

I. IDENTIFICATION DATA

Thesis title:	Methods for speaker identification from an acoustic signal
Author's name:	Mohammed Ali
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Doc. Ing. Petr Pollák, CSc.
Reviewer's department:	Department of Circuit Theory

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	challenging
<i>How demanding was the assigned project?</i>	
Given assignment represents a challenging task. It is not a simple task to manage the theory of speaker identification, deep learning approaches, as well as an orientation in available toolkits, especially, for the thesis of bachelor level.	

Fulfillment of assignment	fulfilled
<i>How well does the thesis fulfill the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.</i>	
The assignment was fulfilled, author has realized a survey of techniques used for speaker identification and within the experimental part he proved an applicability of selected DNN-based system.	

Methodology	correct
<i>Comment on the correctness of the approach and/or the solution methods.</i>	
Based on realized survey, the author focused his attention correctly on x-vector-based speaker identification as nowadays standard and ECAPA-TDNN architecture as a representative of other advanced DNN-based system. After an initial comparison of both systems, he realized a testing of ECAPA-TDNN (as a better system) under real-life conditions. For this purpose he has completed representative set of evaluation speech data from publicly available databases, noisy speech signals were obtained by augmentation using clean speech and noises from other publicly available resources which is standard procedure widely used within such an experiments. In the end, the experiments with various languages and sensitivity to artificial generated speech were realized as well.	

Technical level	B - very good.
<i>Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?</i>	
<p>Author explains clearly the theoretical background of speaker identification. Maybe, some parts could be more brief, e.g. description of anatomy and physiology of speech production, similarly for sampling, details of GMM training, UBM-GMM, and i-vector systems, especially, when just DNN-based techniques are analyzed within further experimental part. I also found several slightly inaccurate formulations, e.g. discrete time (n) must be used in equation (1) as pre-emphasis filter is digital FIR filter, also a weighting window itself realizes just minimum smoothing of the spectrum (description of eq. (2)), etc. On the other hand, this part is very intelligible and given description proves, that author is familiar with basic problems in the field of speaker recognition.</p> <p>Realized experiments are described clearly, crucial information about system and used evaluation data are presented. The amount of realized experiments is not huge, however, it was enough to describe well the behavior of analyzed systems in real-life conditions according to the assignment. I missed a little more detailed discussion about pre-trained models mainly from the point of view coverage of acoustic conditions of used enrollment and testing data. I also missed better description of features used at the input of DNNs. Were 40 FB coefficients always used? Further details like boundaries of full frequency bands are also very important, especially, when various sampling frequencies are used.</p>	

Formal and language level, scope of thesis

B - very good.

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

Concerning formal issues, the thesis has standard and logical organization, the extent is sufficient, the language is good as well. I do not see too many grammatical errors, on the other hand, some typographic incorrectnesses can be found. Typically, the usage of uppercase vs. lowercase letters should be unified, also text could be justified at both sides, the format of equations is slightly worse – currently it is acceptable, of course, but for your further technical report I recommend to use Latex typesetting. The presentation of obtained results could also be better, many lines in figures 13-20 overlap a lot - the range at y-axis could be optimized or some results could be presented numerically in tables as well.

Selection of sources, citation correctness

A - excellent.

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

Author cites relevant resources. The presented list of references is very large, strongly above standards for bachelor thesis. All cited references contain mandatory items, however, the unified format is not always used, e.g. for author names.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

1) I missed English in your experiments with multiple languages (maybe also some other non Asian language). What was a motivation for this choice? Also couldn't be differences in results for particular languages influenced by given selection of rather small amount of speakers in used testing data per language? To confirm observed trend in obtained results, couldn't be suitable to repeat this experiment still with other 10 speakers per language?

2) How did you compute FB features for signals with various sampling frequency in your experiments analyzing sensitivity to artificially generated signals? Did you use the same setup or did you modify it for particular sampling frequencies? I think that this can play very important role when signals with different sampling frequencies are processed. By the way, wouldn't be better for this purpose to realize speaker verification task to guarantee that artificially generated speech is rejected?

Finally, author has realized nice work and he manage to learn a lot from the field of speaker recognition. On the basis of above mentioned notes and minor remarks I classify the thesis by the grade B - very good.

Date: 2.9.2022

Signature: