Automatic Music Transcription for the Thai Xylophone on Embedded Devices

CONTROL SYSTEMS AND INSTRUMENTATION ENGINEERING PROGRAM

Piradej Tantisukharom, Ratirat Kumvej, Advisor: Asst.Prof.Dr.Sarawan Wongsa

Introduction

Automatic Music Transcription (AMT)[1] aims to convert musical audio signals into musical notation using computational algorithms. Current complex AI-based music transcription struggles on embedded devices. This research proposes a new Automatic Music Transcription (AMT) technique specifically for the Thai xylophone (Ranad) that can run on these devices. It addresses the limitations of existing methods trained on Western instruments, including the requirement of extensive retraining. The developed technique could potentially enable applications in :

- Preserving Thai musical heritage: Documenting rare Thai music pieces for future generations.
- Educational Technology: Creating interactive training games for learning and practicing the Ranad.

Methods

In our Thai xylophone music transcription algorithm:

- Real-time display and MIDI file export.
- Comparison with the state-of-the-art Onsets and Frames (OaF) [2] algorithm for advanced AMT.

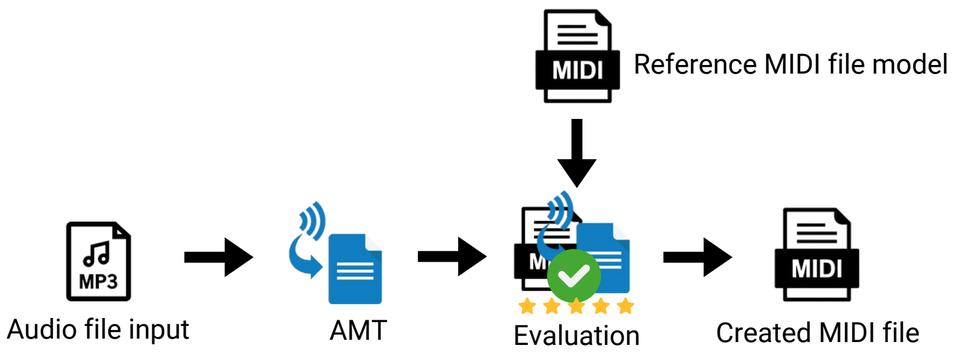

Figure 1 Workflow of the project.

- **Audio file input:** import audio files with polyphonic because the Xylophone has two sounds playing simultaneously, it will create reference files (MIDI) and use the reference files to create estimate files (.mp3) from the synthetic xylophone sounds.
- **Automatic Music Transcription (AMT):** Transcription begins by dividing the song into windows to select parts of the signal for analysis, then taking each window doing Fast Fourier Transform (FFT) and calculating the average power.

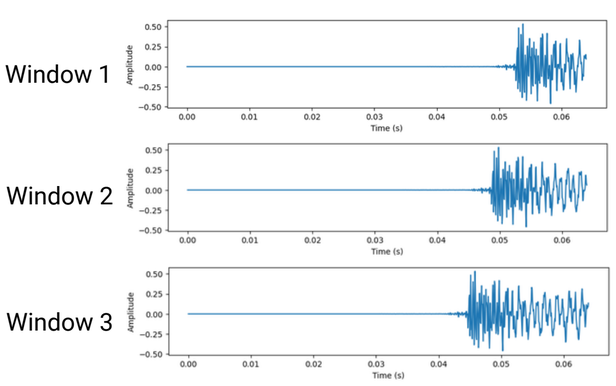

Figure 2 Windowing parts 1 to 3

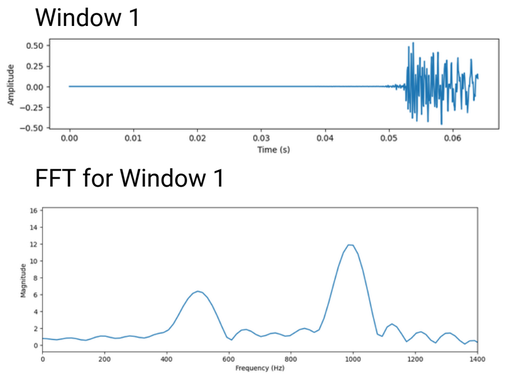

Figure 3 FFT for Window 1

- **Energy Novelty:** The average power is used to detect signals during onsets and offsets. The baseline line, created using EWMA method with lambda (λ) as the weight determinant (Figure 4), helps capture the starting and ending points of music notes. This approach ensures accurate signal detection by considering the varying energy levels (Figure 5).

$$EWMA_t = \lambda \cdot \overline{P}_t + (1 - \lambda)\, EWMA_{t-1}$$

λ = The weight decided by the user.
$\overline{P}_t$ = Value of the current average power.

Figure 4 Exponentially Weighted Moving Average (EWMA) equation.

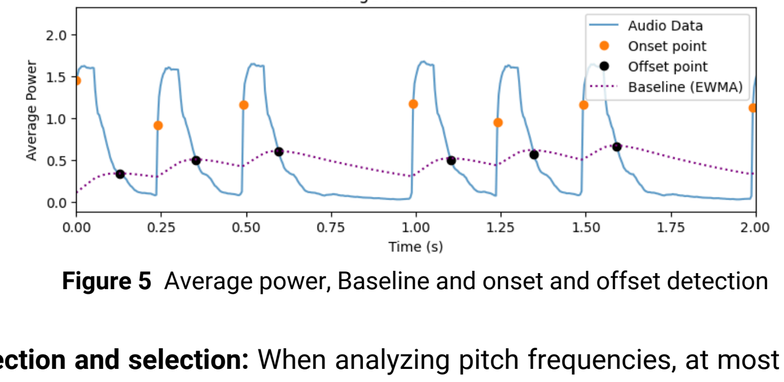

Figure 5 Average power, Baseline and onset and offset detection

- **Pitch detection and selection:** When analyzing pitch frequencies, at most two peaks are chosen as candidates. The second peak is confirmed if its magnitude is at least alpha (α) times that of the largest peak, where alpha is the weight to determine the height of the baseline (Figure 6). The final pitch frequency is determined by selecting the two most common values within a buffer, with the second value within 30% of the first.

