

## Review of Bachelor's Thesis

**Student:** Gregušová Sabína  
**Title:** Evaluation and Optimization of Computational Costs in Speaker Recognition Systems (id 23008)  
**Reviewer:** Silnova Anna, MSc., DCGM FIT BUT

- 1. Assignment complexity** **average assignment**  
The task of this thesis is moderately challenging. The student is supposed to propose an evaluation metric that can be used to assess both the accuracy and the computational costs of automatic speaker verification system. The task requires understanding the state-of-the-art processing pipeline used for speaker verification. Moreover, the student is supposed to modify the existing pipeline to optimize the performance of the system in terms of the newly proposed metric.
- 2. Completeness of assignment requirements** **assignment fulfilled with minor reservations**  
Although the particular realization is somewhat simpler than what was proposed in the task, the objectives of the project are fulfilled.
- 3. Length of technical report** **within minimum requirements**  
The thesis is somehow on a short side. It is not a problem for the experimental part of the work, which is described sufficiently well. However, the theoretical part could be expanded to explain better the concepts used for the experiments.
- 4. Presentation level of technical report** **55 p. (E)**  
The narrative of the thesis is quite chaotic. Especially it is true for Chapter 2, where the author gives a description of the existing approaches in speaker verification. For the reader not familiar with the topic it would be almost impossible to understand what are the processing steps, in which order they are applied and why they are needed. Moreover, there are some factual errors that complicate understanding even more. For example, the section on Gaussian Mixture Models has errors in the provided formulas, does not explain how exactly GMMs are used in speaker verification or why their description is at all needed in the scope of this work. The same could be said about the section on Neural Networks and a few others.  
Consequently, when some of these theoretical sections are referenced in the rest of the text, it only confuses the reader.
- 5. Formal aspects of technical report** **75 p. (C)**  
The thesis is written in English. There are few typos and grammatical errors but they do not prevent the understanding of what was done. Plots and graphs included in the experimental part of the work greatly help the reader to follow the text.
- 6. Literature usage** **75 p. (C)**  
The citations are sufficient and properly reference the existing work.
- 7. Implementation results** **70 p. (C)**  
The developed software is rather simplistic and somewhat lacks more descriptive documentation. However, the topic did not imply any excessive software development. Mostly, reusing and slightly modifying the existing open-source Kaldi toolkit was required. Moreover, as the project is mostly research-related, there is no need for production-quality software.
- 8. Utilizability of results**  
The work aims at formalizing common concerns about the computational complexity and time requirements of the speaker verification systems. It proposes the ideas and provides some initial experiments which pave the road for the future research in this direction. In future, it can be used as a reference baseline if some more elaborate methods are proposed.
- 9. Questions for defence**
  - You propose a Modified DCF metric and Time Constrained Protocol. Do you think they are interchangeable or there are tasks when one should be preferred over the other? If yes, give an example.
  - In your work, you present two approaches to truncate the input utterances to reduce the processing time(A and B as referenced in the text). How do you think, what are the reasons for the approach B (limiting the number of speech frames for x-vector extraction) to outperform A (simply truncating the audio)?
  - Have you considered other strategies to reduce the processing time except for those two described in the thesis? If yes, what were they and what advantages and disadvantages they have compared to those that you've used?

**10. Total assessment**

**65 p. satisfactory (D)**

The work successfully addresses the main objectives posed in the task description. The author indeed proposed evaluation metrics that take into account not only the verification accuracy of the system but also the time required to perform a decision. Also, two approaches were proposed to reduce the processing time. These are evaluated using the proposed metrics. One of the approaches is shown to be superior to the other. However, the proposed approaches are somewhat simpler than the ones mentioned in the task description. Moreover, the work heavily relies on the existing techniques and evaluation metrics used for speaker verification. Poor description of these components does not allow the unprepared reader to fully understand what was done. Moreover, it even does not allow to assess what level of understanding the author herself has in those concepts that she attempts to describe.

Chapter 2 is the one that has the most negative impact on the total grade.

In Brno 9 June 2020

Silnova Anna, MSc.  
reviewer