FP7-ICT-2011-7 - Language technologies (STREP)
Duration: 36 months (1 Nov 2011 – 31 Oct 2014)

<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>Short</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UNIVERSITAT POLITECNICA DE VALENCIA</td>
<td>UPVLC</td>
<td>Spain</td>
</tr>
<tr>
<td>2</td>
<td>XEROX SAS</td>
<td>XEROX</td>
<td>France</td>
</tr>
<tr>
<td>3</td>
<td>INSTITUT JOZEF STEFAN</td>
<td>JSI</td>
<td>Slovenia</td>
</tr>
<tr>
<td>3+</td>
<td>KNOWLEDGE FOR ALL FOUNDATION</td>
<td>K4A</td>
<td>UK</td>
</tr>
<tr>
<td>4</td>
<td>RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN</td>
<td>RWTH</td>
<td>Germany</td>
</tr>
<tr>
<td>5</td>
<td>EUROPEAN MEDIA LABORATORY GMBH</td>
<td>EML</td>
<td>Germany</td>
</tr>
<tr>
<td>6</td>
<td>Deluxe Digital Studios Ltd</td>
<td>DDS</td>
<td>UK</td>
</tr>
</tbody>
</table>
Motivation

• Large on-line repositories of video lectures are being established: VideoLectures.NET, poliMedia, OCW and Opencast Matterhorn-based repositories, etc.

• Video lectures are neither transcribed nor translated due the lack of cost-effective tools.

• Transcriptions and translations are needed to make them accessible to speakers of different languages and people with disabilities. They would also facilitate further search and analysis functions.

• Starting hypothesis: current ASR and MT techniques are not far from achieving acceptable results in educational repositories.
Motivation: applications of transcription

• Accessibility (hearing impairment) or quiet environments
• Translation to any other languages
• Searchability
• Recommendation
• Summarisation
• Video segmentation
• Discovery of content relations
Goal and technology challenges

Goal:
To develop innovative, cost-effective solutions to produce accurate transcriptions and translations in VideoLectures.NET, with generality across other Matterhorn-related repositories.

Technology challenges:
• Improvement of transcription and translation quality by massive adaptation.
• Improvement of transcription and translation quality by intelligent interaction.
• Integration into Matterhorn to enable real-life evaluation.
Outline of the talk

• Demo videos: project overview, intelligent interaction and Matterhorn integration
• Transcriptions and translations at the pilots
• Massive adaptation
• Intelligent interaction with users
• Integration and evaluation
• Try our tools and transcribe your videos
Demo videos

• Project overview
  https://www.translectures.eu/project-summary

• Intelligent interaction with users
  https://www.translectures.eu/web/progress

• Matterhorn integration
  https://www.translectures.eu/web/progress
## Transcriptions and translations at the pilots

<table>
<thead>
<tr>
<th></th>
<th>Language</th>
<th>Nb. of lectures</th>
<th>Nb. of hours</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VideoLectures.NET</strong></td>
<td>English</td>
<td>9148</td>
<td>5900</td>
</tr>
<tr>
<td></td>
<td>Slovenian</td>
<td>741</td>
<td>490</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>9889</td>
<td>6390</td>
</tr>
<tr>
<td><strong>poliMedia</strong> (live)</td>
<td>Spanish</td>
<td>7554</td>
<td>996</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>420</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Catalan</td>
<td>91</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>8065</td>
<td>1069</td>
</tr>
</tbody>
</table>

**VideoLectures.NET:** EnEs, EnFr, EnDe, EnSl, SlEn  
**poliMedia:** EsEn
Massive adaptation

• Massive adaptation of acoustic models
  – Adaptation by neural networks
  – Adaptation by CMLLR and MLLR
• Massive adaptation of language models
  – Adaptation by interpolation
  – Adaptation by lecture slides
  – Adaptation by lecture-dependent in-domain document retrieval
• Massive adaptation of translation models
  – Adaptation by data selection

More on massive adaptation in Muhammed’s talk
Intelligent interaction with users

• Fast constrained search
• Intelligent interaction for transcription
  – Confidence estimation
  – User interaction units
• Intelligent interaction for translation
  – Intelligent interaction at sentence level
  – Intelligent interaction at word level
• Incremental training for translation
Transcription of 23 videos (3h) with no interaction
Transcription of 23 videos (3h) with no interaction

Supervised minutes
WER
18.8
Supervised minutes
WER
18.4 BI
Transcription of 23 videos (3h) with batch interaction

