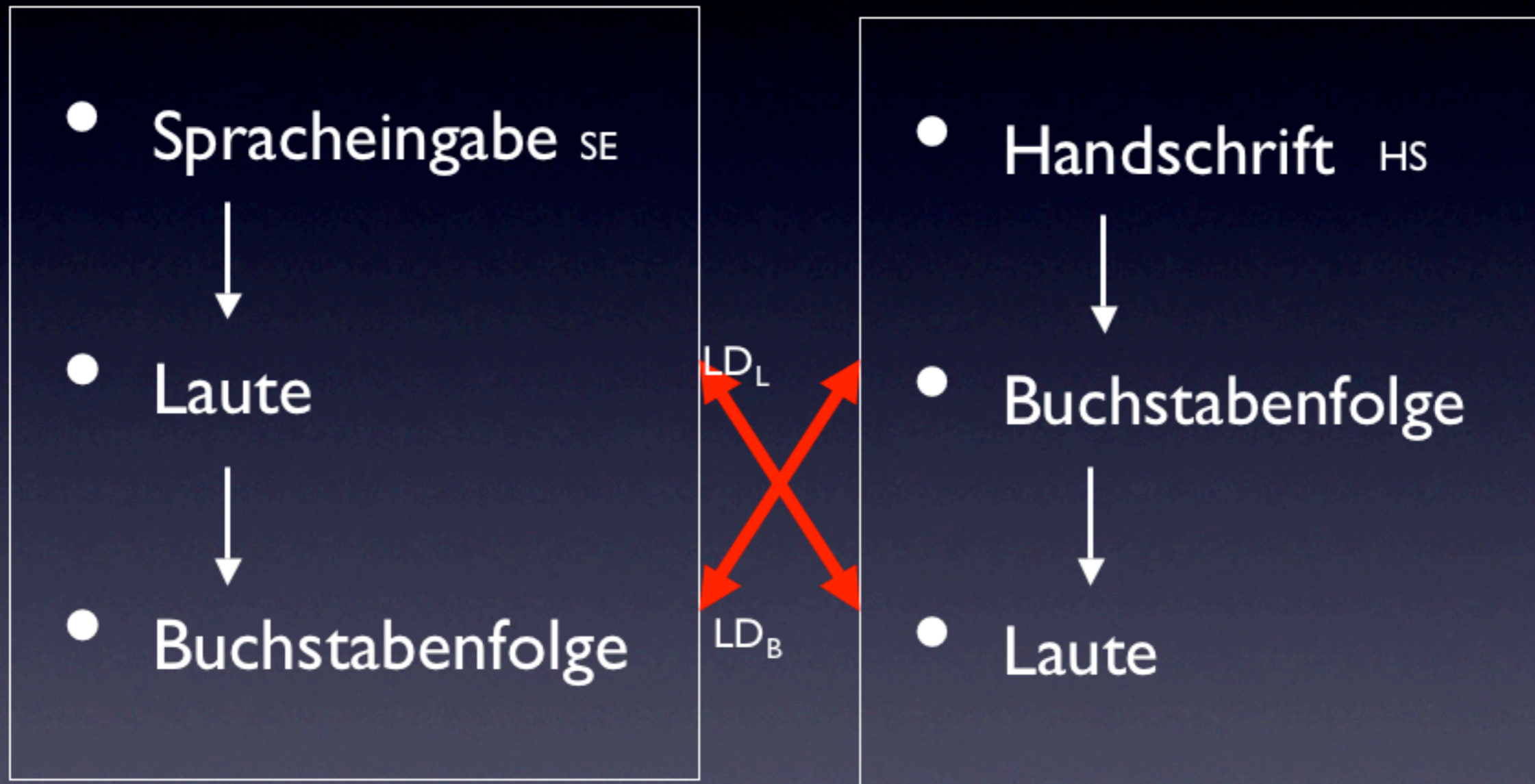
Auswahl

- Jedes Tupel wird anhand verschiedener Eigenschaften bewertet
- Tupel mit höchster Punktzahl aufgenommen
- Alle anderen Tupel nicht aufgenommen

Die Technik

- Kreuzvergleich
- Levenshtein Abstand - Edit Distance
- → Anzahl Mutationen um zwei Strings gleich zu machen

Kreuzvergleich



$$\text{Punktzahl} = LD_{SE} * SE * LD_{HS} * HS$$

Versuch

- 54 Instanzen
- 54 nicht bekannte Wörter
- 18 Schlüsselsätze
- Nicht bekannte Wörter gesprochen

Das Ergebnis

Testanordnung	Fehlerquote
Unimodale Spracherkennung	47.33%
Unimodale Schrifterkennung	19.33%
Multimodale Sprach und Schrifterkennung	16.33%

Inhalt:

1. Einleitung

2. Multimodale Interfaces am Beispiel von:

(I) Vocabulary Recognition

(II) **Affekt - Sensitive Human
Computer Interaction**

3. Fazit & Ausblick

Affekt Sensitive Interfaces

- Computer erkennt den psychischen und physischen Zustand einer Person
- Computer hat die Fähigkeit auf den Zustand einer Person zu reagieren
- Ziel ist das Erzeugen einer künstlichen „emotionalen Intelligenz“

© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com

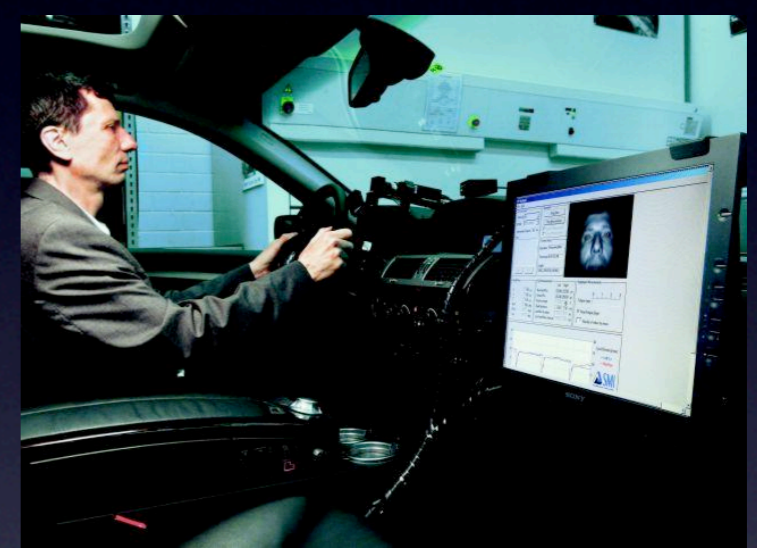
© Mike Baldwin / Corbis
BALDWIN



“Artificial intelligence is overrated. We find most people prefer the user-friendly advantage of artificial stupidity.”

“Artificial intelligence is overrated. We find most people prefer the user-friendly advantage of artificial stupidity.”

Multimodale Interfaces für Zustandserkennung



- BMW forscht an System das den physischen und psychischen Zustand von Fahrern erkennen
- System erfasst Gesichtsmimik, Augenlieder schlag und Fahrverhalten über Lenkrad



