Music Score Page Turner

Chan Yat Long Ariel[1, *], and Cheung Tsz Hong[2, #]
[1]Sheng Kung Hui Tang Shiu Kin Secondary School, Hong Kong
[2]Po Leung Kuk Choi Kai Yau School, Hong Kong
**Contact: cyla.hk@gmail.com*
#Contact: cheune4@cky.edu.hk

***Abstract*— This learning summary introduces the problem of music sheet page turning. A solution to the problem is presented, where the real-time alignment of a performance with images of a music sheet is obtained using an Optical Music Recognition algorithm and Dynamic Time Warping algorithm. This learning summary also summarizes the current solutions to the problem of music sheet page turning, their problems, the aims of our solution, the methodology of our solution, limitations of our solution as well as future directions.**

I. Introduction

Musicians have to play music from a score, and when the score is multiple pages long, it must be flipped. Therefore, it poses a slight inconvenience to flip the score, especially for beginners who struggle with playing; and professionals who may be in the middle of a difficult phrase. This process impacts these musicians, and deters them from playing optimally, discouraging beginners from playing musical instruments.

II. Definition of key terms

A. Optical Music Recognition (OMR):

OMR is a field of research that investigates how to computationally read music notation in documents. One of the key goals for OMR is to produce a MIDI: a file type with a long tradition of computational processing for a vast variety of purposes, representation of music. [1]

B. Chroma features:

In music, a *chroma feature* or *chromagram* describes the twelve different pitch classes. These *chroma*-based features, which are also referred to as "*pitch class profiles*", are powerful tools used to analyse music with pitches that fit into twelve distinct categories, and whose tuning approximates to the equal-tempered scale. Chroma features possess a unique property such that they are able to capture harmonic and melodic characteristics of music, while being robust to changes in timbre and instrumentation. [2]

C. Dynamic time warping (DTW):

DTW is a well-known technique used to find the optimal alignment between two time-dependent sequences under certain restrictions. Intuitively, the sequences are warped in a nonlinear fashion to match each other. Originally, DTW has been used to compare different speech patterns in automatic speech recognition. In fields such as data mining and information retrieval, DTW has been successfully applied to automatically cope with time deformations and different speeds associated with time-dependent data. [3]

D. Hidden Markov model (HMM):

The *HMM* is a statistical model that assumes the system being modelled is a Markov process with unknown parameters. Hidden Markov models are known for their applications to thermodynamics, natural sciences, economics, finance, signal processing, information theory, etc. Most importantly, it can recognise patterns from speech, handwriting, and gestures. [4]

III. Existing solution and problems to address

Currently, the page turner pedal is the most popular method to turn a page without using the hands. When the foot pedal is stepped on, a signal triggers the page turner to turn a page. However, page turner pedals with reasonable quality cost at least 300 Hong Kong dollars, making them less appealing to beginner or casual musicians. Moreover, musicians who play instruments that utilise both feet will have difficulty using page turner pedals.

Notably, "*Enote3*", an application geared towards professional musicians, plans to implement an automatic page turning feature. However, a hefty subscription of 10 Euros a month is required. Despite its premium nature, users can't import their music score of choice into the software. Also, there are applications and GitHub repositories such as "*Playscore2*", [5] and [6], that convert digital images of music scores into a soundtrack. However, these applications do not come with a page turning functionality.

Furthermore, other applications such as "*Millie's Library"* — a South Korean e-book service provider, provides a page turning function that utilises eye-tracking technology. When the user glances at a certain part of the music score, the application turns a page. However, this method may be troublesome for musicians who like to look around during their performance, which is common for professional performers.

Therefore, our solution must incorporate the convenience brought by the foot pedal, the automation, customizability and personalised experience the existing software lacks, whilst maintaining a low economic cost. Simultaneously, it needs to preserve the easy-to-use design principle — evident in the foot pedal, and possess a simple GUI to allow for any technologically challenged musician to use our software with ease.

