



*Chorus: Status and Challenges of  
MMSE Technology  
Joachim Köhler (Fraunhofer IAIS)*

Brussels, 26/27 May 2009



# AGENDA

---

- Landscape of MMSE projects
- Functional Description of a MMSE
- Content Enrichment Methods (Creating metadata)
  - Speech
  - Music
  - Images
  - Video, 3D (talk by Oliver Schreer)
  - Multimodal Analysis
- Conclusion

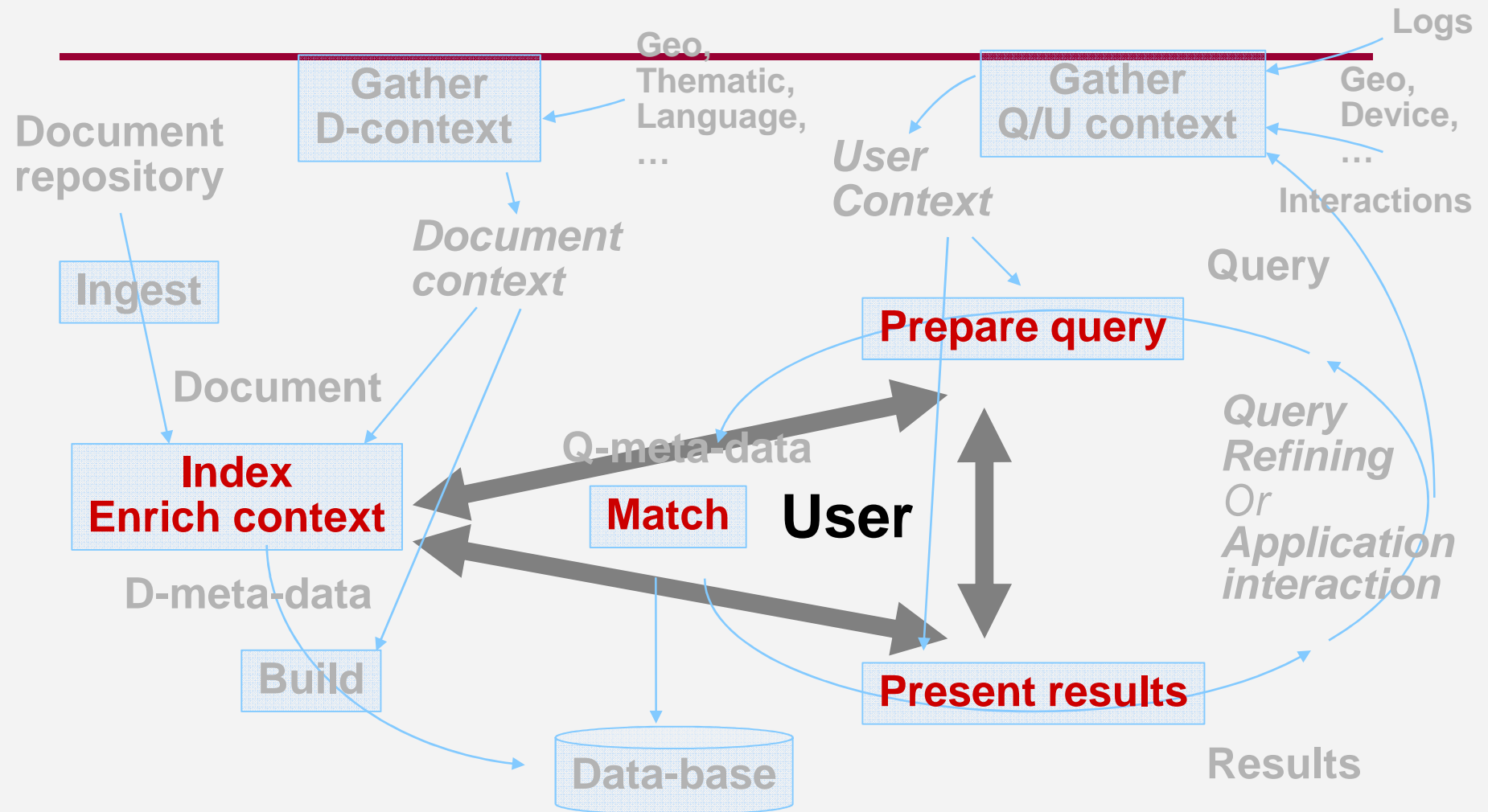
## *Overview about the AV indexing activities in the 9 IST projects*

