

SPEECH RECOGNITION TECHNOLOGIES FOR ELECTRONIC DICTIONARIES

Sergij Černičko, Petr Lokaj, Jiří Janoušek, and Michal Kašpar

Lingea s.r.o., Brno, Czech Republic
michal.kaspar@lingea.com

1. COMPANY INTRODUCTION

LINGEA is the Central European leader in language technology and language data development with experience of over 15 years in helping customers around the world and across private and public sectors.

LINGEA's main activity related to ASRU topics are **language technologies**: Spell Checker, Lemmatiser, Thesaurus, Hyphenator for 16 languages, which help users all around the world when writing, text editing, full text searching etc., whether delivered as part LINGEA dictionary applications for computer and mobile platforms or implemented into third party online and offline solutions.

LINGEA also offers **licensing** of a vast database of language data, which has been developed for over 15 years and now includes data for 60 languages. The data can be used for development of complex language tools as well as for publishing dictionaries.

2. LINGEA AND SPEECH TECHNOLOGIES

LINGEA has detected the potential of speech recognition for its electronic dictionaries and educations product series already in 2005. It has teamed up with BUT Speech@FIT research group, and, during project "Research and development of corpus and speech technologies in new generation of electronic dictionaries" supported by the Ministry of Trade and Commerce of the Czech Republic (MPO), it will introduce voice-search and visualization of pronunciation into its flag-ship products Lexicon and Handylex. The production version of the software runs on the BS-CORE library produced by Phonexia [BS-CORE].

The cooperation with BUT continued in a follow-up MPO project "Multilingual recognition and search in speech for electronic dictionaries" and in "Technologies of speech processing for efficient human-machine communication" supported by the Technology Agency of the Czech Republic. In these projects, the accent is on rapid development of speech recognition and forced-alignment engines for new languages and on mobile platforms.

3. LINGEA ASRU 2013 DEMOS

The demonstrations prepared for ASRU 2013 include voice-search in dictionary on a mobile platform and feedback on user's pronunciation. Both are based on BUT's neural-network phone recognition architecture making use of long-temporal context [Schwarz2009]. From the implementation point of view, the software runs on BS-CORE library produced by Phonexia that provides all speech recognition functions, the only different component being the recognition network.

3.1. Search in mobile dictionaries

The dictionaries are available for 13 language pairs, voice-enabled search is now available for English and Czech. The recognition network is a simple parallel combination of all tokens in the dictionary. The recognizer produces a sorted n-best list that is presented to the user (Figure 1).

Beta version of the voice-enabled search was integrated within dictionaries for mobile platforms iOS, Android and Windows Phone. The final version for end customers is due to be launched in 2014.



Figure 1: User interface of HandyLex with voice search.

3.2. Pronunciation feedback

Correct pronunciation is key in language teaching as well as in ‘operational’ scenarios (business trip or holiday in a foreign country). Electronic dictionaries usually contain reference pronunciation spoken by a native speaker. Some can record user’s one, compare the two waveforms and produce a more or less obscure numerical “quality assessment”. Our decision was to align user-spoken waveform with the reference one, and show him/her the individual phones. This way, the user is able to check the duration and quality of individual phones and eventually improve his/her pronunciation.

The pronunciation feedback works as client-server application as shown in Figure 2. The recognition network is generated by simply chaining the phonemes from reference pronunciation. As Lingea and BUT Speech@FIT use different phoneme-sets, a set of mapping rules was developed. The demo is available in Spanish, German English, and Turkish with user interface shown in Figure 3.

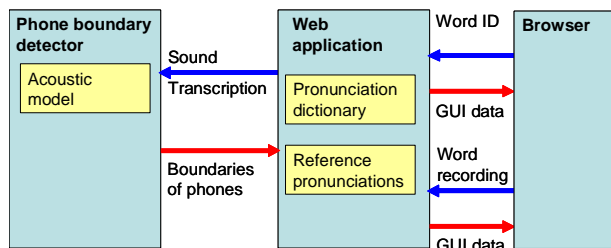


Figure 2: Pronunciation feedback block scheme.

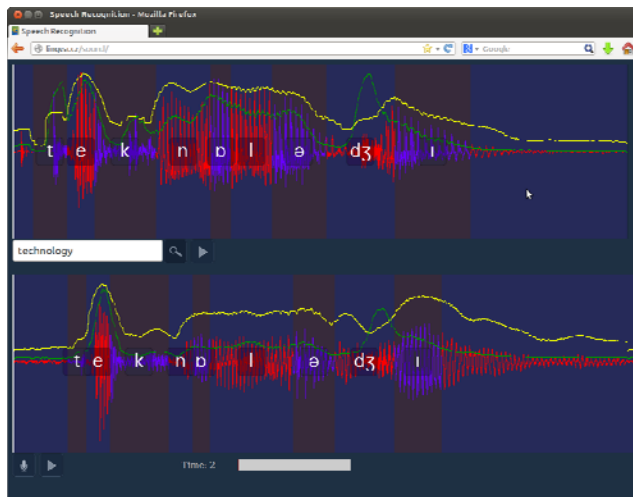


Figure 3: Pronunciation feedback GUI.

4. ACKNOWLEDGEMENTS

This work was partly supported by Technology Agency of the Czech Republic grant No. TA01011328 and by the Ministry of Trade and Commerce of the Czech Republic Project No. FR-TI1/034.

5. REFERENCES

- [BS-CORE] Brno Speech Application Interface Documentation <http://phonexia.com/docs/bsapi/>.
- [Schwarz2009] Petr Schwarz: *Phoneme recognition based on long temporal context*, Ph.D. thesis, Brno University of Technology, 2009.
- [HandyLex] Dictionaries for smartphones and tablets, <http://www.lingea.cz/mobile/>