Supervised minutes
WER
18.0 BI
Transcription of 23 videos (3h) with **batch interaction**

![Diagram showing the transcription process and a graph of WER over supervised minutes with a value of 17.6 BI.]
Transcription of 23 videos (3h) with **batch interaction**

![Diagram showing transcription process with batch interaction](image_url)
Transcription of 23 videos (3h) with \textit{batch interaction}

![Graph showing the decrease of WER with increasing supervised minutes with a value of 16.8 BI.]
Transcription of 23 videos (3h) with batch interaction

Supervised minutes
WER
16.8 BI
Transcription of 23 videos (3h) with intelligent interaction

1. WER 16.5 BI
2. WER 18.5 II

Supervised minutes
Transcription of 23 videos (3h) with **intelligent interaction**

![Diagram of transcription process with speech recognition and intelligent interaction](image-url)
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
18.0 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
17.5 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
17.3 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
17.3 II
Transcription of 23 videos (3h) with intelligent interaction

![Diagram showing transcription process]

- Supervised minutes vs. WER
- WER values: 16.5 BI, 17.2 II

Graph showing the transcription process with movie icons and progress tracking.
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI 17.0 II
Transcription of 23 videos (3h) with intelligent interaction

Graph showing the decrease in WER (Word Error Rate) as the number of supervised minutes increases.

Supervised minutes vs. WER

WER

0  3  6  9  12  15  18

16.5 BI 16.5 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER

16.5 BI
16.4 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
16.0 II
Transcription of 23 videos (3h) with intelligent interaction

![Graph showing WER vs Supervised minutes]

**Supervised minutes**: 18

**WER**: 16.5 BI, 15.7 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
15.7 II
Transcription of 23 videos (3h) with intelligent interaction
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
15.4 II
Transcription of 23 videos (3h) with **intelligent interaction**

![Graph showing WER improvement with supervised minutes]

- Supervised minutes: 0, 3, 6, 9, 12, 15, 18
- WER improvement: 16.5 BI, 15.2 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
15.1 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
14.5 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
14.5 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
14.3 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
14.3 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
14.0 II
Transcription of 23 videos (3h) with intelligent interaction

\[
\begin{array}{c|c|c|c}
\text{Supervised minutes} & \text{WER} & \text{BI} & \text{II} \\
\hline
0 & 16.5 & & 13.6 \\
3 & & & \\
6 & & & \\
9 & & & \\
12 & & & \\
15 & & & \\
18 & & & \\
\end{array}
\]
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
13.6 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
13.4 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
13.1 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
13.0 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
12.6 II
Transcription of 23 videos (3h) with intelligent interaction

Supervised minutes
WER
16.5 BI
12.6 II
Transcription of 23 videos (3h) with **intelligent interaction**
Integration and evaluation

• Integration and evaluation at Videolectures.NET: Matjaz’s talk

• Integration and evaluation at poliMedia: Carlos’ talk
Try our tools and transcribe your videos

On 13 March 2014, UPV opened a free service for public and private organisations to test transLectures technology. Eight organisations have already tried our tools:

- Univ. Carlos III de Madrid, Spain, 117 mins (Es)
- Intl. Centre for Theoretical Physics, Italy, 276 mins (En)
- SpeechWare, Belgium, 180 mins (Es)
- Centre for Dev. of Adv. Computing, India, 55 mins (En)
- UMH Occupational Observatory, Spain, 54 mins (Es)
- RWTH Aachen, Germany (En)
- UPV Linguistic Standardisation Service, Spain (Ca)
- Aseryla project, Spain (Es)
Try our tools and transcribe your videos

www.translectures.eu
EML offers research-based products, solutions and services for
- **Speech transcription** for media data (e.g. lectures, broadcast news), telephone and smartphone
- **Spoken interaction** for speech interfaces

contact: Dr. Marion Mast
email: marion.mast@eml.org
www: www.eml.org
Beyond transLectures

- Multilingual MOOC aggregator platform
- 12 universities and companies in 8 EU countries
- EMMA MOOCs will go live in September 2014
- MOOCs automatically transcribed and translated
- Languages transcribed
  - English, Spanish, French, Italian, Dutch, Portuguese and Estonian
- Language pairs translated
  - Spanish, French, Italian, Dutch, Portuguese and Estonian → English
  - English → Spanish and Italian
- EMMA open for providers wishing to make MOOCs available in multiple languages