Bridging the Semantic Gab

-

Multimodale Suche I

David Caccamo

Bachelor Software- und Internettechnologie

21.06.2006

Gliederung

1. Motivation
2. Multimodal – was bedeutet das?
3. Die Multimodale Suche
   - Definition und Abgrenzung
   - Suchen mit Hilfe multimodaler Interaktion
   - Analyse mit Hilfe multimodaler Interaktion
4. Beispiele im Web
5. Fazit & Ausblick

1. Motivation

Informationsüberfluß in der virtuellen Welt

Informationsdurst der User

Mensch-Computer-Interaktion

Heutige Situation

2. Multimodal – Was bedeutet das?

Table 1. Some human sensory modalities relevant to human-computer interaction.

Modality	Examples
Visual	Gaze
	Image capture
Auditory	Voice input (as part of the language faculty)
	Nonspeech audio output (as part of the music faculty)
Haptic/kinesthetic	
Touch	Pressure
	Texture
Hand movement	Sign language
	3D motions
	Writing motions
	Drawing or other nontextual gestures
Head movement	Lip reading, oral cavity
	Head movement for pointing
	Facial expressions
Other body movement	Movement captured in dance or sports

2. Multimodal – Was bedeutet das?

Logische Kombinationen von Modalitäten:

- Auditory (Voice Input) & Touch (Pressure)

➡ Navigation im Auto

- Auditory (Nonspeech) & Touch (Texture)
- Body Movement & Auditory (Voice Input)

3. Die Multimodale Suche

Definition

Kombinantion natürlicher Eingabemodalitäten

Multimodales System

Multimedialer Output

Die Multimodale Suche

Abgrenzung

1. Suchen mit Hilfe multimodaler Interaktion

2. Analyse mit Hilfe multimodaler Interaktion

Suchen mit Hilfe multimodaler Interaktion

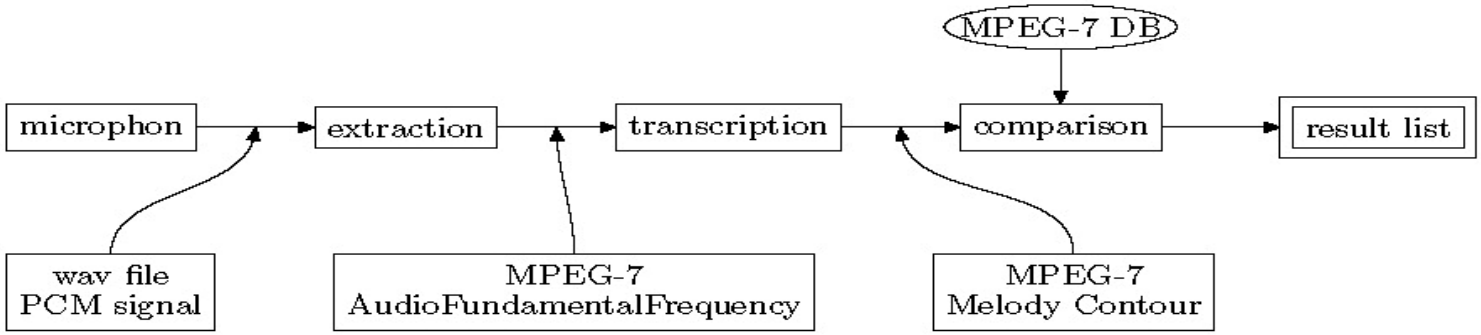
QBH – Query By Humming

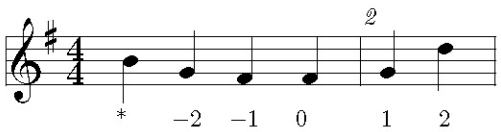
(Audio Engineering Society AES & Fraunhofer Institut)

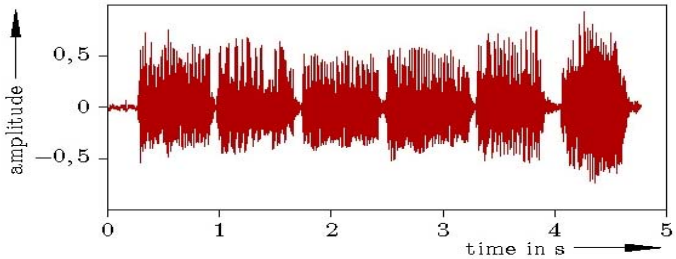
➡ Suche von Liedern durch Summen/Pfeifen über im System abgelegte Metadaten (MPEG-7 Deskriptoren)

