Posudek oponenta diplomové práce

Student: Danko Michal, Bc.
Téma: Identifikace hudby, řeči, křiku, zpěvu v audio (video) záznamu (id 17411)
Oponent: Malenovský Vladimír, Ing., Ph.D., UPGM FIT VUT

1. Náročnost zadání **průměrně obtížné zadání**

2. Splnění požadavků zadání **zadání splněno s drobnými výhradami**

The student has successfully implemented a VAD system based on neural networks. His experiments were mainly concentrated on multi-task learning. The results are compared against a single baseline system, so I'm lacking comparisons against some existing VAD schemes. Also, some conclusions are too simplified and questionable.

3. Rozsah technické zprávy **je v obvyklém rozmezí**

4. Prezentační úroveň předložené práce **73 b. (C)**

The ordering of chapters is OK but in the introduction, some sections are irrelevant, e.g. Section 2.3.1. The biggest problem, however, is that the "state-of-the-art" is completely missing! Also, I'm lacking mathematical background for the multi-task learning (back-propagation).

5. Formální úprava technické zprávy **68 b. (D)**

The student definitely has courage to write the thesis in English because the level is poor. There are many grammatical errors such as starting sentence with "Which". The expressions are often misleading and superfluous. Symbols are sometimes defined after they are used for the first time, e.g. FAcc, Hit rate and BAcc. References are sometimes missing or incorrect, e.g. on page 123, the reference to Fig. 3.8 and Fig. 3.9 is incorrect.

6. Práce s literaturou **79 b. (C)**

The author often references monographies leading to "high-level" descriptions of knowledge that has been known to scientists since long time. Diagrams and illustrations are often taken from external sources instead of being created, hopefully in a better way, by the author himself. Nevertheless, the citations are correct and respect the established practise.

7. Realizační výstup **85 b. (B)**

The implementation looks OK. It's written in Python using several existing tools. However, the code suite from the attached CD cannot be run due to lack of input data.

8. Využitelnost výsledků

It's certainly good that the author successfully implemented the VAD based on neural network.The conclusion that adding auxiliary detection tasks in the NN training helps improving the detection accuracy of the VAD has a value. However, in order to re-use the results or conclusions for further research, the author should perform more rigorous analysis dealing with larger datasets, errors in labelling, different selection of categories and sub-groups of acoustic events, various number of subgroups, etc. Also, it would be good to test the system on some well-established noises, such as car, office, street, babble, cafeteria, ... and perform a thorough analysis in a non-stationary environment.

9. Otázky k obhajobě

1) Can you explain how you back-propagate the error in your multi-task setup? There is no mathematical background in your thesis.2) You often use the expression "events correlated to speech". Can you clearly explain what you mean and how you define such events?3) Do you have any explanation why under-sampling of data helps?

10. Souhrnné hodnocení **72 b. dobře (C)**

The student has shown that he can perform experimental work to achieve the objective of this thesis but I think he could have done more. The level of presentation is mediocre and the scientific impact of this work is not high. My overall evaluation summary is therefore: "AVERAGE".

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 31. května 2016

.....
podpis