SignSpeak
Bridging The Gap Between Signers and Speakers

Philippe Dreuw, and Hermann Ney
dreuw@cs.rwth-aachen.de

ESSV – Sep 2009

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

~~Outline~~

Introduction

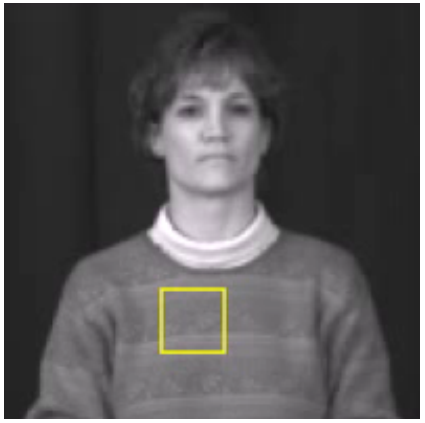
- **New trend in sign language research**
  - **advances of computer technology enabling the easy use of digital video**
  - **continuous spread of Internet**
  - **public interest**
  - **allows for integration of NLP, ASR, and CV research**

- **SignSpeak project (EU funded STREP project)**
  - **better linguistic knowledge of sign languages**
  - **vision-based technologies for sign language processing**
  - **automatic sign language recognition**
  - **automatic sign language translation**

→ Provide new e-Services to the deaf community

Application: Speech-to-Speech

Recognition: Speech-to-Text (Video → Glosses)



⇓

Translation: Text-to-Text (Glosses → Text)

`JOHN FISH WONT EAT BUT CAN EAT CHICKEN`

John will not eat fish but eats chicken

⇓

Synthesis: Text-to-Speech (Text → Audio)

`021.wav`

Sign Languages in Europe

- **Green - Recognised in constitutional level**
- **Orange - Recognised their national sign language by other legal measures**
- **Red - Not recognised at all**

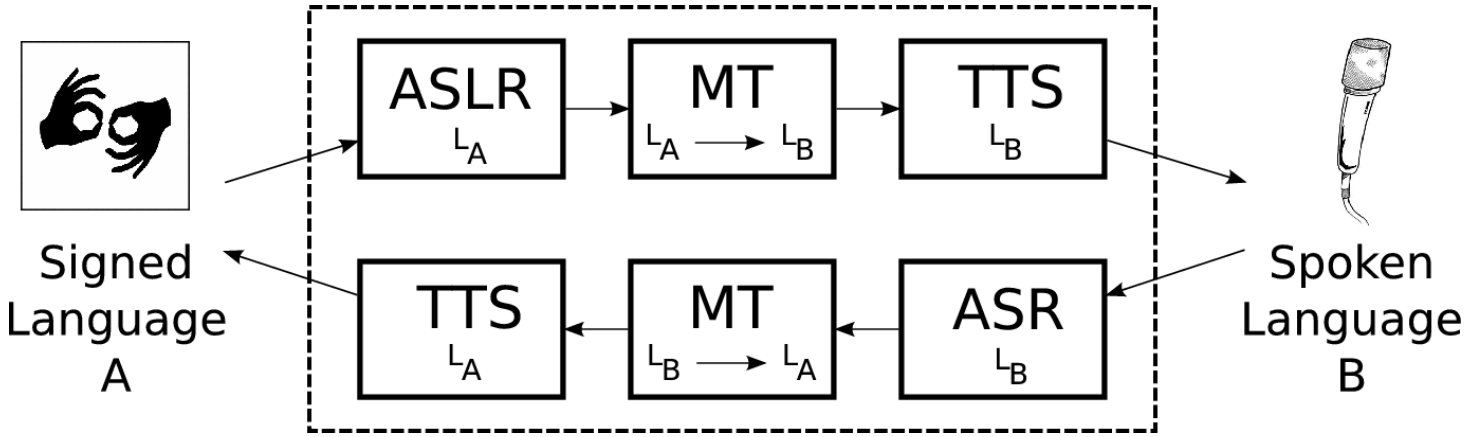
Sign Languages in Europe

- **European Union of the Deaf (EUD)**
  - **non-research partner in SignSpeak**
  - **about 7,000 official Sign Language Interpreters**
  - **estimated about 650,000 Sign Language users in Europe (EUD Survey, 2008)**

