



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

Jim Flanagan

CAIP Center, P.O. Box 1390, Piscataway, NJ 08855-1390, USA

Received 12 January 1996

As requested, I am pleased to remark upon the provocative and wide-ranging paper by Boulard et al.

The title and slant are chosen for dramatic impact, and serious readers will not be led into believing that research should overtly aim for negative results. There are always more things that don't work than do, and life is short.

I resonate with the implication that guiding research into intense competition fetters thinking and discourages risk-taking. It fosters overlap and duplication. Researchers, whose funding may otherwise be jeopardized, tend to work on the same thing – namely, the surething established method of the moment. This channeling has been a failing in contract-supported research in the past. The present result is highly sophisticated recognition algorithms which work reliably when honed for specific data types, but which exhibit frailty under alternate scenarios. This is not to say competitive comparisons are not valuable. They are. But as a central research thrust they are not cost effective in advancing knowledge. And, when duplicated many times, the costs escalate.

Task-specific applications pay off big. Witness the success of voice-controlled routing of telephone calls. But, the cost and effort of training current models is truly onerous. The apparent need to expose systems to databases of ever-increasing size smacks of desperation. It would seem prudent to devote a reasonable portion of contract resources to bold, searching, speculative research that constantly tests

the frontiers of understanding. Most major advances come from such efforts – not from incremental polishing. A significant failure rate signals an appropriate level of audacity.

The collection of speculative issues offered in the paper are admittedly parochial. Mine are, too. Two that I would have wished to see on the list are:

- *New parametric descriptions of speech information.* Such descriptions can transcend linear acoustics and exploit first principles of sound generation from fluid flow and articulatory dynamics. If “gesture” is indeed the underlying information – recall, Sir Richard Paget suggested that speech was invented when ancient man found it inconvenient to “talk with his hands full” – articulatory-based descriptions might provide exceptional robustness and parsimony. Compact parameterization, based upon constraints of speech generation, could contribute to a coalescing (and simultaneous solution) of the problems of speech coding, synthesis, and recognition.
- *Distant-talking operation of speech systems.* As speech technology advances, users will tire of the unnatural need for close-talking sound pick-up – with the attendant inconvenience of body-worn, hand-held, or tethered equipment. Low-cost, high-quality microphones and economical DSP can now be used in quantity for sophisticated microphone arrays. These arrays can achieve sound capture with spatial selectivity in three dimensions, and can mitigate multipath distortion

and multiple interfering sources. Ultimately, sound pick-up for speech interfaces should approach the freedom and naturalness of face-to-face communication.

The paper ranges perhaps farther and is more

lengthy than necessary to make its point. But, the authors have performed a singular service to the speech community in spotlighting critical philosophies and inviting discussion.