

## Recommendation for the defense of Ph.D. thesis

**Ph.D. candidate:** Anna Silnova  
**advisor:** Lukáš Burget

I am happy that Anna Silnova has submitted her thesis under my supervision, and that her long-years efforts have converged to the submission of the doctoral dissertation at FIT BUT.

Anna's thesis is in the area of speaker recognition and speaker diarization. Speaker recognition and diarization techniques usually make use of speaker embeddings – fixed-length vector representations extracted from speech utterances. The focus of Anna's thesis is dealing with uncertainty about the information encoded in the speaker embeddings. One of the main contributions of Anna's thesis is development and experimentation with a variant of Heavy-tailed Probabilistic Linear Discriminant Analysis (HT-PLDA) allowing for a fast evaluation of speaker verification scores. HT-PLDA is first presented as a generative model together with a Variational Bayes algorithm for its training. Subsequently, HT-PLDA model training using many different discriminative training criteria is explored. To my knowledge, some of the discriminative criteria are very novel and were not previously considered for machine learning before. For some of the discriminative criteria, training based on advanced sampling techniques was proposed (e.g., Smart-Dumb/Dumb-Smart algorithm), which is also a very novel approach in the field. The final part of the thesis focuses on designing a neural network architecture for extracting probabilistic embeddings – i.e., embeddings that have the form of a (parametric) probability distribution rather than being just a point in an n-dimensional space. This allows the embedding to directly model the uncertainty about the speaker information contained in the input speech utterance. For each proposed model and technique, an extensive and careful experimental evaluation is carried out to assess its effectiveness.

In my opinion, Anna's thesis represents a solid contribution to this speaker recognition and diarization area, but as indicated above it can find its applications in other areas of machine learning. Based on Anna's prior publications, some of the proposed models/techniques have been already well adopted in the field and found their way even to commercial applications (e.g. HT-PLDA model). I believe that the complex problems, models and algorithms are described in the thesis with the proper mathematical rigor and yet in quite accessible ways, which makes the thesis a good reference document for the research community. It should be noted that many of the ideas and models presented in the thesis were suggested by Dr. Niko Brummer (currently with Amazon), who is a co-advisor of Anna's thesis. Still, significant elaboration on the ideas, further development of the models, and all the experimental work was carried out by Anna herself. In my opinion, the thesis demonstrates Anna's extensive knowledge and orientation in the field.

The thesis actually covers only part of Anna's activities, and it does not include Anna's earlier works (e.g works on age estimation or deep neural network features for i-vector/PLDA-based speaker recognition). She has been involved in numerous speaker- and language-recognition-related projects running in the group. Anna has participated in several top-level multi-week research workshops, namely ASRWIS 2016 in Stellenbosch, South Africa, SCALE 2017 in Baltimore, USA and JSALT 2020 held remotely. In 2019, she spent 4 months at SRI International (Menlo Park, CA), a research organization founded by Stanford University, that has one of the most famous US speaker recognition groups, working with Mitchell McLaren and Luciana Ferrer on robust speaker recognition. Evaluations and challenges are key activities helping our lab to advance its know-how and assess objectively where we stand in the state-of-the-art. Anna participated in many such endeavors: NIST SRE 2016 - 2021, NIST LRE2017 - 2022, VoxCeleb 2019, VoxConverse 2021, SDSV 2020, and others, and in all of them, her contribution was very important for the overall success of our team. Anna obtained the Jack Godfrey's Best Student Paper Award at ISCA Odyssey 2020 The Speaker and Language Recognition Workshop - a bi-annual prestigious workshop of the international speaker and language recognition research community.

Anna is a great team-mate – with her strong mathematical background and friendly attitude, she is always ready to share with others her deep insights into machine learning and extensive know-how in speech processing. She is never afraid to start working on a new challenging task with an uncertain outcome. I really appreciate her sense of responsibility and persistence when working on such a challenging task. She succeeds where many others would get frustrated and give up.

To conclude, **I fully recommend Anna Silnova's Ph.D. thesis for the defense**, I wish her all the best in her professional and personal life, and I am looking forward to working with her in the future, and to enjoying the excellent espressos that she often makes us during the coffee breaks.

In Brno, November 8<sup>th</sup>, 2022

Lukáš Burget, Ph.D.

BUT Speech@FIT, Department of Computer Graphics and Multimedia

Faculty of Information Technology, Brno University of Technology

Božetěchova 2, 612 66 Brno, Czech Republic

Tel: +420 5 41141269 Fax: +420 5 41141290

<mailto:burget@fit.vutbr.cz>, <http://www.fit.vutbr.cz/~burget>