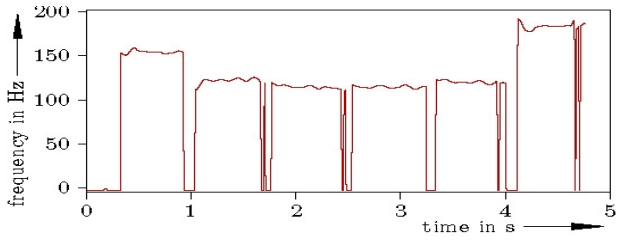
- Input über Mikrofon
- Geeignete Umwandlung / Darstellung
- Vergleich mit den im System befindlichen MPEG-7 Deskriptoren zum Auffinden ähnlicher Lieder

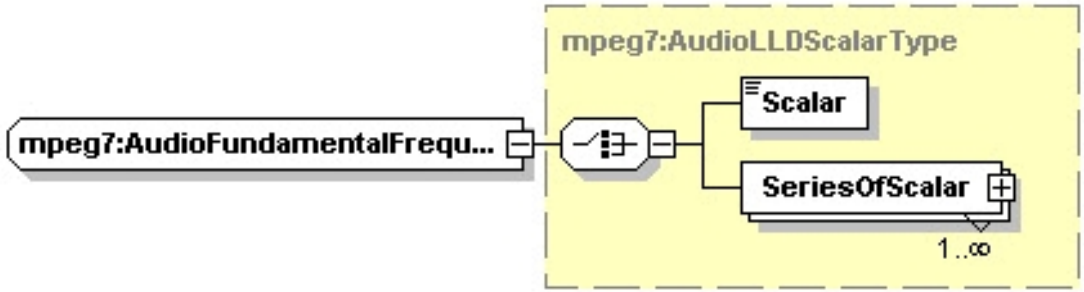
QBH – Query By Humming





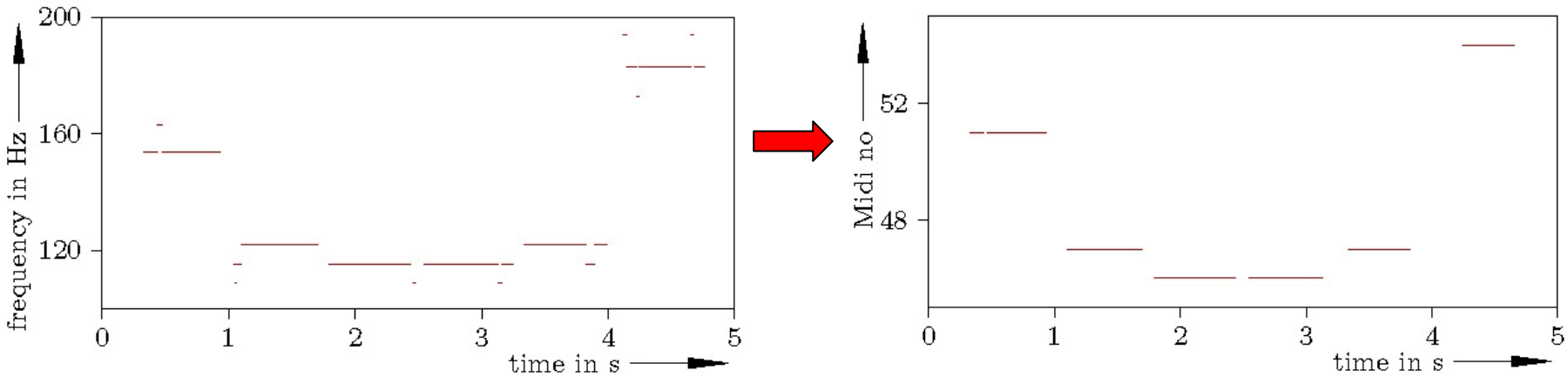






QBH – Query By Humming

Umschreibung in eine zum Vergleich geeignete Darstellung



Berechnung der Notenwerte und ihrer Konturwerte

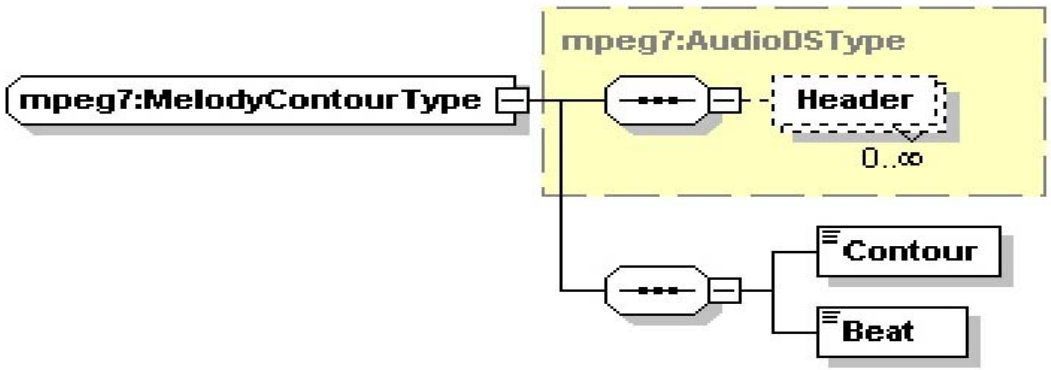