IV. Aims of our solution

As aforementioned, the primary aim of the application is to be able to flip the right page at the right time. To break that down, the application must recognise musical symbols on a digital music score in modern western notation and track the part of the music score that the musician is playing in real time. Logically, the application will know when the musician reaches the end of the page and must therefore turn to the next page automatically.

V. Methodology

This is the general flow of the application, and it has two pipelines: the OMR pipeline, and the DTW pipeline. These two pipelines are then brought together to generate a signal that flips the page at the right moment.

A. OMR pipeline:

The OMR pipeline generates a reference audio from the music score using SheetVision, a Github Repository [5].

In SheetVision, the input image is first binarized, then it uses a template matching algorithm from OpenCV2 to perform template scaling, character classification and classifier thresholding. Afterwards, the recognised musical symbols are sequenced and exported to a .mid file.

We chose to use SheetVision for its simplicity to set up, its reliability, and the output .mid file can be easily converted to a .wav file using simple commands.

Other ways to achieve the results were also explored. Notably [6] is troublesome to set up, and [7] - [8] are not as accurate as SheetVision. Solutions using OMR machine learning models were considered, as [9] - [11] proposed model structures that yielded accurate results. The DeepscoresV2 dataset [12] was chosen to be used as the dataset for training the models for its large number of classes of music symbols. However, it is deemed too challenging for us to understand the dataset structure and train the model, and given time constraints, this method was deemed unsuitable.

B. DTW pipeline:

The DTW pipeline first extracts the chroma feature vectors of the real time music played by the musician, and compares it with the chroma feature vectors extracted from the reference audio.

We have used a python module named *librosa* to perform the chroma feature extraction and DTW. In DTW, a cost matrix C is computed, and an optimal path P is obtained by optimising in C [13].

The assumption a conventional DTW takes, namely the forced matching between the last index from the first sequence with the last index from the other sequence [14] may not hold true, and hence requires manual intervention.

Therefore, the audio snippet consisting of the prior 10 seconds of play by the musician is compared with the reference audio using a sliding window. Simultaneously, the programme tracks if the musician is playing ahead of the reference audio, indicative by the sliding window's output. This results in the differing adjustments of the sliding window in the next call of the DTW. The more ahead the musician is playing with respect to the reference audio, the adjustment gets more subtle. Conversely, the more behind the musician is playing with respect to the reference audio, the adjustment gets more drastic. This way, the DTW pipeline can continue to track which part of the score is being played by the musician with reasonable accuracy.

Notably, the HMM approach was also considered as a candidate to track the score. This was attempted in [15] and [16]. However, it was very challenging to collect sufficient data to train a HMM for this specific use case.

VI. Results

At the end, a page turner software with a graphical user interface using Python3 Tkinter module is created. (Source code on https://github.com/CYL-Ariel/CityUGEF_AIoT_Final_Project). Users can select 2 pages of images of music scores into the software, and use the software for real-time page turning, as demonstrated in https://www.youtube.com/watch?v=qrnlEPvXHVU.

VII. Limitations

The OMR algorithm in the SheetVision repository currently only can recognise a handful of musical symbols on a simple music score, while needing a lot of time to compute. Occasionally it may miss whole lines of musical symbols.

As aforementioned in the methodology of DTW, error is expected as conventional DTW assumes that the last index from the first sequence must be matched with the last index from the other sequence. If the microphone input deviates greatly from the sliding window of the reference audio, then the behaviour of the algorithm will be unpredictable. This leads to the decrease in tolerance of wrong notes.

The above limitations are amplified when noisy input is fed into the algorithms in the reference and real-time pipelines. This may come in the form of distorted music score images, or background noise whilst playing the instrument.

VIII. Future directions

A. Continue developing a fully fledged, reliable app that can be installed on mobile devices, possibly with the following features:
   - Wrong note recognition and tempo analysis,
   - In-app recording function, and
   - In-sync multi-device page turning for ensembles and orchestras.