	Speech/Audio	Image	3D	Video	Text/Semantics
<b>DIVAS</b>	FhG IDMT Sail Labs			Elecard ICCS	
<b>PHAROS</b>	Univ. P. Fabra FhG IDMT Sail Labs	EPFL		EPFL Open Univ., UK	Web Models L3S Research
<b>RUSHES</b>	Brunel Univ.	Brunel Univ.	FhG HHI	Queen Mary Univ. Brunel Univ.	Queen Mary Univ. Brunel Univ.
<b>SAPIR</b>	IBM Univ. of Padova	CNR		Eurix	Xerox
<b>SEMEDIA</b>				Joaneum Research Fundacio Barcelona Univ. P. Fabra UPC Barcelona Digital Video Systems Univ. of Glasgow	
<b>TRIPOD</b>		Dublin City Univ.			Sheffield Univ.
<b>VICTORY</b>			Certh/ITI		
<b>VIDI-VIDEO</b>	INESC Lisboa	U. Surrey UvA ITI U. Florence		UvA ITI U. Florence	
<b>VITALAS</b>	FhG IAIS	INRIA Robotiker		INRIA CWI Certh/ITI	Univ. of Sunderland EADS

## *Overview about the AV indexing activities in national research projects*

	Speech/Audio	Image	3D	Video	Text/Semantics
<b>Quaero (French)</b>	Limisi RWTH Aachen Univ. Karlsruhe VecSys IRCAM	INRIA Univ. J. Fourier Jouve		INRIA LTU Univ. J. Fourier	Jouve Limsi INRIA
<b>Theseus (German)</b>	FhG IAIS M2Any	FhG HHI FhG First Siemens CT	FhG HHI FhG IGD	FhG HHI Siemens	Univ. Karlsruhe FhG IAIS DFKI FZI
<b>iAD (Norway)</b>				Dublin Univ.	Fast/Microsoft
<b>MultimediaN (Dutch)</b>	U. Twente TU Delft	CWI U. Amsterdam		U. Amsterdam CWI TU Delft Philips	U. Twente
<b>SECO (Italian)</b>					DEI (Univ. Milano)
<b>Internet Score Cards (Swiss)</b>					IFAAR

# Functional Description of a MMSE



# Content Enrichment Methods: State-of-the-Art: Speech Analysis

---

- Approaches/Technologies
  - Speech recognition: HMM based LVCSR systems, Spoken Document Retrieval, Subword indexing (**SAPIR**, **VITALAS**, **PHAROS**, **Quaero**, **Theseus**, **MultimediaN**)
  - Speech Segmentation: speaker clustering and recognition (**DIVAS**, **VITALAS**, **Quaero**, **Theseus**, **MultimediaN**)
  - Speech-to-video transcoding (**DIVAS**)
- Systems
  - IST AV-projects: IBM speech system (**SAPIR**), Audiomining System from Fraunhofer IAIS (**VITALAS**), Sail Labs Technology (**DIVAS**), Voxlead from Limsi & Vecsys (**Quaero**)
  - Others: BBN, HTK-Group Cambridge, RWTH Aachen, Nuance, etc.
- Applications
  - Indexing of broadcast news/archives (**VITALAS**, **DIVAS**, **VIDIVIDEO**, **Quaero**, **Theseus**, **MultimediaN**)
  - Podcast/Videocast search (Potzinger, Blinkx)
  - Audio archives (Parliament data, historical archives)

# Best Practice: Speech Search

Suchen | Browsing | Hilfe

Suche   « Zurück zur Übersicht  
Asset 8 von 67: 080213\_h19\_p  
[vorheriges](#) [nächstes](#) Fraunhofer Institut Intelligente Informationssysteme

00:00 18:11

**Treffer**

<b>Treffer 1</b> Start: 02:45 Match: 0% 	... Barack <b>obama</b> hat bei den amerikanischen vorwahl auch diesen Dienstag abgeräumt ...
<b>Treffer 2</b> Start: 02:59 Match: 0% 	... kann ein republikaner seine führungsrolle weiter ausbauen damit für <b>obama</b> bei den demokraten im rennen um die präsidentchaftskandidatur im ...
<b>Treffer 3</b> Start: 03:09 Match: 0% 	... bundesstaaten in folge hatte gewonnen zuletzt mit klaren vrsprung hat <b>obama</b> zurzeit auf der siegerstraße auftritt in wisconsin wohl nächste ...
<b>Treffer 4</b> Start: 03:26 Match: 0% 	... die zeit und auch junge wähler und schweizer haben für <b>obama</b> gestimmt und auch bei den frauen hatte diesmal gewann ...
<b>Treffer 5</b> Start: 03:31 Match: 0% 	... und auch bei den frauen hatte diesmal gewann erstmals hat <b>obama</b> einen deutlichen vrsprung bei den wahl delegierten wendepunkt in ...
<b>Treffer 6</b> Start: 04:21 Match: 0% 	... republikanern für sich entschieden und ist eben wahrscheinlicher kandidat dagegen <b>obama</b> oder clinton und mit einer weiter ...

