Review of Bachelor's Thesis

Student: Odehnal Ondřej
Title: Unsupervised Evaluation of Speaker Recognition System (id 24991)
Reviewer: Plchot Oldřich, Ing., Ph.D., DCGM FIT BUT

1. **Assignment complexity** — **less demanding assignment**
   This work concentrates mainly on applying known clustering methods on top of the scores obtained from a given pretrained speaker recognition system. The work requires extensive experimentation and and data analysis rather than core algorithmic implementation. A careful review of contemporary methods for clustering and its evaluation is also necessary for a successful completion of this assignment.
2. **Completeness of assignment requirements** — **assignment almost fulfilled**
   The main goal of the work, which is unsupervised evaluation/assessment of the speaker recognition system was fulfilled and very well experimentally analyzed on a commercial production system. I have slight reservations to the fulfillment of the second point of the assignment, which should be designing the methods of automatic quality control in evaluation data. The work contains section 2.7 discussing the obvious attributes affecting the quality, but there are no experiments later in the work to asses which measures are the most important and how do they affect the accuracy of the unsupervised system evaluation.
3. **Length of technical report** — **in usual extent**
   The length of the work is within normal parameters.
4. **Presentation level of technical report** — **80 p. (B)**
   The thesis is logically structured with rather extensive theoretical introduction. Chapters are logically following each other. Some subsections of the theory are probably not necessary for sufficiently introducing the necessary apparatus to the reader.
5. **Formal aspects of technical report** — **90 p. (A)**
   This thesis is written in a good English. Typography is without any major issues. Perhaps, I would recommend to use smaller figures and the font that is similar in size to the font in the thesis text.
6. **Literature usage** — **100 p. (A)**
   The citations are in order and used extensively throughout the work.
7. **Implementation results** — **75 p. (C)**
   The output is the methodology supported by the extensive experiments. Experiments are well designed, described and executed. Results of the experiments are discussed. As stated above, I am lacking some experimentation with quality measures.
8. **Utilizability of results**
   The main focus of the work is experimentation and evaluation of clustering methods in order to evaluate their ability to estimate the performance of the systems on an unlabeled dataset. I believe that the methodology is usable in the practice. Both for development/scientific systems and for commercial use. The work does not necessarily bring new discoveries, but is practically useful.
9. **Questions for defence**
   In your work you focused on predicting the EER which is a calibration independent technique. Would you have some comments or results that would address the actual performance such as actual DCF as defined by NIST? In other words would you be able to estimate quality of calibration in an unsupervised way?
10. **Total assessment** — **80 p. very good (B)**
    Overall, I liked the whole work. It was well experimentally executed with carefully analyzed results. The outputs are interesting and useful. I was lacking an experimentation addressing the quality measures and therefore I recommend the mark B.

In Brno 3 June 2022

Plchot Oldřich, Ing., Ph.D.
reviewer