

Workshop CLSP 97 at the Johns Hopkins University

Activity report of Prof. Chris J. Wellekens

1 Objectives

Invited by Professor Jelinek to participate at the Workshop 97, I collaborated in the team Acoustic Modeling. Several topics were proposed by our group leader Professor Andreas Andreou.

1. Heteroscedastic LDA applied to the training for improvement of discrimination
2. Long term filtering of the feature vectors (long window analysis: up to 1 sec according to Dr Hermansky).
3. Vocal track normalisation
4. Remap training (continuation of research of the previous year)

In the first days, members of the group (Hermansky, Luettin, Wellekens) agreed to cover the problem of long window data based filtering while Dr Minani worked on the channel adjusting using a trained bias operator. C.Kamm worked on the vocal tract normalization and Prof. Andreou on LDA training (in continuation of the thesis of N.Kumar) with Dan Fain.

2 Filtered feature coefficients

Since the computation of Delta and Delta-delta coefficients is based on a rough filtering process, we decided to make exploratory tests on filtered feature coefficients on a small data based I usually use for quick test of new algorithms. These results showed that coefficient low pass filtering leads to some improvements of the recognition scores.

3 Trained filtering of feature vectors

The idea to design an optimal filter led me to develop a new algorithm for training the HMM parameters of phonemic models simultaneously with those of the filter. I wrote the draft of a paper to describe this new training algorithm based on a parametric description of ideal FIR LP, BP and HP filters from their cut-off frequencies. The text of this paper as of the end of the workshop is appended. The first effort was to build the algorithm for a LP filter common to all states and all parameters. Later, the theory and the programs were extended to dedicated filters for each state and each entry of feature vectors. Tests were mainly performed on the small database since the objective was to program and test new algorithms.

The convergence of the algorithm was quite difficult to obtain for the following reasons:

- due to the fact that, at each training iteration, the filter is modified, the filtered features can be amplified with the updated filter gain. The global score could then show irrelevant improvements due to the gain updates. A normalization constraint on the impulse response power was included in the process.
- the cut off frequencies were the solutions of a fixed point iteration process. However, several solutions exist corresponding either to a maximization or minimization of the scores. Only a maximization of the global probability had to be detected by considering the convexity of the curve.

The draft of the paper describing the new algorithm contains results obtained respectively with fixed cut-off frequencies (untrained) and with trained cut-off frequencies. The convergence has still to be improved and no significant conclusions can yet be drawn up to now. Additional work requires:

- improve the search of the optimal cut-offs: best maximum among several solutions and rejection of minima.
- extend the results to a larger data based and eventually to SWITCHBOARD.

The program requires substantial modifications to the Viterbi algorithm. For that reason, HTK was not used. Work is going on in my lab to finalize this work and publish the results on large data bases.

4 Publications and Communications

During the workshop, I made a presentation of the paper and gave the available results. At the end of the workshop, the advancement was presented together with the other results of the acoustic team. Texts of the papers and transparencies used for the presentation were archived in Prof Andreiou files.

During the workshop I attended to lectures of invited speakers and also to most of the presentations of the other teams.

5 Evaluation of the workshop

This kind of workshop is certainly the best opportunity to collaborate I ever had. It is by far much more efficient than a workshop where people only have discussion during a couple of days. The excellent workstation network contributed significantly to make the work very extensive since we never had any failure that stopped our tests.

However some improvements could be brought either on the definition of the objectives or in the preparation of the work.

Regarding the objectives, one may wonder if increasing the recognition scores on SWITCHBOARD using mainly HTK the flexibility of which is questionable is the best and most meaningful research objective. Clearly, speech recognition research follows this kind of process where researchers define new targets in terms of complexity of the data bases (mainly size and speaking modes) but new ideas do not necessarily come from work on huge and huge databases. If the objective is to obtain new ideas, the researchers should make tests on small or medium sized data bases and more students could be in charge of the transfer of this ideas to specific LVCSR programs. This comment applies more specifically to our group on acoustic modeling.

Regarding the preparation of the workshop, I have two comments:

- Part of the participants were already active at the previous workshop using the same database and tools. For these participants, the data

bases were familiar and easy to access. As a consequence, it is important to give new participants more informations on the prerequisites.

- Research topics should be better identified and discussed long before the beginning of the workshop. In that way, participants have time to prepare by readings and discussions with future co-workers to the expected work.

The accomodation conditions were excellent. Major efforts were made to make this 6-week stay at Johns Hopkins comfortable and efficient. Lodging was OK at a walking distance of the lab. The organization of some entertaining activities like movies or group dinners was an excellent initiative. The everyday activity was perfectly organized and advertized. Support and administrative staff was very kind and helpful.

Appendix: Paper written during the workshop