Word Boundary Detection Based on Phoneme Sequence Constraints[1]

György Szaszák and Zsolt Németh

Most of ASR applications are based on statistical approach, often using Hidden Markov models to model phonemes. For the purpose of speech recognition, words in the dictionary of speech recognizers are mapped to phoneme model sequences. In case of continuous speech recognition even word networks are connected together during the decoding process, which needs much more computation performance and which also might lead to more confusions due to the longer speech patterns increasing the number of canditate word sequences. In ASR, the Viterbi algorithm is used for recognition to find the most probable match. Using word level segmentation or word boundary information in ASR systems can decrease the searching space during the decoding process thus increase recognition accuracy.

In human speech perception word boundary detection depends on many factors, including prosodic, semantic and sintactic cues [4]. The interaction among these factors is difficult to model, hence prosodic or sintactic information are usually handled and processed separately. Speech production is a continuous movement of the articulating organs, producing a continuous acoustic signal. In human speech processing, linguistic content and phonological rules help the brain to separate syntactic units, such as sentences, phrases (sections between two intakes of breath), sintagms or even words.

Since the mid-eighties there have been several trials to exploit prosodic information in human speech [5, 6]. In the Laboratory of Speech Acoustic of the Budapest University of Technology and Economics, we have already investigated word boundary detection possibilities based on prosodic features, reliing mainly on fundamental frequency and energy level data derived from the acoustic signal [1]. We have also trained a prosodic HMM based word boundary segmentator in order to use it as a front-end module for Viterbi decoding in ASR [2].

Using phoneme sequence information is an alternative for word boundary detection. Phoneme sequence constarints can be derived by matching a complete set of 3 phoneme sequences that can occur accross word boundaries [3, 8]. Phoneme assimilations must be taken into account over word boundaries.

In this paper we would like to investigate the use of phoneme sequence constraints in word boundary detection for Hungarian language. We have collected a phonetically rich, large Hungarian text database from the Internet. This material was transcribed phonetically taking into account all possible assimilations over word boundaries, and also the pronunciation variants of words [7]. We matched against this corpus complete sets of 2 or 3 phoneme sequences in order to get a frequency statistics of each phoneme sequence element word internally and over word boundaries. In this way we obtained a frequency database for all possible phoneme sequences in Hungarian. Based on this information we derived a set of rules (eg. phoneme sequence constraints) which can be used for word boundary detection for Hungarian language. We also investigated the adaptability of the method for phoneme based speech recognizer's output that contains some phoneme confusions, since phoneme recognition accuracy in ASR - just like in case of human listeners - is around 70-80%. A new challenge is the paralel use of prosody based and phoneme sequence based word boundary detection methods, which might be of interest in our future work. In this paper we would like to present our methodology and evaluate our results obtained by phoneme sequence based word boundary detection for Hungarian language.

References

[1] K. Vicsi and Gy. Szaszak. Automatic Segmentation of Continuous Speech on Word and

[1]The work has been supported by the Hungarian Scientific Research Foundations OTKA T 046487 ELE IKTA 00056.

Phrase Level based on Suprasegmental Features. In: *Proceedings of Forum Acusticum* 2005 Budapest, Hungary, 2669-2673.

[2] Gy. Szaszák and K. Vicsi. Folyamatos beszéd szószintű automatikus szegmentálása szupraszegmentális jegyek alapján, *MSZNY 2005*, Szeged, Hungary, 360-370.

[3] J. Harrington, G. Watson, and M. Copper. Word Boundary Identification from Phoneme Sequence Constraints in Automatic Continuous Speech Recognition, 1998, COLING: 225-230.

[4] S.L. Mattys and P.W. Jusczyk. Phonotactic and Prosodic Effects on Word Segmentation in Infants, *Cognitive Psichology*, 1999, 38: 465-494.

[5] Di Cristo. Aspects phonétiques et phonologiques des éléments prosodiques, *Modeles linguistiques Tome III*, 2:24-83.

[6] M. Rossi. A model for predicting the prosody of spontaneous speech (PPSS model), *Speech Communication*, 1993, 13:87-107.

[7] K. Vicsi and Gy. Szaszák., Examination of Pronunciation Variation from Hand- Labelled Corpora, *TSD 2004 Brno*, Czech Republic: 473-480.

[8] K. Müller. Probabilistic Context-Free Grammars for Phonology. *Workshop on Morphological and Phonological Learning at Association for Computational Linguistics 40th Anniversary Meeting (ACL-02)*, University of Pennsylvania, Philadelphia, USA.