

Using Principles of Language Modeling in Acoustic-Phonetic Decoding

David Andrš¹

¹ University of West Bohemia, Faculty of Applied Science, Department of Computer Science and Engineering, Univerzitní 8, 306 14 Pilsen, Czech Republic

E-mail: andrsd@kiv.zcu.cz

Keywords: decoder, language models, czech

Abstract: The purpose of the decoder in speech recognition task is to derive one or more utterance hypotheses from the stream of phoneme probability estimations. The phoneme estimations are computed for each part of the speech signal (called *frame*).

Nowadays, the decoders used in speech recognition task are based on two basic algorithms, Viterbi algorithm and A* search. Both algorithms use language models but Viterbi algorithm can handle only bigram language models. The A* search can handle language models of higher order but the crucial part of it is the evaluation function. It directly determines how fast the search will be. Another disadvantage of the Viterbi algorithm is its poor performance on on large vocabularies.

In this paper, we propose a new decoding method, which combines principles used in Viterbi algorithm with principles of language modeling used in A* search. The method uses a vocabulary of predefined words that can be included in the hypothesis, thus the resulting hypotheses are not restricted by a grammar or language models. The language models will be used in further phase – *rescoring*.

The method utilises two level model – a segment and a word model. Words are automatically divided into segments that consist of phonemes. During the decoding, the recognition proceeds in a reverse direction. The segments are built from phonemes and then the words are built from the segments.