Noten chrom. Skala: $f(n) = f_0 * 2^{n/12}$

Abweichende Töne: $c(f) = 1200 * \log_2 * (f/f1)$

$|c(f)| > 50$ c => neuer Ton

Event > 80 ms => neuer Ton

Contour value	Change of c(f) in cents
-2	$c \leq -250$
-1	$-50 \leq c < -250$
0	$-50 < c < 50$
1	$50 \leq c < 250$
2	$c \geq 250$



Analyse mit Hilfe multimodaler Interaktion

Interactive Storytelling

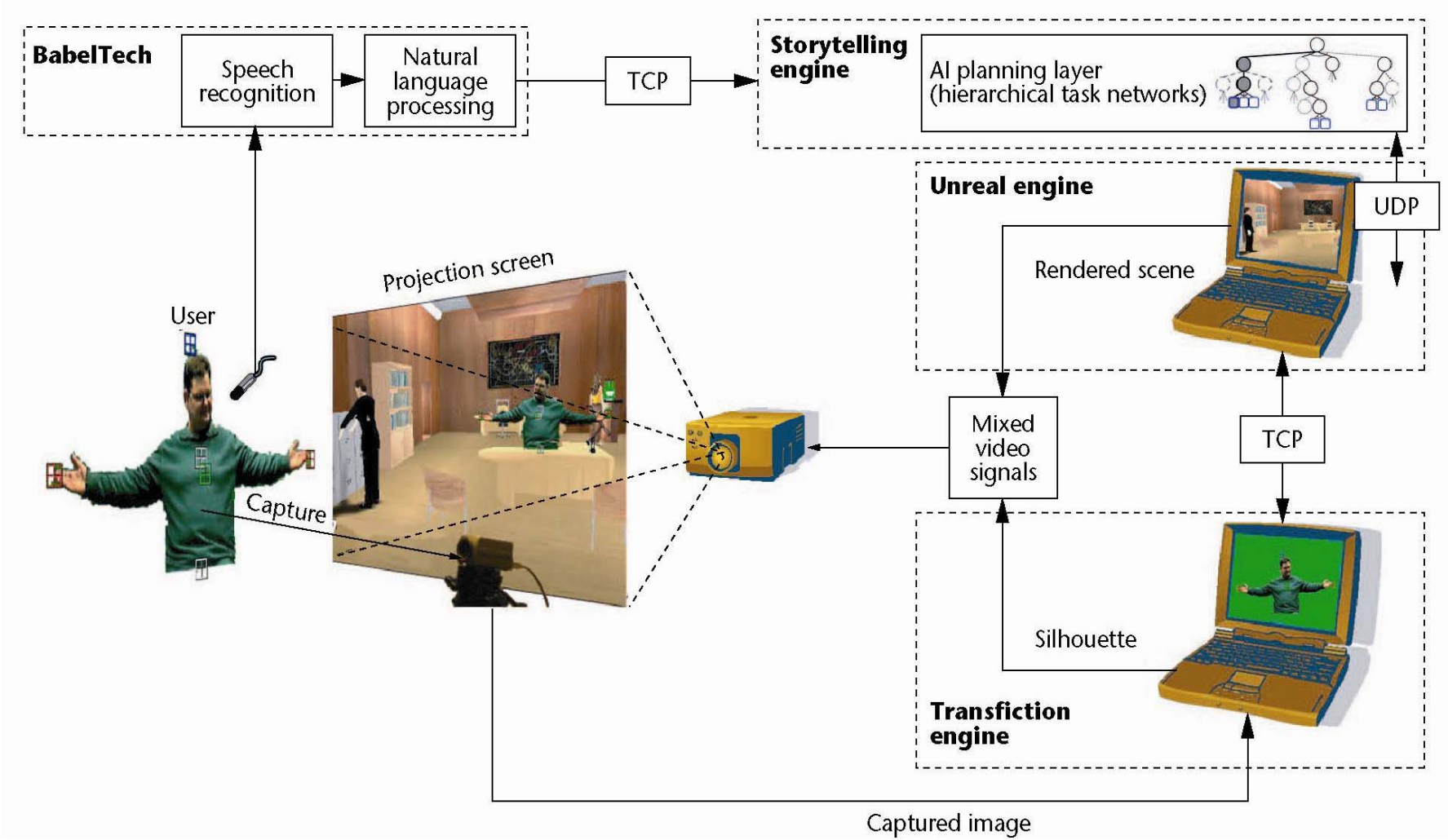
(Universität Teesside/UK & Zentrum f. graph. Datenverarbeitung)

