Ortfon2 - tool for orthographic to phonetic transcription

Dawid Skurzok1,2, Bartosz Ziółko1,2

1. AGH University of Science and Technology, Kraków, Poland, Department of Computer Science, Electronics and Telecommunications
2. Techmo, Kraków, Poland

Purpose: Grapheme to phoneme transcription is widely used in Text-To-Speech and Automatic Speech Recognition (ASR) systems. There are two methods of performing such transcription automatically: rule based and data driven. Rule based method are based on context decisions provided by linguists. The number of rules is huge and using them is not obvious. One of our aims was to create both fast and flexible system while preserving easy and intuitive interface. Many existing tools for language processing are written using some scripting language. It significantly degradates speed efficiency and creates many problems during integration with systems created in compiled languages, especially in embedded system. Moreover some high secure environments prohibit installation of any scripting language.

Method: We chose C++ to make our tool fast and easy to integrate with SARMATA ASR system. To provide simple and easy interface for rules editors we created a small declarative language on top of a CSV file format. File in this format can be created in every spreadsheet editor. Expressions of our language can just be placed in the cells of spreadsheet. To use in Ortfon2, CSV file is compiled and converted to an internal file format. Compilation process optimize much of the transcription rules to minimize its further processing.

The first cell in the transcription rules table denotes a letter for which this table may be used. First column contains prefixes for this letter and first row consists of suffixes, both in form of orthographic expressions. Every other cell contains a phoneme, a sequence of phonemes, a neutral symbol or an empty symbol. Every table has also a number specified by a user which can be used while debugging. Phonetic neutral symbol means that a given letter does not produces a phoneme.

Ortfon2 uses its own specialized parsing engine to compare text input with an orthographic expression. It may produce a transcription using an ASCII name for symbols, as well as UTF-8 encoded IPA and Slavonic symbols.

Results: The quality of transcription was measured by evaluating ASR system performance. ASR using first version of Ortfon recognized correctly 95% of phrases, while using the second version recognized 97%. The test was performed on 10 000 phrases. Ortfon2 transcribes more than 36 thousands of words per second.

Acknowledgments: This work was supported by LIDER/37/69/L-3/11/ NCBR/2012 grant.