A Hierarchical Evaluation Methodology in Speech Recognition

Gábor Gosztolya and András Kocsor

Research Group on Artificial Intelligence
of the Hungarian Academy of Sciences and University of Szeged
H-6720 Szeged, Aradi vértanúk tere 1., Hungary
{ggabor, kocsor}@inf.u-szeged.hu
http://www.inf.u-szeged.hu/speech

Extended Abstract

Automatic speech recognition (ASR) is a pattern classification problem [1, 2] in which a continuously varying signal has to be mapped to a string of symbols (the phonetic transcription). Besides the identification of speech segments to grammatical phonemes [3], efficient searching in the induced hypothesis space [4, 5] is of great importance as well. This work is connected to both areas: first we give a hierarchical scheme of the Hungarian phonemes (see Fig.1.), then we try to exploit this structure in the search process.

For this hierarchical classification we used traditional grammatical features (voicedness, roundness, etc.) to characterize phonemes, which came from the physical articulation of speech sounds/phonemes.

Since the hypothesis space in ASR is generally a search tree, standard tree search methods can be applied. In addition to these algorithms, the characteristics of the speech recognition problem led to the development of search techniques especially suitable for ASR hypothesis spaces [6]. Here we propose a *multi-pass* search method [7] (which belongs to the speech-related family) using the above-mentioned hierarchical partition. In general, multi-pass methods work in two (sometimes more) steps: in the first pass the less likely hypotheses are discarded by using some condition requiring low computational time. Then, in the second pass, only the remaining hypotheses are examined by more complex, reliable evaluations, which will approximate the probabilities of the hypotheses more closely. (In the common search methods – such as *Viterbi beam search* method [8] – the first pass is omitted, so more hypotheses are scanned in the second pass, making the process more time-consuming.) Our method employs conditions using grammatical features for the first pass, because they are strongly associated with the phonemes, so they alleviate the solution of the decision-making task.

Finally, after examining the results, we found that with the proposed hierarchical evaluation methodology we were able to significantly decrease the run time of our speech recognition system [9].

References

1. K. Fukunaga, *Statistical Pattern Recognition,* Academic Press, New York, 1989.

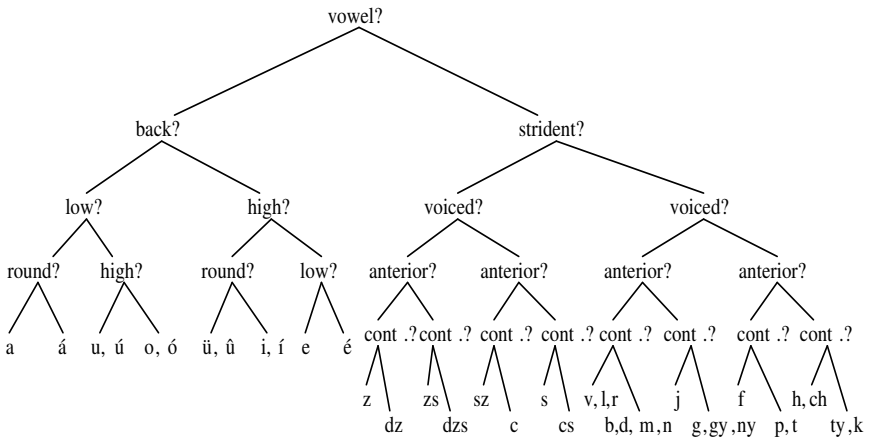

Fig. 1. A relatively detailed hierarchical scheme of the Hungarian phonemes

2. C.M. Bishop, *Neural Networks for Pattern Recognition,* Clarendon Press, Oxford, 1995.
3. A. Kocsor, L. Tóth, A. Kuba Jr., K. Kovács, M. Jelasity, T. Gyimóthy and J. Csirik, *A Comparative Study of Several Feature Space Transformation and Learning Methods for Phoneme Classification,* International Journal of Speech Technology, Vol. 3, Number 3/4, pp. 263-276, 2000.
4. G. Gosztolya, A. Kocsor, L. Tóth and L. Felföldi *Various Robust Search Methods in a Hungarian Speech Recognition System,* Acta Cybernetica 16., pp. 229-240., 2003.
5. G. Gosztolya and A. Kocsor, *Improving the Multi-Stack Decoding Algorithm in a Segment-based Speech Recognizer,* Proceedings of the 16th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE 2003, LNAI 2718, pp. 744-749, Springer Verlag, 2003.
6. F. Jelinek, *Statistical Methods for Speech Recognition,* The MIT Press, 1997.
7. X. Huang, A. Acero and H.-W. Hon, *Spoken Language Processing,* Prentice Hall PTR, 2001.
8. P.E. Hart, N.J. Nilsson and B. Raphael, *Correction to "A Formal Basis for the Heuristic Determination of Minimum Cost Paths",* SIGART Newsletter, No. 37, pp. 28-29, 1972.
9. A. Kocsor, L. Tóth and A. Kuba Jr., *An Overview of the Oasis Speech Recognition Project,* Proceedings of ICAI '99, Eger-Noszvaj, Hungary, 1999.