

**Ing. Miloš Cerňak, Ph.D.**  
**Voice Technologies & Systems**  
**R&D Laboratory, IBM**  
**Praha, Czech Republic**

## **Review of Doctoral Thesis**

**Author:** Martin Karafiát  
**Title:** Study of linear transformations applied to training of cross-domain adapted large vocabulary continuous speech recognition systems  
**Reviewer:** Ing. Miloš Cerňak, Ph.D.

### **1. General description and the topicality of the thesis**

The thesis consists of eight chapters, the presentation and original contribution is claimed immediately at first pages of the thesis. Chapters 5 and 7 are devoted to description of the claimed contributions. The thesis is written on 68 pages, followed with 55 references. Author concludes the thesis on last 2 pages. The structure of thesis conforms to principles and requests to the structure of scientific thesis. The description of the work-around sounds sometimes very briefly, but the thesis is written rather concisely.

The author has studied and used appropriate number of bibliography sources used and quoted in the thesis. It is the evidence of the deep theoretical knowledge and very good orientation in the problem discussed in the thesis.

Concerning to the topic of the thesis, I appreciate the topic of linear transformations and their application to meeting transcriptions. Meeting transcriptions as a part of Rich Transcription evaluation belong to the pinnacles of speech recognition, something that we have aspired to from the beginning. With transcription, anything that is said can be converted to the text, and further processed.

The author has performed good orientation a speech recognizer composition (Chapter 4) and wide knowledge of linear transforms in feature-space (Chapters 3, 5 and 7).

**The thesis fulfils the formal requests on good level. The topic of thesis is current and relevant in the context of up-to-date research in large vocabulary cont. speech recognition.**

### **2. Claims of the thesis**

The aims and claims are described in the introduction. Chapters 2-3 and 6 are theoretical background of the research. Chapter 4 presents baseline recognition system. Two main claims are visible throughout the thesis:

1. Proposal and testing MAP-Smoothed and Silence Reduced HLDA with the aim to increase robustness in meeting speech recognition. Chapter 5 investigates the claim.
2. Linear transforms applied to NB-WB adaptations with the aim to effective porting of telephone speech DB into meeting domain. Chapter 7 explains the used approach.

**Aims and methods are clearly described; author represents the ideas and knowledge with sufficient theoretical background. The aims were fulfilled, methods of research work are appropriate to the aims and claims formulated in the thesis.**

#### **4. Publication activity**

The author published his work from both parts of the thesis on top scientific conferences. I expect that this work will be extended into future journal papers. In the publication list there are more multi-authored works (7 – 11 authors), where it is really hard to estimate Martin's contribution. On the other hand, in the important works dealing with the topic of the claims of the thesis, Martin is the first author on the author list.

From the publication list I see that he is able collaborate in the both local and international scientific teams. Reading his thesis I can just say that he is able to use acquired knowledge from international projects in his own research.

The core of the thesis was sufficiently published. Martin presented in his work a great scientific erudition.

#### **5. Questions**

1. Why improvements obtained by smoothing techniques in HLDA are so quite small? Do we reached "technological limit", or do you expect further improvements in the future?

2. It is well known that the problem of ASR robustness can be solved using 3 approaches: the first approach works on signal level (speech enhancement), the second approach works on feature level (feature normalization), and the third one works on models (adaptation). You used the adaptation for porting of telephone speech database into meeting domain. Then, you combined the adaptation with HLDA in feature level, improving WER of the recognition task.

Finally, my hypothetical question: Including also present techniques on signal level; can it become further improvement of the task (in the terms of a specific method combination as you presented e.g. in HLDA adapted system)? If not, why?

#### **6. Conclusion**

In my opinion, the thesis by **Martin Karafiát** fulfils all the conditions for gaining the PhD. degree in Information technologies; therefore it is recommended.

Praha, 05.01.2009