Idee:

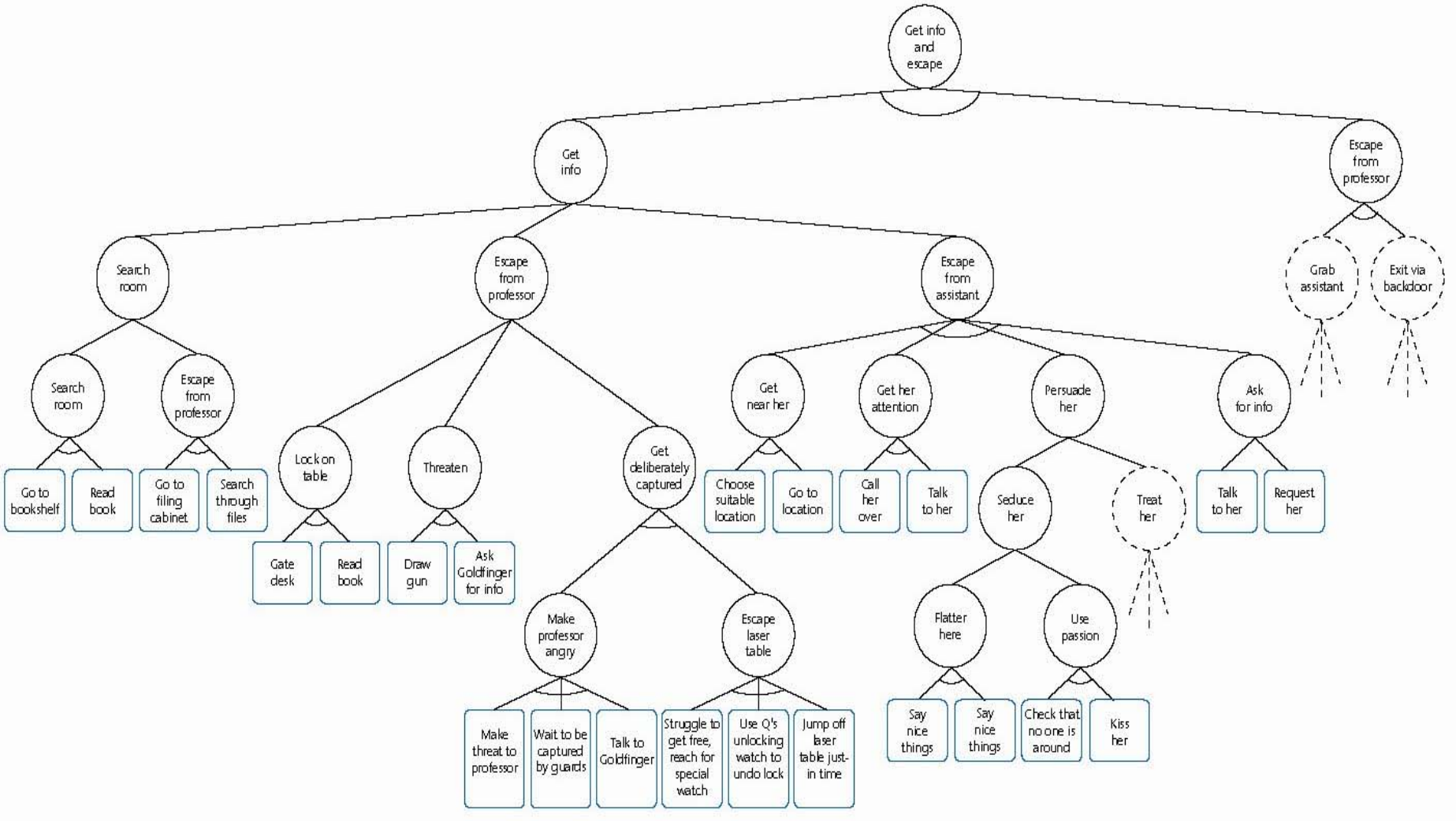
Außenstehender User wird in virtuelle Welt projeziert und interagiert mit KI

