ulm university universität
u ulm

Universität Ulm – Ulm University | 89069 Ulm | Germany

Communications Engineering – Dialogue Systems

Prof. Dr. Dr.-Ing. Wolfgang Minker

Albert-Einstein-Allee 43
89081 Ulm, Germany

Tel: +49 731 50-26254
Fax: +49 731 50-12-26254
wolfgang.minker@uni-ulm.de

PhD Thesis Review

Topic: Automated Lipreading with LipsID Features
Candidate: Ing. Miroslav Hlaváč

	Points * Weight	Σ	Remarks
Objective - Consideration of the framework - Motivation of the work - Definition of the individual field of research	**3,0 * 2** 3 3 3	**6,0**	This thesis to be presented by Miroslav Hlaváč affiliated with the University of West Bohemia aims at creating a novel visual feature set for automatic lip-reading systems. The ultimate goal consists of improving the performance of automatic speech recognition systems using such an additional visual information channel.
Solution of the task - Application of the state-of-the-art - Working method, systematics and determination - Approach - Completeness of the solution - Scientific quality (relevance) of the results	**3,6 * 3** 4 4 4 3 3	**10,8**	The main contributions of the thesis may be summarized as follows: Based on the thorough analysis of existing visual speech features Miroslav Hlaváč has developed a new feature set for training a neural network enabling visual/audio-visual speech recognition. He also proposed a method for appropriately extracting these features. The new feature set has been implemented into a state-of-the art system for visual speech recognition for performance evaluation. Miroslav Hlaváč has consistently shown concerns on the reproducibility of his work with respect to the state-of-the-art.
Written presentation - Presentation of the aims of work - Presentation of the theoretical background - State-of-the-art description - Presentation of the results - Presentation of the progress achieved - Evidence to alternative solutions - Recommendations and perspectives - Structure of the document - Clarity and intelligibility - Clarity (tables and graphics) - General appearance - Appropriate length - Factual correctness - Coverage of the bibliography - Appropriateness of the cited literature	**2,3 * 2** 3 3 3 3 3 1 1 3 2 2 2 2 3 2 2	**4,6**	Mr. Hlaváč has carried out a thorough state-of-the-art analysis. He has demonstrated that his research is significant for the community, original and novel. The research objectives are well formulated and the relevant background theories required for the understanding of the thesis have also been introduced. The document is complete and well structured. To a large extent it is clearly written and correct in its form and contents.

Significance of the work - Contribution to the progress - Obtained findings o technical o scientific - Industrial feasibility - Societal benefit	**2,6 * 3** **2** **3** **2** **3** **3**	**7,8**	The presented work can be considered significant since visual speech recognition may be used in a wide area of applications and by a large variety of user groups. The presented results reflect the quality of the submitted work and would certainly merit a higher number of publications in top-ranked international conferences and well-established journals. The thesis concludes with an exhaustive summary, discussion and some future perspectives.
Overall	**2,9**	**29,2**	

Basis for assessment

Degree of fulfillment	Points
outstanding	4
above-average	3
average	2
below-average	1
insufficient	0

To summarize, in his thesis work Miroslav Hlaváč investigates and evaluates novel visual features to be integrated into a visual/audio-visual speech recognition system. The general aim is to increase the recognition results.

Based on the document that has been made available to me for appraisal, I can approve the thesis of Ing. Miroslav Hlaváč for defence.

Ulm, November 27th, 2019

(Prof. Dr. Dr.-Ing. Wolfgang Minker)

Review of Dissertation Thesis

Automated Lipreading with LipsID Features

submitted by Ing. Miroslav Hlaváč

at the Faculty Faculty of Applied Sciences, University of West Bohemia

The thesis was completed in 2019. It is written in English and has 107 pages.

Scientific background of the thesis:

Automatic Speech Recognition (ASR) is a well-established branch of modern computational science and developing and designing of systems for visual or Audio-Visual Speech processing and Recognition (AVSR) is one of the subareas of this branch. It is useful to use visual part of the speech namely in Large Continuous Speech Vocabulary (LVSCR) systems where information from visual part can improve resulting recognition rate. Very important part of AVSR is visual speech features extraction. The large variety of different approaches gives the researcher a great opportunity to study visual speech features extraction methods and algorithms from different points of view. Usually, visual features performance and robustness are the most important criteria discussed in comparative studies.

Main objectives of the work:

In his thesis, Miroslav Hlaváč focuses on design of a new method for visual speech features extraction (LipsID) based on state-of-the-art methods where artificial neural networks have been used. The state-of-the art is described well but the description of visual speech features is relatively poor. Sometimes, Mr. Hlaváč runs into unnecessary details. In his thesis, he gives a good overview of all aspects concerning of AVSR in Chapter 6, where he proves adequate knowledge of all related work done by other authors. In the chapter that precedes it (i.e. in Chapter 5), he briefly summarizes basic principles of artificial neural networks. Chapter 4 describes statistical models (active shape model, active appearance models...) which can be used in AVSR task. The most commonly used datasets are presented in Chapter 7. The third part of thesis deals with visual speech features analysis, new LipsID visual features and lipsreading experiments (Chapter 10). Chapter 10 is the biggest weakness of the dissertation thesis. There are only few experiments (in my opinion) and the experiments are described very shortly in approximately 6 pages. Very similar range has Chapter 7 – datasets. I wonder why designed LipsID features have not been tested on other datasets and why LipsID features have not been compared with other commonly used (in AVSR) visual features.

Methods and methodology:

The thesis demonstrates that Miroslav Hlaváč has become familiar with all the basic concepts and algorithms used in AVSR. The new described method (LipsID) for visual features extraction is well designed and it is shown from several results. LipsID features can improve resulting recognition rate in comparison to LipNet results.

Results and achievements:

In his thesis, Miroslav Hlaváč presents results from several experiments to demonstration of his system for visual features extraction based on artificial neural networks. The results which are shown in tables (Chapter 10) are good enough but I would expect considerably more experiments in such dissertation thesis. Otherwise the results are presented clearly, with reasonable discussion. Mr. Hlaváč writes in the conclusion of his dissertation thesis that he would like to developed a lipreading end-to-end system which will be available on Git so everyone can replicate the

experiments. I suppose that it will be a difficult job, but I am wondering how this resolution will turn out.

Scientific contribution of the thesis:

The thesis comes with a new idea for visual speech features extraction. The algorithm and method with LipsID features appears useful. The text of the thesis is clear, well structured and contains minimum errors. Some of the taken images (e.g. figure 17, 21..) are relatively small and it is quite difficult to see what is there.

As to Miroslav Hlaváč's publishing activity: His list contains 16 published works, which looks like a good number, *however, none of them occurred on a major and prestigious speech conference*, like Interspeech or ICASSP.

Conclusion:

The author of the thesis proved to have the ability to perform research and to achieve original scientific results. I recommend the thesis for presentation with aim of receiving the degree of Ph.D.

Questions:

1) Why have not been LipsID features compared with other commonly used (in AVSR) visual features in single experiments?

In Liberec, November 22nd, 2019

doc. Ing. Josef Chaloupka, Ph.D.
Institute of Information Technology and Electronics
Faculty of Mechatronics, Informatics and Interdisciplinary Studies
Technical University of Liberec