  → **the number of sign language users might be much higher!**

- **European Parliament - 7th June 2009 - Ádám Kósa (HU)**
  - **first ever deaf person and sign language user was elected as an MEP**

SignSpeak: Research and Challenges



- **Six components-engine necessary to build a Sign-To-Speech system**

- **SignSpeak** `http://www.signspeak.eu`
  - **linguistic research in sign languages**
  - **environment conditions and feature extraction**
  - **modeling of the signs**
  - **statistical machine translation of sign languages**
  - **languages and available resources**

Linguistic Research in Sign Languages

- Linguistic research on sign languages started in the 1950 (Tervoort et al., Stokoe et al.)
- Recognition of sign languages as an important linguistic research object
  - 1970, USA
  - 1980, Europe
  - since 1990, worldwide

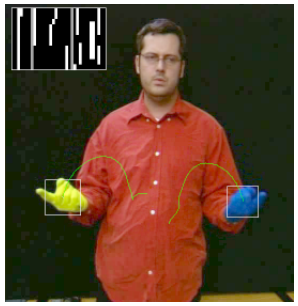
  $\longrightarrow$ 2004, foundation of the Sign Language Linguistics Society
- Vision-based linguistic research
  - small sets of elicited data (Corpora) recorded under lab conditions
  - often either too small and spontaneous, or too constrained

Sign Language Recognition

- **What features do we need?**
  - **manual components: hand motion / form / orientation / location**
  - **non-manual components: mimic, eye gaze, body / head orientation**

→ **should be extracted from input signal**

- **Different approaches / assumptions**
  - **special hardware**
  - **computer vision**

