***Shanghai Normal University***
*No. 100 Guilin Road*
*Shanghai, 200234*
*People's Republic of China*

Professor, Yanhua Long, Ph.D
Shanghai Normal University
Email: yanhua@shnu.edu.cn
Tel: +86 18502194179

**June 22, 2023**

**Review of habilitation thesis of Ing. Jiří Málek, Ph.D.**

To whom it may concern,

The habilitation thesis entitled "Compensation of real-world distortions in speech signals" focuses on three research topics that deal with compensating various distortions commonly found in real-world speech recordings. The thesis is a compilation of papers written by the habilitant between 2013 to 2022. He had undertaken several research projects provided by the funding agencies GAČR and TAČR, and the manuscripts were created as an output of these projects.

The main topic of the habilitation thesis is the extraction and enhancement of target speech from noisy mixtures under real scenarios. The thesis first summaries the proposed methods using spatial beamformers and cancellation filters to enhance the target speech under multi-channel microphone array conditions, especially when the target speaker is moving in the noisy mixture recordings. In addition, the habilitant also includes the works of using supervised speaker identification to guide the target speech extraction under blind source separation scenarios like CHiME. Then, works of single-channel speech enhancement techniques were also included in this thesis, the representative work was the denoising autoencoder with convolutional neural networks for speech enhancement.

All of these works are excellent, focusing on the challenging issues of speech enhancement and target speech extraction in complex noisy environments, and the publications about these works are in top journals in the field of speech, such as the IEEE Transactions on Audio, Speech and Language Processing, Signal Processing, etc. However, most of the experiments in these papers were conducted on publicly available simulated datasets. One question I would like to pose to the discussion is: "Which of these proposed methods have been applied to actual project engineering, how were they utilized, what specific challenges did they solve, and what were their overall impacts on the projects' progress?"

The second part of the thesis focuses on improving the robustness of automatic speech recognition. Dr. Málek has been involved in proposing new methods such as the

multi-condition training and data augmentation. These methods were original and proved effective to addressing the noise robustness of both under-resourced and conversational telephone speech ASR tasks. Moreover, although these works published in the year of 2018, the habilitant stated that they can easily be applied to nowadays end-to-end ASR frameworks. For this part, I have one question for the expert discussion: What's the advantage and disadvantage of the proposed data augmentation methods over the state-of-the-art techniques like specAugmentation, speed perturbation, reverberation simulation, etc? Do they show complementary information for ASR acoustic modeling?

The third part of the thesis presents Dr. Málek's contribution to the compensation of nonlinear distortions, one for the blind compensation of soft clipping distortion and the other is the nonlinear echo cancellation in speech recordings. These works were published in two conference papers, and have been verified under both simulated and real-world conditions.

In conclusion, the habilitation thesis provides a comprehensive overview of high-quality research works by Dr. Jiří Málek and his co-authors. Despite most papers being re-printed, Dr. Málek has published more than 45 papers in the field of speech signal processing, indicating his strong expertise in the area. Overall, based on this thesis, I confirm that the candidate Dr. Málek is truly deserving of the habilitation as an Associate Professor.

Best Regards,

**Dr. Yanhua Long**