

To

Sylva Sadovská
Faculty of Information Technology
Brno University of Technology
Božetěchova 2
612 66 Brno
Czech Republic

Pittsburgh; December 1, 2014

Re: Evaluation of Michal Fapšo's thesis

Dear colleagues,

I am writing to you today in response to your request for an evaluation of the thesis "Query-by-example Spoken Term Detection" by Ing. Michal Fapšo, which you sent me on October 15, 2014. Thank you for giving me the opportunity and honor of commenting on Ing. Fapšo's PhD thesis; I thoroughly enjoyed the duty - the thesis contains an overwhelming amount of interesting results, so the lecture was informative in many aspects, and I am confident in my evaluation.

Search and retrieval involving speech and language in low-resource settings is a timely research topic, as indicated by the presence of the Babel project and the QUESST and NIST STD 2006 evaluations. During his work, Ing. Fapšo has either directly contributed, or provided fundamental techniques that were applied by others, to all these efforts, which taken together define the state-of-the-art in the field. His thesis presents a thorough analysis and comparison of several fundamentally different approaches to "query-by-example" search. I have not seen such a comparison elsewhere, and in fact, my own "overview" paper "Metze et al.: Language Independent Search in MediaEval's Spoken Web Search task. Computer, Speech, and Language, 2014", cites the candidate's work extensively and would have benefited from the know-how and insights presented here. The candidate's work therefore constitutes a sizable and relevant piece of work that describes several significant advances that the candidate invented or implemented himself.

I will not repeat the thesis document outline here, but I find the structure of the document concise and well balanced. In several places, the use of the English language could be slightly improved, but the comprehensibility is never impaired, so I regard this not as an issue. In light of the above, I agree that the candidate's three main contributions are as follows, and will discuss them briefly here:

An analysis of STD evaluation metrics

One of the main problems in current Spoken Term Detection (STD) research is the uncertainty on how to best evaluate them. Metrics such as DET/ ROC curves, TWV and its variants, EER, precision/ recall, or Cnxe all have their benefits and disadvantages. The proposed FOM (and its development-friendly variants) are well suited for research purposes, in particular for QbE STD, and have since been picked up by other research groups. The candidate's discussion of metrics in Section 2.3, and the wealth of results presented in the following section demonstrate excellent understanding of the STD problem, the underlying issues, and make a convincing case for using FOM (variants), which other researchers have since picked up on (QUESST is currently discussing metrics for the next year's evaluation).

Implementation, analysis and improvement of two WFST-based QbE systems

WFST-based STD systems, while generally performing very well, have in the past been applied only to situations where the target language was known, and a reasonable amount of linguistic knowledge is available. By directly generating phone lattices from multi-lingual phone recognizers, rather than word- or syllable-based recognizers, the candidate proposed a novel way of performing *language independent search* using WFSTs, with no dependence on language-specific knowledge sources whatsoever. The present work shows for the first time which assumptions and changes to make so that this approach is competitive to pre-existing solutions, while requiring fundamentally less resources to be available.

Comparison and analysis of DTW, GMM/ HMM and WFST QbE systems in several [...] setups

The correctness and significance of the previous result is convincingly shown on a number of tasks using a wealth of metrics as discussed above. Several competing approaches have been implemented, resulting in hard baselines and compared with the proposed novel approach. The results on MediaEval, where the BUT systems performed excellent overall and were the most robust submission on the evaluation data, demonstrate the rigor of the candidate's experimental method and the strength of the results that have been achieved.

I would also like to point out that the candidate has made parts of his code available on github, which greatly facilitates its use by others, and advances scientific progress. To conclude, I would therefore like to give the following responses to the concrete questions laid out in your request:

Q: Is the topic appropriate?

A: Yes, this is a timely and substantial contribution in a very active research field. Many of the techniques and comparisons presented by the candidate advance the state-of-the-art, as shown by the results achieved in the evaluations in which he participated (NIST STD 2006 and MediaEval SWS 2012).

Q: Is the work original?

A: Yes, the work is original. The candidate's meaningful evaluation of STD metrics is a significant contribution, on which others have already been able to build. Similarly, the candidate has shown great ability in collaborating with others to prepare meaningful comparisons and analyses of many different STD systems - more than he could possibly have implemented by himself only.

Q: Has the core of the doctoral thesis been published at an appropriate level?

A: Yes, in line with the collaborative nature of his work, the candidate has published an appropriate amount of papers in international peer-reviewed conferences (SLT, ACM Multimedia, INTERSPEECH), evaluation workshops (MediaEval), and international journals (ACM TOIS).

Q: Does the list of the candidate's publications imply that he is a person with an outstanding research erudition?

A: Yes, the candidate has authored and co-authored a number of significant papers on his thesis topic in relevant conferences over a period of several years. The relevance and volume of this scientific output speaks greatly to the quality of the underlying work, and the candidate's role in its creation.

My analysis is based entirely on the written document that I was provided with, and my knowledge of the MediaEval and NIST evaluations which I am familiar with. If possible, I would like the candidate to elaborate a bit on the following issues and provide answers in writing, or at the defense:

- There are other metrics for evaluating STD as well, for example *Cnxe*, see Luis J. Rodriguez-Fuentes and Mikel Penagarikano, "MediaEval 2013 Spoken Web Search Task: System Performance Measures", Technical Report-2013-1, Dept. Electricity and Electronics, University of the Basque Country, May 30, 2013, <http://gtts.ehu.es/gtts/NT/fulltext/rodriguezmediaeval13.pdf> - what is the candidate's opinion on this metric (advantages, disadvantages?), and how does it relate to the metrics he proposed and evaluated? Could the use of probabilities (rather than just detections vs non-detections) be useful for the analysis of STD systems?
- Some of the presented approaches are indeed inspired by Parada's work, as the candidate acknowledges. It would be good if there was an explicit discussion of the differences between these two works, and their respective advantages and disadvantages, to bolster the claims of novelty required for a thesis.
- Could the conclusion be rewritten and focused a bit? It starts with "In this work, we have presented three query-by-example STD systems and two query-by-example STD *baselines*" - this formulation seems to be just an oversight, but a thesis' contribution is to go beyond the

baseline, and improve a task. So, it would be good to re-iterate at this point the unique and novel contributions by the author quantitatively (not just qualitatively), and give examples or answers to the question “what is possible now that was not possible before”?

In sum, as I hope the above discussion shows, *I am convinced that the total of the presented work in context and the quality of the doctoral thesis meet the requirements of the relevant PhD degree*, so such title should be awarded. I would recommend the candidate submit an abridged, first-authored version of the thesis to a journal, in order to make its contents and his contributions known to a wider audience. Please do not hesitate to contact me with any further questions you might have, I would be happy to elaborate and support the candidate’s case in any way.

With best regards,



Prof. Dr. Florian Metze