CLARIN Annual Conference 2016

Aix-en-Provence, 26-28 October 2016
Polish Read Speech Corpus for Speech Tools and Services

Danijel Koržinek, Krzysztof Marasek, Łukasz Brocki

Polish-Japanese Academy of Information Technology, Warsaw, Poland

October 28, 2016, Aix-en-Provence, France
About Clarin-PL

- B-type Clarin centre operating in Poland since 2013
- Run by teams from 6 Polish universities:
  - Wrocław University of Science and Technology
  - Institute of Computer Science PAS
  - PJAIT
  - Institute of Slavic Studies PAS
  - University of Łódź
  - Wrocław University
- It deals with various topics, including:
  - computer linguistics, social linguistics, language translation, language history and speech
- [http://clarin-pl.eu](http://clarin-pl.eu)
Speech resources at Clarin-PL

- **Motivation:**
  - lots of data used by HSS community exists in the form of audio
  - processing and analyzing this data is difficult and can be expensive

- This segment of the project is fully developed by PJAiT

- Consists of 2 main areas:
  - speech data
  - speech tools

- [http://mowa.clarin-pl.eu](http://mowa.clarin-pl.eu)
Speech data

- Speech corpora are expensive and hard to obtain, e.g.:
  - commercial: “CSLU”, “Speecon”, “GlobalPhone” and “Babel”
  - domain-limited: “Pelcra corpus of spontaneous speech” and “Spelling and Numbers Voice database”
  - restricted due to copyright or other rules and limitation

- Our goal was to create a free general-purpose speech corpus
We recorded and annotated \( \sim 56 \)h of studio quality read speech and \( \sim 13 \)h of telephone quality read speech.

The main purpose of these corpora is the development of speech processing tools.

It is available in two forms:

- EMU database (currently old format)
- Kaldi ASR baseline system

It is available on a liberal license (CLARIN PUB+BY+INF+NORED)

http://mowa.clarin-pl.eu/korpusty/
Emu database (old)
Emu database web service

- In development!

(source: http://ips-1mu.github.io/EMU.html)
## Kaldi baseline results

<table>
<thead>
<tr>
<th>WER %</th>
<th>experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.06</td>
<td>mono</td>
</tr>
<tr>
<td>17.56</td>
<td>tri1</td>
</tr>
<tr>
<td>16.75</td>
<td>tri2a</td>
</tr>
<tr>
<td>15.75</td>
<td>tri2b</td>
</tr>
<tr>
<td>13.50</td>
<td>tri3b</td>
</tr>
<tr>
<td>13.10</td>
<td>tri3b-sp</td>
</tr>
<tr>
<td>12.88</td>
<td>tri3b-20k</td>
</tr>
<tr>
<td>12.41</td>
<td>tri3b-mmi</td>
</tr>
<tr>
<td>11.64</td>
<td>+wide beam</td>
</tr>
<tr>
<td>7.37</td>
<td>+large LM rescoring</td>
</tr>
<tr>
<td>3.23</td>
<td>oracle of wide beam</td>
</tr>
<tr>
<td>9.25</td>
<td>TDNN</td>
</tr>
<tr>
<td>5.91</td>
<td>+large LM rescoring</td>
</tr>
<tr>
<td>2.83</td>
<td>oracle</td>
</tr>
<tr>
<td>8.91</td>
<td>LSTM</td>
</tr>
<tr>
<td>5.78</td>
<td>+large LM rescoring</td>
</tr>
<tr>
<td>2.61</td>
<td>oracle</td>
</tr>
</tbody>
</table>
Kaldi baseline usage

- everything available on Github:

  https://github.com/danijel3/ClarinStudioKaldi

- usage:
  1. download and install Kaldi
  2. git clone ... above repository
  3. modify path.sh and cmd.sh (if neccessary)
  4. ./run.sh
Other speech data

- The recorded corpus, while useful for tool development, lacks certain features for the actual study of language.
- We are currently working on other in-domain corpora:
  - PELCRA spontaneous speech corpus (with UŁ)
  - Polish Parliament (with UŁ and IPI PAN)
  - Kroniki - historical videos with news and current events (with University of Wrocław)
Speech tools diagram

- Speech signal
- Audio Analysis
- Voice Activity Detection
- Grapheme ↔ Phoneme
  - phonemes
  - adaptation
- Speaker Diarization
- Recording transcript
- Language Modelling
  - words
- Alignment
  - Speech segmentation
- ASR
  - Automatic transcription
- KWS
  - Detected keywords
Other similar Clarin initiatives in other countries:
- WebMAUS by LMU (speech segmentation)
- AVATech by Max Planck Institute and Fraunhofer Institute (video/audio processing, speech segmentation, VAD and speaker diarization)
- TTNWW (speech transcription services for Dutch)

We developed speech tools available as web services:
- Grapheme-to-phoneme conversion
- Speech alignment
- Speaker diarization
- Voice activity detection
- Keyword spotting
- Speech transcription

http://mowa.clarin-pl.eu/mowa
Grapheme-to-phoneme conversion

- Converting text from its orthographic into phonetic form
- Uses SAMPA phonetic alphabet
- Rule-based system
- Allows multiple word pronunciations
- [http://mowa.clarin-pl.eu/transcriber](http://mowa.clarin-pl.eu/transcriber)
Text-to-speech alignment

- Given a transcription and an audio recording, we can calculate accurate alignment on word and phoneme level
- Also works on long audio (up to ~30 minutes)
Voice activity detection

- “Naive” methods are easily deceived
  - thresholding, energy, 0-cross, running average,…
- Uses a trained acoustic model to reject non-speech events
  - knocks, noise, music, …
- Difficulties with para-linguistic noise
- Uses a frame-based RNN model
- Has very high recall (>99%), but precision is still an issue (lots of noise can be misclassified as speech)
- Classification of non-speech was also attempted
Speaker diarization

- Multiple levels of speaker recognition:
  - speaker change detection
  - speaker diarization (← this was done)
  - speaker identification
- Currently based on LIUM speaker diarization system
- Results are provided in the form of speech segmentation
Keyword detection

- Often we don’t need a full transcript
- We can provide a list of keywords with the audio file and the system will generate a list of likely occurrences and their location
- It uses an ASR system with a general LM
- OOV words are modelled using syllables
Speech transcription

- Probably one of most sought after services
- Speech recognition works best when limited to a specific domain
- We provide a demonstration system for now, but would like to expand to specific domains useful in HSS research
- Based on the Kaldi toolkit for speech recognition
Selected applications

- Speech alignment/segmentation was used to annotate the Pelcra corpus of spontaneous speech
- Alignment was also used in the study titled “Respeaking - the process, competences and quality”
- Attempts were made to transcribe social science interviews
Future plans

- Additional corpora will be annotated and delivered on their respective platforms
- Usability improvements through integration with the EMU web platform
- Development of a transcription service aimed at HSS research
- Facilitation of cooperation with more partners in the HSS community
Contact

- Danijel Koržinek - danijel@pja.edu.pl
- Krzysztof Marasek - kmarasek@pja.edu.pl
- Łukasz Brocki - lucas@pja.edu.pl