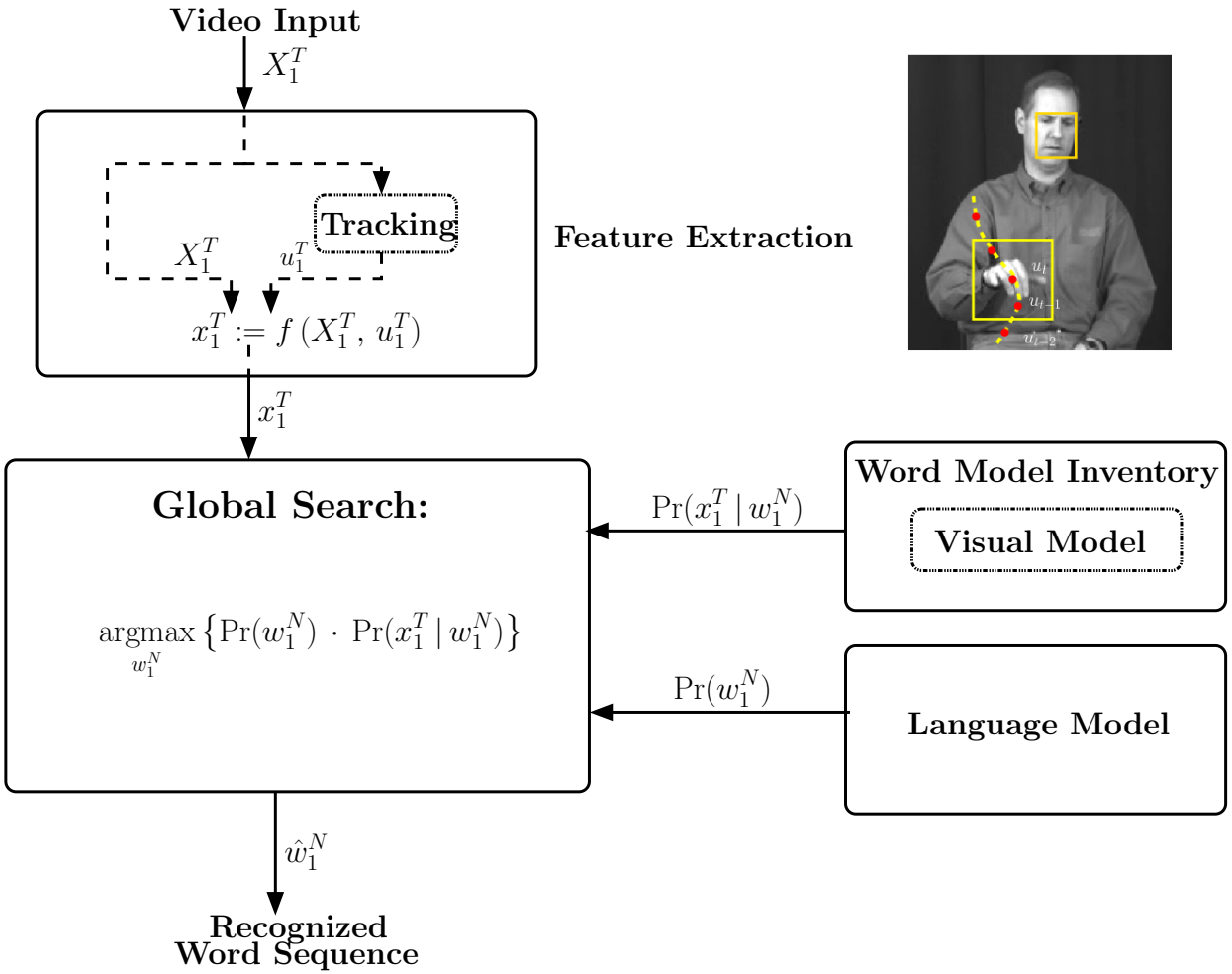




→ **only the vision-based approaches do not restrict the way of signing**

→ **different problems arise in feature extraction**

System Overview

- **Bayes' decision rule used in ASLR**
  - **with tracking framework and feature extraction as a pre-processing step.**



Speech and Sign Language Recognition

- **At least four crucial problems have to be solved in ASR/ASLR:**
  1. **preprocessing and feature extraction of the input signal,**
  2. **specification of models and structures for the words to be recognized,**
  3. **learning of the free model parameters from the training data, and**
  4. **search the maximum probability over all models during recognition.**

- **Similarities**
  - **temporal sequence of sounds or gestures**
  - **languages and dialects**

- **Main Differences Between Signed and Spoken Languages**
  - **simultaneousness**
  - **signing space**
  - **3D coarticulation and movement epenthesis**
  - **silence**

