Posudek oponenta diplomové práce

Student: Profant Ján, Bc.
Téma: Robustní rozpoznávání mluvčího pomocí neuronových sítí (id 21835)
Oponent: Rohdin Johan A., Dr., UPGM FIT VUT

1. Náročnost zadání **obtížnější zadání**

The task is difficult because
1, It involves many different datasets. These data sets are large and real, and therefore difficult to process.
2, It requires the student to understand and use many different quite complicated methods.

2. Splnění požadavků zadání **zadání splněno**

All assignments were fulfilled.

3. Rozsah technické zprávy **splňuje pouze minimální požadavky**

The thesis seems to meet the minimum requirement. Moreover, the thesis is written in a concise way so that every page is very informative.

4. Prezentační úroveň předložené práce **90 b. (A)**

Generally, the thesis is well organized and easy to read. The results of the experiments are presented in appropriate detail.
In a few cases I would have liked to see more details about the methods, e.g.,
* What was the adaptation method for PLDA?
* In the experiments about processing speed, in "fs2i" what does it mean that i=0? How is it different from skipping the odd feature vectors (fs1)?

5. Formální úprava technické zprávy **95 b. (A)**

Overall the thesis is very good in this respect. There are only minor typos which do not limit the comprehensibility of the work.

6. Práce s literaturou **85 b. (B)**

The student has used the available state-of-the art toolkits as starting point and improved upon them. Therefore the student did not solve problems that were already solved.
Generally, the thesis contains many relevant references. In a few cases though, I was missing some:
* In some cases, e.g. for bottleneck features, some representative work were cited whereas it could have been good to cite also the earliest work.
* It would have been good to cite works similar to x-vectors, e.g., d-vectors and the work by Gautam Bhattacharya "Deep Speaker Embeddings for Short-Duration Speaker Verification" in Interspeech 2017.
* Chapter 2.5 (before 2.5.1) This part is very similar to Bishop's presentation of Neural networks. This should have been stated.

7. Realizační výstup **95 b. (A)**

The experiments are carefully planned, systematic and properly evaluated. The experimental set-up is well documented.

8. Využitelnost výsledků

The thesis systematically evaluated previously proposed methods as well as some proposed modifications of, e.g., the neural network architecture. The results will definitely be useful for us in future speaker recognition evaluations.

9. Otázky k obhajobě

* What were the most important things that made x-vectors work so well compared to other approaches/architectures for DNN embeddings?
* Do you think more end-to-end approaches with joint training of embedding extractor and feature extractor or the backend will beat the x-vector approach in future?
* Regarding Section 5.3.2. If K is very large, the clusters will be formed by very few segments and their representative x-vector could be quite random. Doesn't this mean that one of the cluster may match the enroll speaker very well just by chance? Could a better approach be derived?

10. Souhrnné hodnocení **90 b. výborně (A)**

The most important part of the thesis is the experiments, i.e., Chapter 4 and 5. Here, scenarios similar to the most

recent and (expected future) evaluations are thoroughly evaluated. Several different systems are evaluated and best ones are very competitive for the evaluated conditions. The experiments are systematic with clear conclusions.

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

V Brně dne: 6. června 2019

Rohdin Johan A., Dr.
oponent