Review of a PhD Thesis Proposal

Title of the proposal: *Handwritten Optical Music Recognition*
Author of the proposal: Jan Hajič jr.

The thesis proposal submitted by Jan Hajič is in fact an extensive and detailed report on his research in the field of Optical Music Recognition (OMR) that aims to automatically read music scores. In Section 1 the author first explains the motivations and details of the OMR task. He describes main difficulties and different approaches to the task. He states that the overall goal of his thesis is to create an OMR system capable of offline reading handwritten common western music notation (CWMN). The task of OMR systems is very complex and generally includes several main stages (see Section 2):

- image preprocessing,
- music symbol recognition,
- musical notation reconstruction,
- final representation.

Among relevant methods that the student takes into account are neural networks for OMR, multimodal learning models combining music images and audio data, user feedback, evaluation techniques, and ground truth datasets. Hence the handwritten OMR is a topic that encompasses a wide variety of difficult subtasks, and in this sense the thesis goal seems to be rather ambitious. As I am not an expert in the given field, I do not try to assess whether the declared goals of the thesis is realistic. However, the proposal as a whole is clearly written and proves that Jan Hajič is familiar with the newest relevant scientific contributions in the field, and is able to analyze the state-of-the-art and the key issues.

In Section 3 Jan Hajič describes his important decision about further development and reacts to the crucial need of OMR benchmarking. He decided first to systematically build a ground truth dataset of handwritten music, with focus to two research questions:

- the definition of the ground truth for a music image,
- choice of data point in the benchmark dataset.

In fact, this research line finally led to establishing an original theoretical result in OMR objectives. Jan Hajič suggests that so called notation graphs can be used as a good ground truth for an OMR dataset and shows that notation graphs' information content is both necessary and sufficient for both reprintability and replayability of musical scores. While the idea of a notation graph is not new, the full imlementation seems to be an original contribution of the student. An important practical result of Jan Hajič's recent work is ground truth dataset MUSCINA++ thoroughly described in Section 3. In the next Section 4 the author analyses the evaluation issues and a data-driven approach to OMR end-goal evaluation. He searches for a suitable automated metric for OMR evaluation so that it really ranks as better systems those that should be ranked better. In fact, he describes his own already published proposal of a robust, cumulative, data-driven methodology for creating such a metric.

Future plans described in Section 5 focu on two experimental research lines:

- notation syntax-aware offline handwritten OMR, and
- augmenting it by including the audio modality.

The plan of recognition experiments is both clear and, in my view, demanding. In Section 6 the author concludes that while the work described in this thesis proposal does not implement OMR solutions yet, considerable contributions have been already made to OMR as a field, especially by providing a new kind of benchmarking: both evaluation procedures and, even more importantly, ground truth dataset.

My conclusion is that, without any doubt, the thesis proposal is well written, both in structure and clarity of argumentation and in the use of English. The reported achievemets seem to be a good starting point for the further development of an OMR system. The student is convincing also as to his future plans, yet the declared goals may be really demanding.

Prague, April 24, 2017

RNDr. Martin Holub, Ph.D.

Institute of Formal and Applied Linguistics