



Comments on “Towards increasing speech recognition error rates” by H. Boullard, H. Hermansky, and N. Morgan

Renato De Mori

School of Computer Science, McGill University, 3480 University Street, Montreal, Quebec H3A 2A7, Canada

Received 8 January 1996

This paper is a valuable effort to stimulate discussion on basic issues of a problem for which interesting solutions have been found which on one side made research results applicable in practice and on the other side have shown evidence of imprecision and incompleteness.

A fundamental problem for me is the modeling and integration of various sources of knowledge. The paper seems to insist on acoustic analysis and modeling and on the scoring theories that should be used for integrating acoustic and language models. More attention is probably deserved by the number, type, training methods of other models for lexicon, language, syntax, semantic and discourse. For these models it is very important, if the goal is automatic speech recognition or understanding, to conceive scoring methods for the hypotheses they generate in such a way that scores from different models can be effectively integrated.

An important question seems to concern the value of a model that is sound, based on acceptable grounds, but is not ready to be integrated with actual systems or has been integrated without producing significant performance improvement. Examples could be better phoneme recognition systems which do not allow to obtain better word recognition, grammars with a larger coverage on a corpus which do not reduce perplexity, local parsers with which new probabilistic scores can be derived without increasing dictation performances, methods for automatic

learning semantic interpretation knowledge which is not better than manually derived knowledge.

If integration of different knowledge sources is based on probabilistic scores, the following consideration seems important.

The Bayes rule is used, in today popular systems, for introducing two models based on which it is possible to compute probabilistic scores for signal interpretation. Models are trained independently, each one is trained in a sub-optimal way. So the probabilities computed with these models are just approximations of the true ones making the entire scoring system imprecise. Usually this imprecision leads to performance degradation which are partially compensated by heuristics like the introduction of an exponent for the language model probability.

The Neural Network solution does not avoid this problem and has an interest in the fact that it computes different probabilities with a different model that can be more accurate in taking, for example, signal history into account or in modeling certain types of probability density functions.

Another important question is about the experimental evaluation of new methods. The available corpora have been conceived for specific purposes. Can they be used for evaluating new components developed for other purposes?

What is the value of a completely new set of ideas tested on a standard digit corpus with higher error rates than the best reported in the literature? How

much higher is acceptable? What is the value of a relative improvement if the best result is poorer than the best published one? How important is to know “what worked better or what did not work”?

When the error rate is sufficiently low (e.g. less

than 10%) is it important to evaluate new ideas based on the overall error rate, rather than measuring, for example, improvements on specific classes of words (e.g. key words, monosyllabic words, acoustically confusable words, digits, letters)?