B. Further research in the following topics:
   - A faster, more accurate OMR algorithm or model that can recognise a larger set of musical symbols could be explored, previously considered OMR models as mentioned in the Methodology section can also be reconsidered.
   - Comparison between HMM and DTW solutions for page turning applications.

IX. Conclusions

In this summary, we have explained the background and existing solutions of the problem of automatic music sheet page turning. We have also introduced our solution to this problem and discussed the methodology, results, and limitations of our solution. Music sheet page turning is a challenging problem to tackle and in the future we will explore possible improvements or alternatives to our current solution.

Acknowledgement

We would like to express our gratitude towards Dr Ray, TAs and mentors for motivating us to learn about Artificial Intelligence, Internet of Things, entrepreneurship and giving guidance and mentorship for our future. We would also like to thank the EDB for funding this programme, as well as our schools SKHTSK and PLKCKY, for giving us this opportunity and supporting us throughout this fruitful journey.

References

[1] J. Calvo-Zaragoza, J. Hajič Jr. and A. Pacha, *Understanding Optical Music Recognition*, ACM Comput. Surv. 53, 4 (2020)
[2] A. K. Shah, M. Kattel, A. Nepal, D. Shrestha, *Chroma Feature Extraction*, Available: https://www.academia.edu/42216949/Chroma_Feature_Extraction
[3] Springer, Berlin, Heidelberg, *Dynamic Time Warping. In: Information Retrieval for Music and Motion (2007)*, Available: https://doi.org/10.1007/978-3-540-74048-3_4
[4] "Hidden Markov Model", *Wikipedia*, Wikimedia Foundation 12 February 2023, https://en.wikipedia.org/wiki/Hidden_Markov_model.
[5] C.. Pratt, *SheetVision*, (2018), GitHub repository, https://github.com/cal-pratt/SheetVision
[6] BreezeWhite, *oemer*, (2022), GitHub repository, https://github.com/BreezeWhite/oemer
[7] A. A. Hemdan, *Orchestra*, (2021), GitHub repository, https://github.com/AbdallahHemdan/Orchestra
[8] A. A. Aashrafh, *Mozart*, GitHub repository, https://github.com/aashrafh/Mozart
[9] J. Calvo-Zaragoza, J. J. Valero-Mas and A. Pertusa, *End-to-End Optical Music Recognition Using Neural Networks*, International Society for Music Information Retrieval Conference, 2017
[10] L. Tuggener, I. Elezi, J. Schmidhuber, T. Stadelmann, *Deep Watershed Detector for Music Object Recognition*, International Society for Music Information Retrieval Conference, 2018
[11] Eelco van der Wel, K. Ullrich, *Optical Music Recognition with Convolutional Sequence-to-Sequence Models*, International Society for Music Information Retrieval Conference, 2017
[12] L. Tuggener, Y. P. Satyawan, A. Pacha, J. Schmidhuber, T. Stadelmann, *DeepScoresV2*, 2020, Available: https://zenodo.org/record/4012193#.Y-m_texBzdo
[13] L. Tuggener, Y. P. Satyawan, A. Pacha, J. Schmidhuber, T. Stadelmann, *The DeepScoresV2 Dataset and Benchmark for Music Object Detection*, 2020 25th International Conference on Pattern Recognition (ICPR)
[14] "Dynamic Time Warping", *Wikipedia*, Wikimedia Foundation 12 February 2023, https://en.wikipedia.org/wiki/Dynamic_time_warping,
[15] N. Orio, F. D'échelle, *Score Following Using Spectral Analysis and Hidden Markov Models*, hal-01161055 , **version 1** (08-06-2015)

[16] B. Pardo and W. Birmingham, *Modeling Form for On-line Following of Musical Performances*, Proceedings, The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, July 9-13, 2005, Pittsburgh, Pennsylvania, USA