

Summary

In this thesis we first adapt a N-best solution search method to the CNET speech recognition system then we develop a segmental and a syntactical post-processing to recover the correct solution in the proposed N-best list.

The N-best solutions for each pronounced word (or sentence) are found using a two pass algorithm: the forward pass uses the Viterbi algorithm and the backward pass uses the A* algorithm. This method applied at the acoustical level, allows us to have the phonetic segmentation of the solutions. For a small N, the N-best solutions search time is an affine function of the number of developed solutions.

The syntactical post-processing is used if the syntax of the application cannot be implemented in the decoding process directly. One solution to get around this problem is to use an overgenerating syntax in the language model and to develop the N-best solutions. Afterwards in a syntactical post-processing the first solution within the N proposed, which respects the real syntax of the application is declared correct. This approach leads to a small substitution error rate on the spelling recognition city name task.

For isolated and connected word recognition tasks, for which the application syntax is directly integrated in the decoding process, we developed a segmental statistical post-processing. The N-best solutions are described as sequences of phonetic segments. We introduce a new method in the statistical modelling of each segment which involves two models: one for the correct segmentation and one for the incorrect segmentation. The segmental score of each solution is the combination of the scores obtained with these two models. This new method allows us to get a 15% to 25% reduction in the error rate, depending on the databases, when compared to the HMM alone.

