More than 35,000,000,000,000 words are spoken on Earth every day.

Some of them matter.

Petr Schwarz

Speech Transcription + Key Word Spotting

Gender Identification

Speaker Identification

Age Estimation

Language Identification

Speech Quality Estimation

PHONEXIA
SPEECH ENGINE

Diarization

Voice Activity Detection

Governmental use cases

Intelligence
Use Cases

PHONEXIA
SPEECH PLATFORM

Forensic
Use Case

PHONEXIA
VOICE INSPECTOR

LEA / Police
Audio Investigation

PHONEXIA
ORBIS

Commercial use cases

Over 25 years in Speech processing

60+ people

Projects in 60 countries around the world

Based by BUT researchers in 2006

ELG Benefits

- Training data for speech transcription in many languages
  - speech data + transcripts
  - Textual data for specific language domains
  - Textual data with annotation - named entities, digits/dates/times in spoken and normalized textual forms …

- The transcripts and annotations do not need to be 100% correct, can be from automatic research or commercial systems

- It helps significantly if all the data have the same format. We license it from many sources and the major part is data unification.

- The data format should be uniform across languages too.

Contributions to the ELG and its marketplace

- We can offer speech data mining tools as executables or cloud-based APIs
  - spoken language recognition
  - gender recognition
  - speaker identification
  - speech transcription / keyword spotting
  - voice activity detection
  - speaker diarization
  - age estimation

- Currently we offer 15 languages over Amazon Web Services

- We can offer all the tools for automatic transcription / annotation of the available data on ELG to make it more valuable

- We collect some data in collaborative research projects

Motivation and goals to engage in the ELG, benefits and positive effects

- Reduced costs of data acquisition
  - currently we use many providers – LDC, ELRA, Appen, SpeechOccean, DataTang …
  - data collection through some collaborative efforts
  - data cleaning and format unification through some collaborative efforts
  - data extended with some metadata (for example speech data transcribed by several research and commercial recognizers)
- New ways to the market – for example, a unified way to offer/promote the technologies in multiple cloud services, including European ones
- Shared development of data cleaning / processing tools
- Shared development of models
- Shared development of technologies
- Sharing of computational resources (related costs)
- Help with legislation and regulations around data

Potential for Collaboration with the ELG and participants

- Current trends in big pre-trained models and self supervised training (in all areas – speech, NLP, dialog, video …) makes it harder for small companies to compete with big ones
- The resources (both human and computational) could be shared to collect/clean data and to train such models
- Several companies may find down-stream tasks that could be addressed through the same pre-trained models

Connection among data, tools, models and community

- An interesting trend and huge potential is in bringing data, tools, models and the community to one place together
- There are many pieces over internet – data repositories, code repositories, online tools, egg examples in many research toolkits (Kaldi, SpeechBrain …), PyTorch examples
- A great example is Hugging Face
- With community (and computational resources), all the above lives