Tuning a CMU Sphinx-III Speech Recognition System
for Polish Language

Marcin Płonkowski, Paweł Urbanowicz
The John Paul II Catholic University of Lublin, e-mail:plomyk@kul.pl

CMU Sphinx-III [1] is one of the most popular speech recognition systems. There are many publications describing the system in relation to the English language and some publications in relation to other languages [2, 3]. For the Polish language, there are some publications describing the CMU Sphinx [4]. However, there are no publications describing many parameters of this system and its testing in relation to the Polish large vocabulary continuous speech recognition.

In this paper authors describe parameters which may be tuned to obtain the best performance and accuracy for a large vocabulary continuous speech recognition task. These parameters are: beam pruning, absolute pruning, fast GMM computation, language weight, word insertion penalty, number of lexical tree instances [5]. Behavior of certain parameters should be similar regardless of the language speech recognition. However, some parameters will have a different impact on the accuracy of the Polish speech recognition as compared to the English speech recognition [6].

There are a lot of differences between Polish and English from the point of view of the speech recognition problems. English is a positional language and Polish is highly inflective. According to the DLIFLC Categories of Language Difficulty, Polish ranks in Category III out of four levels of difficulty, together with other Slavic languages [7]. Hence, the Polish speech recognition is a very difficult task.

The speech recognition task requires a lot of tests and experiments in order to minimize the number of errors and thus improve accuracy. On the other hand, increasing the accuracy has a bad influence on the recognition time making it impossible to decode speech in real time. Therefore, sometimes it is worth reducing the accuracy of the speech recognition to speed up time. This is particularly important in mobile devices which do not have the same performance as desktop computers. So it is important to find a compromise between the accuracy and performance of the speech recognition system.

References
[1] CMU Sphinx: The Carnegie Mellon Sphinx Project. http:// http://cmusphinx.sourceforge.net/, Apr 2013.
[2] Varela A. , Cuayuhuitl H., Nolazco-Flores J. A., Creating a Mexican Spanish Version of the CMU Sphinx-III Speech Recognition System, CIARP 2003, LNCS 2905, p. 251–258, 2003.
[3] Satori H., Harti M., Chenfour N., Arabic Speech Recognition System Based on CMUSphinx, IEEE Proceedings of ISCIII'07, Morocco, p. 31 - 35, 2007.
[4] Janicki A., Wawer D. Automatic Speech Recognition for Polish in a Computer Game Interface, In proceeding of: Federated Conference on Computer Science and Information Systems - FedCSIS 2011, Szczecin, Poland, 18-21 September 2011, p. 711-716.
[5] Sphinx-3 s3.X Decoder, http://www.cs.cmu.edu/~archan/s_info/Sphinx3/doc/s3_description.html.
[6] Vertanen K.: Baseline WSJ acoustic models for HTK and sphinx: Training recipes and recognition experiments, http://www.keithv.com/pub/baselinewsj/baseline_wsj_recipes.pdf (2006)
[7] Lowe, P. Jr., Zero-Based Language Aptitude Test Design: Where's the Focus for the Test? Applied Language Learning, 9(1&2) 1998, 11-31.