Automatic Sign Language Recognition

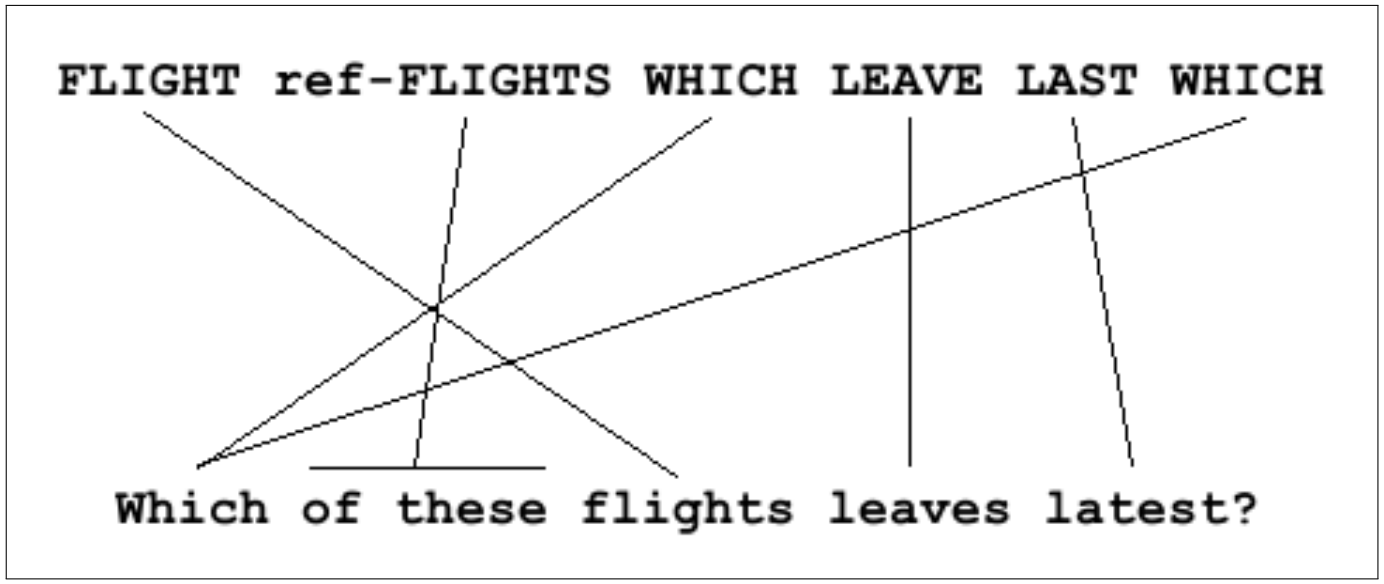
- **Problems in current SOTA approaches:**
  - capturing, tracking, segmentation, ...
  - most systems: very person dependent, recognition of isolated signs
  - modeling of the signs
    - no suitable decomposition of words into sub-word units
    - co-articulation effects (3D) and dialects not modeled
  - lack of data, no publicly available corpora

- **SignSpeak approach/setup: similar to speech recognition**
  - recognition of continuous sign language
  - training with sentences (unknown word boundaries)
  - person independent training and recognition
  - focus on sub-word unit modeling
  - large datasets, will be publicly available

→ use RWTH-ASR large vocabulary speech recognition system

Sign Language Translation

- statistical machine translation requires
  - better linguistic knowledge for phrase-based modeling and alignment
  - large bilingual annotated corpora
- challenges
  - reorderings
  - references in signing space



Available Resources within SignSpeak

- **Corpus NGT** `http://www.corpusngt.nl`
  - core of the SignSpeak data
  - 72 hrs, Sign Language of the Netherlands
  - first large open access corpus for sign linguistics in the world
  - 92 different signers

- **RWTH-Phoenix**
  - 1.5 hrs of German Sign Language
  - weather forecast news, 1k vocabulary size, 10k words
  - 11 signers

- **Other:**
  - RWTH-BOSTON: American Sign Language
  - ATIS: Irish Sign Language
  - SIGNUM: German Sign Language

How Should Sign Language Corpora be Created?

- Existing corpora must be extended to achieve good performance
  - recognition and translation
- Important: new annotations should be more domain specific
  - vocabulary size $<$ 4k
  - Token/Types ratio $\approx$ 20
  - singleton ratio $<$ 40%
- Comparison to limited domain tasks Vermobil II (ASR) and IWSLT (SMT)

	Boston104	Phoenix		Corpus-NGT		Vermobil II	IWSLT
year	2007	2009	2011	2009	2011	2000	2006
recordings	201	78	400	116	300	-	-
running words	0.8k	10k	50k	30k	80k	700k	200k
vocabulary size	0.1k	0.6k	2.5k ?	3k	$>$ 5k ?	10k	10k
T/T ratio	8	15	20 ?	10	$<$ 20 ?	70	20
Performance	10% WER					15% WER	40% TER

Application Scenarios

- **Sign Language**
  - Telefónica I+D, industrial partner in SignSpeak
  - interested in the basic research for possible exploitation
    - communication platform
    - e-learning
    - automatic transcription of video e-mails

- **Automotive**
  - intersection assistant - head pose estimation
  - fatigue detection - eye gaze estimation
  - smart airbags - upper body tracking

- **Games**
- **Medical Sector**
- **Surveillance**

Thank you for your attention

Philippe Dreuw

dreuw@cs.rwth-aachen.de

http://www.hltpr.rwth-aachen.de/