Methods

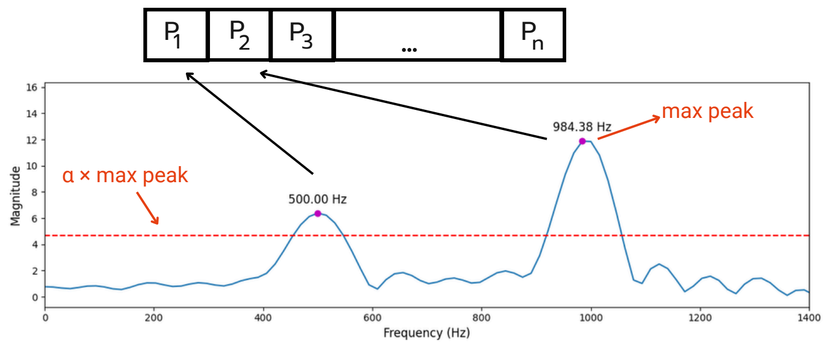

Figure 6 Detect pitch and store them in pitch buffer.

When recording is stopped, the values of onset, offset and multi-pitch are used to create a file as MIDI.

- **Compare and Evaluation :** After getting the MIDI file obtained from AMT, it will be compared with reference files (MIDI) to evaluate performance using Python's library called mir_eval.

Results

- **Songs : Jingle Bells and Groa Nai** from synthetic Thai xylophone sounds.

Table 1. The optimal values of lambda and alpha for detectly onset-offset and frequecy peaks.

	Value	Precision (%)	Recall (%)	F1-score (%)
λ	**0.009**	98.52	98.55	98.54
α	**0.6**	97.15	97.52	97.34

Table 2. Comparison between the proposed algorithm and the OaF model.

Song	Algorithm	Precision (%)	Recall (%)	F1-score (%)
Jingle Bells	Proposed	**99.63**	**98.88**	**99.25**
	OaF	53.16	68.77	59.97
Groao Nai	Proposed	**91.67**	**92.37**	**92.02**
	OaF	34.55	50.38	40.99

The proposed algorithm performs more effectively than Onsets and Frames (OaF) due to the requirements of retraining of Onsets and Frames (OaF) that were trained using piano datasets.

- Compare output MIDI notes between the proposed file and reference MIDI file.

1. Jingle Bells

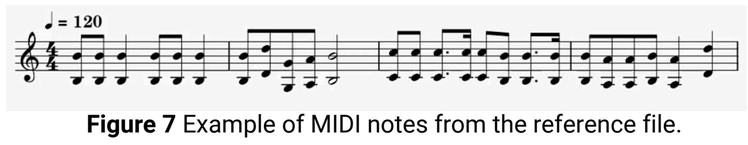

Figure 7 Example of MIDI notes from the reference file.

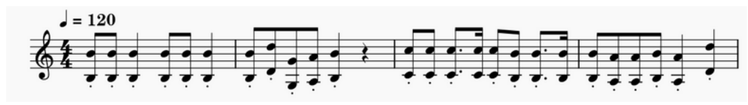

Figure 8 Example of MIDI notes from the proposed file.

2. Grao nai

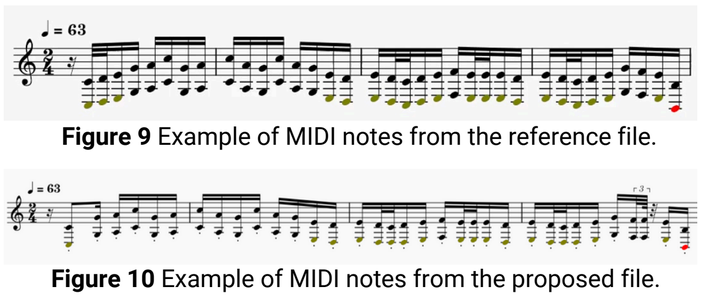

Figure 9 Example of MIDI notes from the reference file.

Figure 10 Example of MIDI notes from the proposed file.

Conclusion

Our proposed algorithm for Automatic Music Transcription for Thai xylophone detects onsets/offsets by an energy-based novelty technique and a spectral-based multi-pitch detection method. It outperforms the deep-learning-based OaF approach, which was trained on piano sounds. Because of its simplicity, the created method is appropriate for real-time deployment on embedded devices.

References

[1] Emmanouil, B., Simon, D., Zhiyao, D. and Sebastian, E., 2019, Automatic Music Transcription: An Overview, Vol.36, IEEE Signal Processing Magazine, IEEE, pp. 20-30.

[2] Curtis, H., Erich, E., Jialin, S., Adam, R., Ian, S., Colin, R., Jesse, E., Sageev, O. and Douglas, E., Onsets and Frames: Dual-Objective Piano Transcription [Online], Available: https://arxiv.org/abs/1710.11153