KI interpretiert Gesten & Sprache **(multimodaler Input)** des Users und kann darauf reagieren

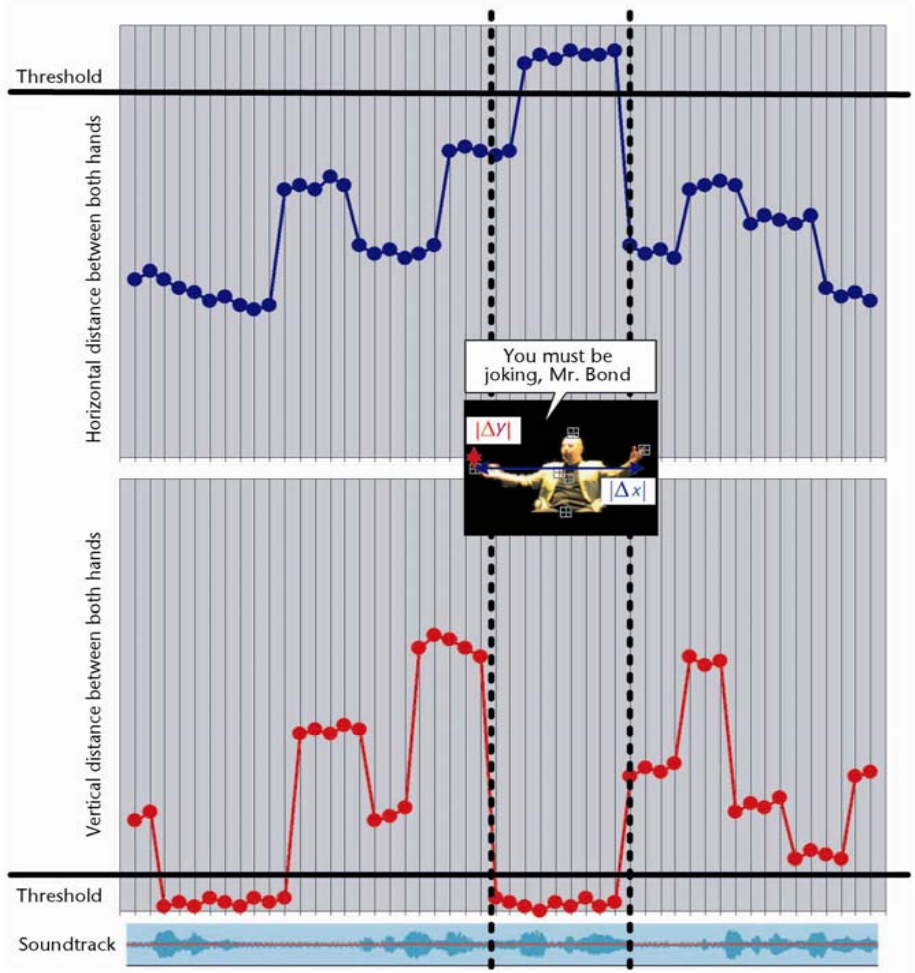
Das Gesamtsystem



Der Entscheidungsbaum für die Storytelling Engine



Die Gestenerkennung anhand der Silhouette



Gestenerkennung

Transfiction Engine **TCP (Koordinaten der Silhouette)** → System

Gestenlexikon & gelieferte Koordinaten → mögliche Interpretationen

Bei mehrdeutigen Gesten → alle Interpretationen werden assoziiert

Spracherkennung

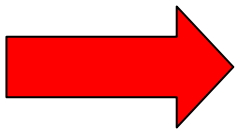
Spracherkennungssoftware zum Ausfiltern der Hauptaussagen

Schlüsselwörter (W-Wörter) & Verben-Erkennung

Gestenerkennung

&

Spracherkennung



Herausfiltern der Hauptaussagen

4. Beispiele im WEB

http://musicline.de/de/melodiesuche/input

http://www.melodyhound.com/

http://www.zgdv.de/zgdv/zgdv/departments/z5/Z5Projects

5. Fazit & Ausblick

Pro:

- Gute Ansätze und vor allem Interesse der Nutzer vorhanden

Contra:

- Systeme sehr aufwendig und noch unzureichend entwickelt
  - MPEG-7 & Kontexterkennung zu schwach
  - Systeme noch zu unflexibel, User zu sehr eingeschränkt