Problem



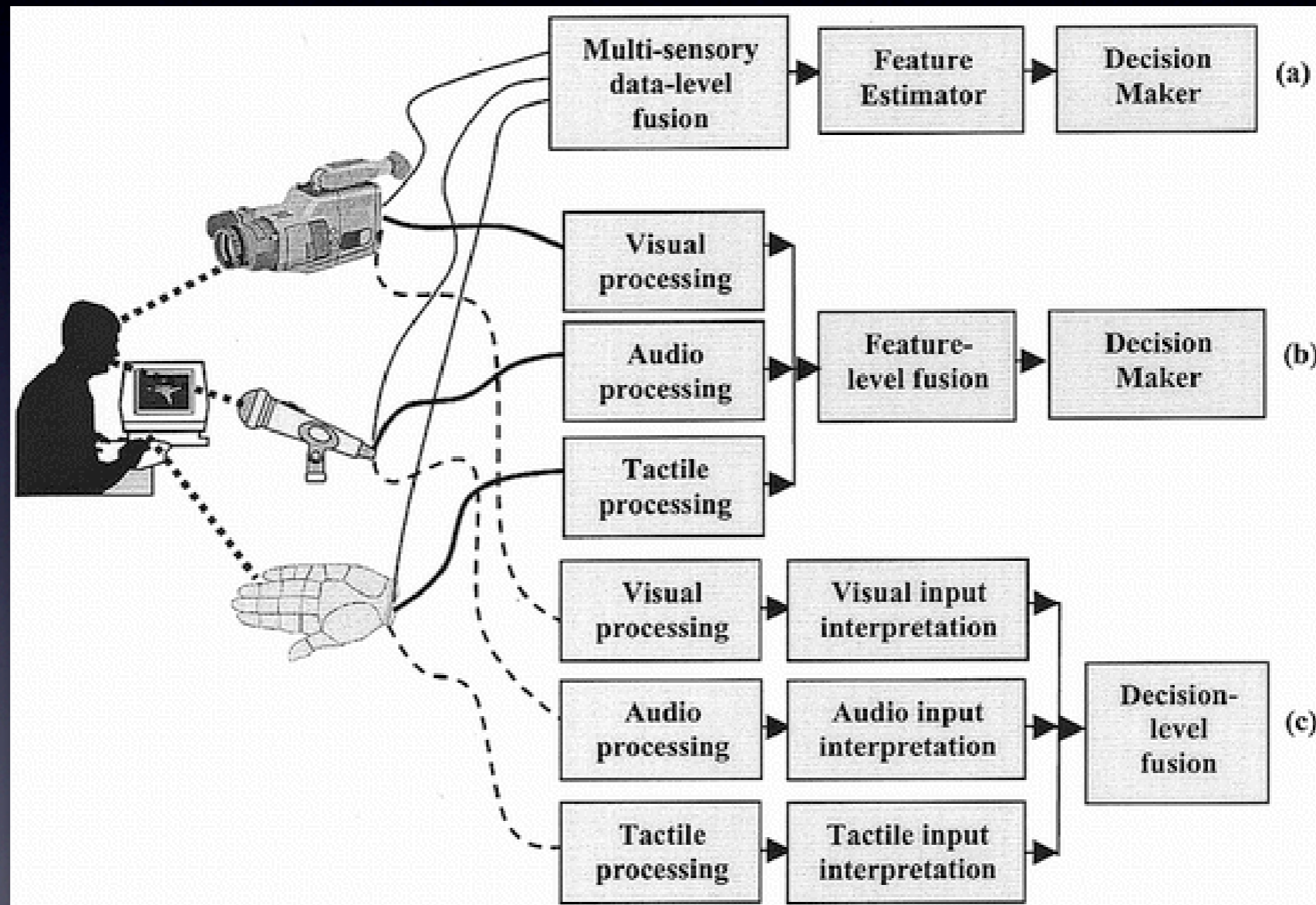
- Lässt sich ein Zustand wie Müdigkeit für einen Computer beschreiben?
- Welche Beobachtungen (z.B. Mimik, Gestik, Sprache) geben einen Aufschluss über den Zustand einer Person?
- Wie lassen sich verschiedene Beobachtungen zu einer Person miteinander kombinieren?

Beispiel: Abbildung von Emotionen auf Sprache

	Happiness	Anger	Fear	Sadness
Pitch	Increase in mean [8], [5], range [41], [5], variability [5]	Increase in mean [27], [5], range [5], variability [5]	Increase in mean [5], range [118], [5]	Decrease in mean [118], [5], range [118], [5]
Intensity	Increased [8], [5]	Increased [8], [5]	Normal [24]	Decreased [27], [5]
Duration (speech rate)	Increased rate [27], [5] Slow tempo [8]	Increased rate [27], [5] Reduced rate [118]	Increased rate [5] Reduced rate [109]	Reduced rate [27], [5]
Pitch contour	Descending line [41]	Descending line [5], stressed syllables ascend frequently & rhythmically [41], irregular up & down inflection [27]	Disintegration in pattern and great number of changes in the direction [24]	Descending line [27], [5]

Pantic, M.; Rothkrantz, L.J. M.; Toward an Affect-Sensitive Multimodal Human Computer Interaction. In: Proceedings of the IEEE, Vol. 91, September 2003

Verbinden von Information von multimodalen Interfaces



Pantic, M.; Rothkrantz, L.J. M; Toward an Affect-Sensitive Multimodal Human Computer Interaction. In: Proceedings of the IEEE, Vol. 91, September 2003

Konzepte des Systems

- $1+1 > 2$ Zwei kleine Inputs von verschiedenen Quellen sind stärker gewichtet als ein starker Input
- Interpretation der Signale ist Kontext abhängig
- System muss mit Widersprüchen umgehen können

Inhalt:

1. Einleitung

2. Multimodale Interfaces am Beispiel von:

(I) Vocabulary Recognition

(II) Affekt - Sensitive Human Computer
Interaction

3. Fazit & Ausblick

Fazit und Ausblick

- Traditionelle Eingabemethoden sind für viele Geräte ungeeignet
- Verbindung von multimodalen Eingabemethoden ist komplex, kann aber Schnittstellen sehr robust machen
- Benutzer möchten je nach Situation ein anderes Interface benutzen

Zeit für...



&



Next iPod: Gesten Interface

