REPORT ON DOCTORAL THESIS

Title of the thesis: TRAP-Based Probabilistic Features for Automated Speech Recognition

Ph.D. candidate: Ing. František Grézl

Reviewer: Doc. Ing. Petr Pollák, CSc.
Czech Technical University in Prague,
Technická 2, 166 27 Praha 6, Czech Republic

The work of František Grézl deals with the extraction of features for automated speech recognition based on temporal patterns (TRAPs). The importance of this research is evident. Automated speech recognition (ASR) or more generally speech technology came in real life in many applications and looking for features describing speech signals reliably and robustly in all relevant details is really important task. TRAP features represent then perspective description of speech signal which is currently studied very intensively in the speech research community over the world.

The work is well and logically structured into 9 chapters and presented main thesis goals can be summarized following way.

1. *theoretical goal* - to provide a general overview of TRAP features and to study their possible modification from several further discussed points of view,
2. *experimental and application goals* - to realize the testing of designed techniques and to implement them into real ASR systems.

It can be said, that these goals of the thesis are dissertable and that they were well accomplished. I see the original contributions of this thesis in following realized tasks.

- Firstly, it is the study and the evaluation of possible modification of basic TRAP features from several aspects as dimensionality reduction, critical-band spectrogram (CRBS) modification, and band merging. Presented techniques have slightly different contribution to base-line system, i.e. band merging and dimensionality reduction have brought better improvement of WER than CRBS modifications.
- Secondly, studied combinations of TRAP systems showed that improvement description of speech can be obtained by methods from simple multi-streams combinations based on averaging or vector concatenation, up to more sophisticated approaches based on band-conditioned classification. The best technique from this group brought approx. 30-40% relative improvement of WER (lowest WER=4.1) with respect to baseline TRAP system (6.6%).
- Thirdly, the testing of behaviour of TRAP features on noisy speech is very interesting results. Performance tests of new designed techniques in noisy environment did not show significant contribution in the case of noisy speech with respect to baseline TRAP system. On the other hand, realized experiments proved that base TRAP features itself overcome significantly standard MFCC features, especially for SNR as 15, 10, or 5 dB (these values can be assumed as typical for noise level in standard medium quality speech collected in real environment).
- Fourthly, combination of TRAP features with standard cepstral features and its application in the LVCSR system was realized. Approx. 2% improvement of WER was achieved by

combination of MFCC and TRAP features based on Heteroscedastic Linear Discriminant Analysis (from 47.5% to 45.4%) in the task of meeting speech recognition. Conversation telephone speech recognition was another task, where application of PLP and TRAP based feature combination brought improvement of WER from 37.2% to 33.6%.

- Finally, it can be said generally, that presented results proved that there is a potential to improve TRAP based feature extraction and proposed methods are possible solutions.

Originality of this thesis is proved also by many works of the author which were published at leading international conferences and workshops as: Interspeech, ICSLP, ICASSP, NIST Workshops, TSD. Related work realized within NIST speaker recognition evaluation 2006 were published in journal IEEE Transactions on Audio, Speech, and Language Processing.

Concerning the formal issues, the thesis is written very well. Problems are clearly and precisely explained. The bibliography is relevant and the most important ongoing international research is mentioned. Especially, I appreciate the conclusions of the theses completed by comments with respect to current related research over the world.

Overall, the thesis of František Grézl describes and demonstrates significant amount of realized work over his Ph.D. study and it shows his capability of independent and original research activity. From my point of view, this thesis overcomes many similar works, especially from the point of view good balance between theoretical work and applications of designed techniques in real systems. I am sure that the thesis contributes the progress in given research field.

On the basis of above mentioned facts, **I do recommend** the thesis for the presentation with the aim of receiving the Doctoral degree at Brno University of Technology.

For a discussion I have two additional questions:

– In chapter 3 you have discussed tests for statistically significant differences in obtained results. But some of presented results are bellow or very close to derived limits. Don't you think that designed techniques (or at least some of them) should be tested also on the basis of other data sets to confirm obtained results?

– You have realized the experiments with noisy speech data without any noise suppression technique which is for sure out of the scope of this work. But do you have some ideas how some noise suppression technique can be included into TRAP based feature extraction?

In Prague, 12th October 2007