[Alle Informationen anzeigen](#)

© 2009 Flowplayer Ltd

# Advanced Content Enrichment Methods: Speech

## Applied Technology

- Statistical based speech recognition (Hidden-Markov-Models)
- Speech segmentation
- Speaker recognition (Gaussian Mixture Models)

## Trends

- Hybrid speech retrieval systems (combination of word and subwords, like phonemes and syllables), indexing of lattices
- Robustness (Background noise, double talk, etc.)

## Challenges

- Variety of languages (multilingual applications)
- Lexicon dependency of existing ASR systems
- Domain dependency (broadcast news versus meetings)
- Very complex audio sounds (composition of different sounds)

## Promising Directions

- audio modelling and mining (i.e. using of anchor models)
- phonetic or syllable based approaches, new forms of context dependency



# State-of-the-Art: Music Analysis

---

- Approaches/Technologies
  - Music segmentation: Spectral Flatness (MPEG-7 Audio), Genetic Algorithms (**DIVAS**, **PHAROS**, **Quaero**, **Theseus**)
  - Music retrieval and Recommendation (SOMs) (**SAPIR**, **Theseus**)
- Systems
  - IST projects: Fraunhofer IDMT (**DIVAS**, **PHAROS**), M2Any (**Theseus**), IRCAM (**Quaero**)
  - Others: Barcelona Music & Audio Technologies, FhG AudioID, PlaySom (Univ. Vienna), SyncPlayer (Univ. Bonn), etc.
- Applications
  - Indexing of music collections
  - Query by humming
  - Audio-music identification
  - Recommendation engines

## *Best Practice: Music Search*

- Jointly develop AudioID-Sonos Technology - a generic framework for content-based semantic annotations and recommendations (FhG IDMT)
- SDK and business solutions available from Mufin GmbH



# Advanced Content Enrichment Methods: Music

## Applied Technology

- Low level feature extraction:
  - Appl. 1: audio fingerprinting (find identical audio segments)
  - Appl. 2: audio matching (find similar audio segments)

## Trends

- Towards music semantics (users and music scientists):
  - genre classification, artist recognition, mood detection, automatic playlist generation, music recommendation
  - music summarization, rhythm and melody extraction

## Challenges

- High complexity of music signal (e.g. polyphonic music)
- Strong subjective component for mood and genre classification

## Promising Directions

- Moving from controlled audio situations to more complex situations
- Robust extraction of high-level information
- Machine Learning algorithms for time-series data

# State-of-the-Art: Image Analysis

- Approaches/Technologies
  - Low level image processing (histograms, shapes, textures, MPEG7-visual, SIFT) (**SAPIR**, **VIDIVIDEO**, **VITALAS**, **SMEDIA**, **TRIPOD**, **Rushes**, **Quaero**, **Theseus**, **MultimediaN**, **IM2**)
  - Image similarity measurements (**Rushes**, **VIDIVIDEO**, **VITALAS**, **LIVE Theseus**, **IM2**)
  - Relevance Feedback (**Rushes**, **SMEDIA**, **VITALAS**), etc.
- Systems
  - IST projects: INRIA (**VITALAS**), Univ. of Amsterdam & Univ. of Florence (**VIDIVIDEO**), etc.
  - Others: IBM (QBIC), Webseek, MPEG-7 search system (Univ. Munich), IKONA (INRIA), etc.
- Applications
  - Content based retrieval in image collections
  - Object recognition, Face recognition (security, photo collections)
  - Automatic annotation of image collections with keywords and textual descriptions

# Best Practice: Image Search

Vitalas - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://rcc.robotiker.es:8080/vitalas/ Google

WebLab Gmail - Boîte de récepti... Mailing list archives Serveur de developpe... www.weblab-project.org

method=fusedSearch: query=championship AND textconcept:belgian AND visualconcept:car AND success:1:1  
similaritySearch:weblab://formatted/07591779

