



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

Jean-Paul Haton

CRIN / INRIA, Nancy, France

Received 9 February 1996

I very much enjoyed the paper of H. Boullard and his colleagues. I found it bright and stimulating, and I agree with the main conclusions that are drawn. Below are only brief and minor comments that result in part from discussions about this paper with my colleague Yifan Gong.

I was neither really surprised, perhaps in part because I know Hervé, nor shocked by the form of the paper. In a sense, I preferred the initial, more controversial, version that was presented at Eurospeech'95 in Madrid [1].

The plan of the article is a little bit confusing. In particular, Section 4, devoted to the search for efficient features for AQSR, contains a large number of topics that are not really useful in the framework of the papers, since they have always been decreasing the error rates (dynamic features and dynamic cepstrum, RASTA processing, cepstral mean subtraction, etc.). Perhaps, this material could have been moved to the state-of-the-art section.

In Section 5.1, the authors address the important issue of speech dynamics modelling by stochastic segments. This is a promising area in which few groups have been working so far. Almost all of them are cited ... except my group in Nancy. It seems to me, but I am not sure that I'm in position to be

completely objective, that our Stochastic Trajectory Model, STM has interesting, and, to my knowledge, unique characteristics [2,3].

As mentioned in Section 5, some recent works have tried to bridge the gap between knowledge-based acoustic-phonetic decoders and statistical systems. I would more strongly advocate for such approaches. Experimental comparative studies that we have conducted in Nancy confirm that pure knowledge-based models remain top-ranked candidates for increasing the error rate ... However, I remain convinced that incorporative speech knowledge (of a nature and in a manner that still need to be determined) should be an ingredient of the necessary breakthrough towards long-term improvements in ASR.

References

- [1] H. Boullard (1995), “Towards increasing speech recognition error rates”, *Eurospeech'95, Madrid*, pp. 883–894.
- [2] Y. Gong and J.P. Haton (1994), “Stochastic trajectory modeling for speech recognition”, *ICASSP'94, Adelaide*, pp. 57–60.
- [3] M. Afify, Y. Gong and J.P. Haton (1995), “Stochastic trajectory models for speech recognition: an extension to modeling time correlation”, *Eurospeech'95, Madrid*, pp. 515–518.