

SPRAAK: Speech Processing, Recognition and Automatic Annotation Kit

Kris Demuynck



Katholieke Universiteit Leuven - ESAT

SPRAAK: what

- State-of-the-art speech recognition system
 - Facilitate (research) collaboration by means of a common platform
 - Facilitate development of niche products, demo systems
 - Focus on Dutch
- STEVIN Project
 - One of the STEVIN priorities
 - period: 01/12/2005-31/05/2008
 - partners: ESAT, RU Nijmegen, U Twente, TNO

SPRAAK: aims

- Modular toolkit (plug&play) for **research on speech recognition**, allows researchers to focus on certain aspects only
- Customizable recognizer with a **simple interface, useable by non experts**

SPRAAK: aims, application developers

- Ready-to-use system that is simple to use
- A high-level API: start/stop, set parameters, load/change acoustic models and language models, audio control, ...
- Client/server setup available
- Only limited knowledge of ASR required; only limited programming skills
- Example configurations
- High level scripts (training, alignment, ...)

SPRAAK: aims, researchers

- Dig deeper in internals, use low-level API
- Low-level API available for C and Python
- Variable amount of programming work, eg. feature extraction vs. tweaking of decoder
- Complete access to all routines
- Detailed knowledge of software package required
- Modules implemented using object oriented paradigm; scripts in Python

SPRAAK: aims, researchers

- Standalone tools can be built with scripts (eg. alignment, enrollment, training of acoustic models, ...)
- Use the low-level API if drastic changes are required eg. changing the decoder, ...
- Prototype new concepts in Python
- Low-level API (Python) also interesting for teaching ASR: allows insight in ASR internals, visualisation of intermediate results (eg. individual gaussians), change specific parts without worrying about the rest
- Good programming skills required

SPRAAK: examples

- Keyword spotting: indexing of archives, call centers, intelligence agencies
- CALL (computer assisted language learning): pronunciation training
- “conventional” applications for other (small) languages (eg. Frisian)
- Transcription/segmentation of speech
- Applications in niche markets, not served by commercial products
- Research oriented applications
- Fundamental research

SPRAAK: design decisions

- Object oriented paradigm (but still C)
 - well defined objects with clear interfaces
 - provide 'hooks' for extensions
 - allow plug-and-play facilities
- Efficiency (CPU, memory)
 - reuse of resources
 - efficient implementation of all core components
- Future proof
 - multi-threading
 - multiple input/output formats (extensible)
 - highly configurable (no fixed configuration)

SPRAAK: current status

- First release is ready (licensing model not final)
- Low level (object oriented layer, I/O, MT-safe, own name space, improved Windows integration, ...) is up and running
- Improved & new high level scripts:
 - parallel training \Rightarrow considerable speedup
 - grammar compiler
 - text normalization (language modeling)
- Example scripts for training
- Low level documentation: ready
- High level documentation: draft version
- High level API: concepts are ready

SPRAAK: future

- Still one month of full-time development
- Further development in follow-up projects (STEVIN and others)
- SPRAAK workshop (to be planned)
- TODO list:
 - high level documentation: final version + integration
 - high level API: implementation
 - scripts for robust alignment (follow-up projects: NEON, HATCHI, BATS)
 - cleanup of some API's (acoustic model, lexicon, search)
 - more examples on how to use the API's, scripts, ...

SPRAAK: IPR

- **For research:** freely available (distribution cost), without support (open source model)
 - If support is required: involve SPRAAK partner(s) in project with provided budget for support
 - SPRAAK partners can deliver expertise wrt addition of new functionality
- **Commercial use:**
 - licences; royalties put in fund for support and extension of the package ⇒ whole SPRAAK community benefits from it.
 - support and new developments: SPRAAK partner(s)
- **SPRAAK consortium:**
 - ensure the longevity of SPRAAK
 - ensure that enhanced/new functionality ends up in SPRAAK
 - income: commercial licenses and support contracts

SPRAAK: distribution

- Open source license: through NTU/ESAT website after registration, free
- Binaries, demo recognizers: TST-centrale, distribution cost
- Commercial contracts: via SPRAAK consortium, prices to negotiate

SPRAAK: conclusions

- SPRAAK is (almost) ready
- Is a platform (toolkit), not a final product
- Suitable for research projects (collaboration between groups) and applications that are not served well by commercial products