CIRMMT Inter-Centre Research Exchange Report

Student: Martha E. Thomae
Location: Department of Software and Computing Systems, University of Alicante
Exchange Supervisor: José Manuel Iñesta
CIRMMT Supervisor: Ichiro Fujinaga
Dates: June 10th – August 10th

For this research exchange, I stayed at the University of Alicante in Spain for two months where I worked in close collaboration with Jorge Calvo-Zaragoza and David Rizo. My current dissertation research involves the digitization, recognition, and encoding of the music content of Colonial Guatemalan manuscripts written in mensural notation. The goal of this inter-centre exchange was two-fold: (1) become familiar with the technologies developed by this research group to apply them to my dissertation research (2) begin work on a collaborative project applying deep learning techniques for the automatic translation of mensural notation into modern values.

Calvo-Zaragoza and Rizo have worked on an end-to-end approach for OMR based on deep learning—specifically using convolutional recurrent neural networks (CRNNs)—which turns the traditional multi-stage OMR workflow into a single step. This approach was recently tested on a Spanish music manuscript written in mensural notation (see Calvo-Zaragoza, Toselli, and Vidal 2019), the results of which are now being corrected using MuRET—a Music Recognition, Encoding, and Transcription tool still in development by Rizo. Given the relationship between Spain and Guatemala, I guessed that the CRNN model trained on Spanish polyphony could make the encoding process more efficient for my sources. This assumption turned out to be true—the performance of the pre-trained CRNN model was satisfactory and we even improved it further by expanding the training set to include a few pages of the Guatemalan manuscript. The data from these extra pages was generated by correcting the previous result of the CRNN using MuRET.

The end-to-end CRNN model returns the sequence of notes as a string of tokens, where each token encodes the label of the music symbol (e.g., C clef, quarter note, half note, sharp) and its position within the staff (i.e., the number of the line or space where the symbol lies) without any musical meaning. Rizo and Calvo-Zaragoza refer to this type of encoding as an "agnostic representation." This is defined in opposition to a "semantic representation", in which musical meaning is conveyed—e.g., a series of accidentals is reinterpreted into a key signature, the position of the notes in the staff is interpreted as the pitch of the notes, and notes followed by dots are reinterpreted into dotted notes. The agnostic representation includes enough information to determine the pitch and duration of the notes when dealing with common Western music notation (CWMN). However, this is not true when dealing with mensural notation. Contrary to CWMN, the duration of the notes in mensural notation is not absolute but rather depends on the preceding and following notes. In mensural notation, the same note shape can be used twice, even next to each other, and convey different durational values. In other words, the label of the note symbol is not enough to convey its duration. Because of this, in addition to the symbol label and pitch, the semantic representation of a sequence of mensural notes needs to include the durational value of the notes to capture its full musical meaning.

This leads me to the second goal of the internship, developing a deep-learning-based approach for the automatic translation of mensural notation which takes the agnostic representation of a sequence of mensural notes as input and transforms it into a semantic representation. An initial solution to this problem was proposed in my master's thesis (Thomae Elias 2018), where I implemented a heuristic approach following rules outlined in various treatises, including Apel's "The Notation of Polyphonic Music, 900–1600." In this internship, we approached this problem from a different perspective, aiming to determine if machine learning can deal with exceptions of the rules as well as scribal errors that might be present in the music sources.

For this automatic translation of mensural notation, I worked on the implementation of an encoder-decoder sequence-to-sequence model with attention. A sequence-to-sequence model aims to map a fixed-length input sequence into another fixed-length output sequence, where the lengths of the input and output sequences might not match. This model consists of two parts: an encoder and a decoder, both of which are made of one or more recurrent neural networks (RNNs). The encoder part reads the input sequence and produces a vector that summarizes the input. This vector is then fed into the decoder part. The decoder generates an output sequence based on this information. The sequence-to-sequence model has been used for translation problems, where the output sequence corresponds to the translation of the input sequence. In this case, we are using it for translating the agnostic sequences into their corresponding semantic sequences. The addition of an attention mechanism allows us to visualize which tokens of the input (agnostic) sequence were the most relevant in predicting each of the tokens of the output (semantic) sequence.

For training this model, pairs of agnostic and semantic sequences are needed as ground truth. I finished the implementation of the model and, given the absence of a large dataset in mensural notation to train the model, I used the Primus dataset (a large set of incipits in CWMN) to verify that the implementation was correct. Now we are working on generating the mensural notation pairs of agnostic and semantic sequences to test the performance of this approach—encoder-decoder sequence-to-sequence model with attention—in the translation of mensural notation.

This research exchange program pushed forward my dissertation research as well as expanded my knowledge of deep learning. I am thankful to CIRMMT for the financial support they provided so that I could have the experience of working closely with this group of experts, using deep learning to advance OMR and the technologies required to facilitate the encoding of mensural music. Future work will involve the recognition and encoding of the whole Guatemalan choirbook manuscript and the generation of the agnostic vs. semantic pairs of sequences for the training set of the implemented encoder-decoder model.