Towards Web Services for Speech Recording and Annotation

Christoph Draxler
draxler@phonetik.uni-muenchen.de
BAS Bavarian Archive for Speech Signals
LMU Munich
BAS

- hosted by University of Munich (LMU)
- Florian Schiel and Christoph Draxler
  - corpora: Verbmobil, SmartKom, SpeechDat,…
  - tools: MAUS, Balloon, SpeechRecorder,…
  - services: consulting, validation, development,…
- spin-off company BAS-Services
Observations

- large speech and multimodal corpora
- incremental and continuous corpus growth
  - horizontal: new signal data
  - vertical: additional annotations
- traditional approaches do not scale up
Primary Data

- recorded signals
- immutable
- time-aligned
- multiple recording channels
- audio, video, sensor
- optimal technical quality
Secondary Data

- annotations of signals
- extensible, modifiable
- multi-tier annotations
- orthographic, phonemic, phonetic, prosodic, syntactic, semantic, discourse…
- complex inter-tier relationships
- derived data
Tertiary Data

- meta-data
- describes speech database
- resources
  - human resources, processing resources
- documentation
  - reports, charts, log files
- contracts
Issues

- distribution of expertise and workload
- interoperability
- long-term availability

- the project has terminated…
- the lead programmer found a permanent position… in Ouagadougou
- the data will outlive the people!
Requirements

- comprehensive data model for corpora
- extends beyond data and annotations
- includes workflow
- allows data migration
- independence and extensibility
- public data formats and APIs
Workflow and Tools

- **Corpus specification**
  - Balloon
  - XML-editors

- **Recording**
  - SpeechRecorder
  - audacity

- **Signal processing**
  - Emu/libassp
  - sox
  - htk
  - sfs

- **Annotation**
  - WebTranscribe
  - Emu
  - MAUS
  - Balloon
  - Praat
  - ELAN
  - ANVIL
  - EXMARaLDA
  - Transcriber
  - NITE Toolbox
  - ...

- **Processing**
  - Balloon
  - IMDI tools
  - XML-editors
  - PDF-editors
  - Web-editors
  - DB-tools

- **Exploitation**
  - Emu/R
  - MapApplets
  - COREX
  - SQL/R
  - Excel

Text, audio, video, sensor, derived signal data, signal processing, (XML-)text, XML, PDF, CSS, scripts.
WikiSpeech

- web-based client-server architecture
- clients
  - creation, annotation, exploitation
- server
  - access, storage, management
Workflow and Tools

corpus specification

recording

signal processing
derived signal data

annotation

processing

exploitation

text

audio video sensor

(XML-)text

XML, PDF CSS, scripts

Balloon SpeechRecorder Emu/libassp WebTranscribe Emu MAUS Balloon Praat ELAN ANVIL EXMARaLDA Transcriber NITE Toolbox …

XML-editors

PDF-editors

Web-editors

DB-tools

Emu/R MapApplets

WikiSpeech DB
Recording Client

Speech Database Server

- demographic data
- signal data
- prompt item
**Annotation Client**

- search criteria
- annotation data

**Speech Database Server**

- signal data
- prompt item
- existing annotations
- application resources
Exploitation Client

Speech Database Server

- text query
- map location
- signal data
- annotations
- meta-data
SpeechRecorder

• platform independent
• multi-media prompts
• multi-channel recordings, local and remote
• support for multiple views and displays
• localisable
• free
WebTranscribe

• extensible framework for web-based annotation
• supports “select - annotate - save” workflow
• platform independent
• plug-in architecture
• functionality implemented in editor modules
• multiple editors supported
Annotation workflow

annotate signal

annotation correct?

yes

no

annotate signal

annotation correct?

yes

no

next annotation?

yes

no

save annotation

your favourite editor
WebTranscribe editors
## Track Record

<table>
<thead>
<tr>
<th>Project</th>
<th>Speakers</th>
<th>Characteristics</th>
<th>Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ph@ttSessionz</td>
<td>1000</td>
<td>age 13-18</td>
<td>40+ (D)</td>
</tr>
<tr>
<td>ISASP</td>
<td>10</td>
<td>aspirated plosives</td>
<td>1 (IS)</td>
</tr>
<tr>
<td>LearnersCorpus</td>
<td>24 + 12</td>
<td>L2 English</td>
<td>2 (UK, D)</td>
</tr>
<tr>
<td>VOKALE</td>
<td>20</td>
<td>CI &amp; control</td>
<td>D</td>
</tr>
<tr>
<td>VOYS</td>
<td>300</td>
<td>Scotland</td>
<td>10 (UK)</td>
</tr>
<tr>
<td>CZECH</td>
<td>10</td>
<td>Czech prosody</td>
<td>2 (D, CZ)</td>
</tr>
</tbody>
</table>
VOYS

- Voices of Young Scots
- 10 recording locations in Scotland
- 30 speakers per location

Uses
- technology development
- documentation of regional accents
She took the butter, paused, and put it on the middle shelf.
WikiSpeech Conclusions

- modern and flexible implementation
- stable and proven data model
- extensible (annotation editors!)
- (free) service provided by BAS
- you provide specs and speakers
- we do the rest…
Links

• LearnersCorpus, VOYS, VOKALE
  • webapp.phonetik.uni-muenchen.de/wikispeech
• SpeechRecorder application
  • www.bas.uni-muenchen.de/Bas/software
WikiSpeech Demo

- http://webapp.phonetik.uni-muenchen.de/wikispeech
- login as user „org0014“ with password „Ma0lxc“
- select one of the demo projects
- select new session and fill in the form
- start recording session
- send your comments to draxler@phonetik.uni-muenchen.de