## VITALAS

Video & image indexing and reTrievAl in the LArge Scale

login | preferences | about | my stored searches

championship

reset search  
advanced visualization

from [ ] to [ ]

visual concept: car

textual concept: belgian

FFFFFF

0-15 of 15 results



The screenshot shows a web browser window displaying the Vitalas search interface. The search query is 'championship'. The visual concept is set to 'car' and the textual concept is 'belgian'. A color wheel is visible, and a small thumbnail of a rally car is shown. Below the search area, a grid of 15 search results is displayed, all showing a yellow and white rally car on a dirt road. The browser window title is 'Vitalas - Mozilla Firefox' and the address bar shows 'http://rcc.robotiker.es:8080/vitalas/'. The browser's menu bar includes 'Fichier', 'Édition', 'Affichage', 'Historique', 'Marque-pages', and 'Outils'. The browser's toolbar includes navigation buttons and a search engine dropdown set to 'Google'. The browser's tab bar shows several open tabs, including 'WebLab', 'Gmail - Boîte de récepti...', 'Mailing list archives', 'Serveur de developpe...', and 'www.weblab-project.org'. The Vitalas logo is prominently displayed in the center of the page, with the tagline 'Video & image indexing and reTrievAl in the LArge Scale' below it. The search results are displayed in a grid format, with 15 results shown. The first row contains 8 images and the second row contains 7 images. All images show a yellow and white rally car driving on a dirt road. The browser window also shows a status bar at the bottom with a pink flower icon.

# Advanced Content Enrichment Methods: *Image*

## Applied Technology

- Content based image retrieval
- Robust feature extraction (e.g. SIFT)

## Trends

- Semantic search and concept-based detection in complex background
- Interactive search and agent interfaces
- New learning models (e.g. neuro-based learning )
- Multi-user environments (towards a common query model)

## Challenges

- Semantic gap (low level versus high level features and models)
- Image segmentation (only solved for restricted domains)
- Interpretation: this refers to the user model and context

## Promising Directions

- New machine learning algorithms based on human perception and cognition
- High-performance computing
- Folksonomies

# *State-of-the-Art: Video Analysis*

---

- Approaches/Technologies
  - Shot detection, keyframe generation (DIVAS, Rushes, SAPIR, VIDIVIDEO, VITALAS, Quaero, Theseus, MultimediaN, IM2)
  - Object tracking based on motion based features, closed captions recognition, etc. (Rushes, VIDIVIDEO, VITALAS, Quaero, Theseus, MultimediaN, IM2)
  - Object detection and recognition (ANN, Adaboost, SIFT) (VIDIVIDEO, VITALAS, SMEDIA, VITALAS, Quaero, Theseus, MultimediaN, IM2)
  - Video annotation and summarization (Rushes, SMEDIA, VITALAS, Quaero, Theseus, MultimediaN, IM2)
  - Video event detection (SMEDIA, VITALAS, VIDIVIDEO)
- Systems
  - IST projects: Univ. Amsterdam & Univ. Florence (VIDIVIDEO), Joaneum Research (SMEDIA), CERTH/ITI (VICTORY), VITALAS (INA/INRIA, CERTH-ITI), Fraunhofer (Theseus),
  - Others: Virage, TrecVideo-participants, Informedia, Univ. of Marburg, etc.
- Applications
  - Indexing of broadcast material, media observation, Indexing of videocast material
  - Recommendation Engines, Video fingerprinting, logo detection, security, etc.
  - 3D video (Rushes, VICTORY, Theseus)

# Advanced Content Enrichment Methods: Video

## Applied Technology

- Concept detection based on low level features (TrecVid)
- Shot based classification
- Combining features using SVM or other classifiers

## Trends

- Incremental improvements:
  - from SVM to more sophisticated kernel methods
  - more robust and efficient features such as keypoint/codebook features
- Inclusion of temporal dimension (using the dynamics in videos)

## Challenges

- Broad-domain applicability
- Lack of training data

## Promising Directions

- Leveraging social tagged media as substitute for training data
- High-performance computing
- Video behavior analysis



# Multi- and Cross Modal Search (Univ. of Marburg: Videana)



Videana Demo  
Semantic Video Search

Concept: Meeting Search

- Flag
- Flag Usa
- Food
- Football
- Golf
- Government Building
- Government Leader
- Graphics
- Grass
- Hassan Nasrallah
- Horse
- Horse Racing
- House
- Hu Jintao
- Indoor
- Kerry
- Lahoud
- Male
- Maps
- Meeting



## *Multi- and Crossmodal MMSE Challenges and Limitations*

---

- Amount and quality of data and annotations
- Availability of keyword vocabulary
- Low-level representations of information
- Quality of machine learning algorithms
- Lack of bootstrapping mechanism
- Context information to reduce the domain space

## Conclusion

---

- More efficient and robust multimedia indexing techniques (especially for “object” based search)
- New media search paradigms (from keyword based search to event/context based search)
- Modeling implicit and explicit feedback to empower personalization and recommendations
- Exploiting user annotations (weakly labeled data)
- More informative user interfaces for MMSE (exploration and experienced based)
- Scalability (algorithms and network infrastructure)

*Thanks for your attention*

<http://www.ist-chorus.org/>

More Information:  
Deliverable D2.2