Automatic Music Transcription

Github repository : github.com/OliverIgnetik

Australian National University

Student: Oliver Ignetik
E-mail: u5012063@anu.edu.au

Supervisor: Dr Parastoo Sadeghi
Examiner: Dr Rodney Kennedy

Abstract

This project explores the application of Non-negative Matrix Factorization (NMF) on monophonic music samples and Neural Network (NN) models on polyphonic music database. The models performed well but need more work to incorporate context into the algorithms.

Background

The capability of transcribing music audio into music notation is a fascinating example of human intelligence. It involves analyzing complex auditory scenes, recognizing musical objects, forming musical structures and checking alternative hypotheses. Automatic Music Transcription (AMT) refers to the design of computational algorithms to convert acoustic music signals into some form of music notation. It is a challenging task and considered an unsolved problem in signal processing and artificial intelligence. This problem is particularly challenging in polyphonic music where even the most advanced systems are far behind meeting the accuracy of trained musicians [1].

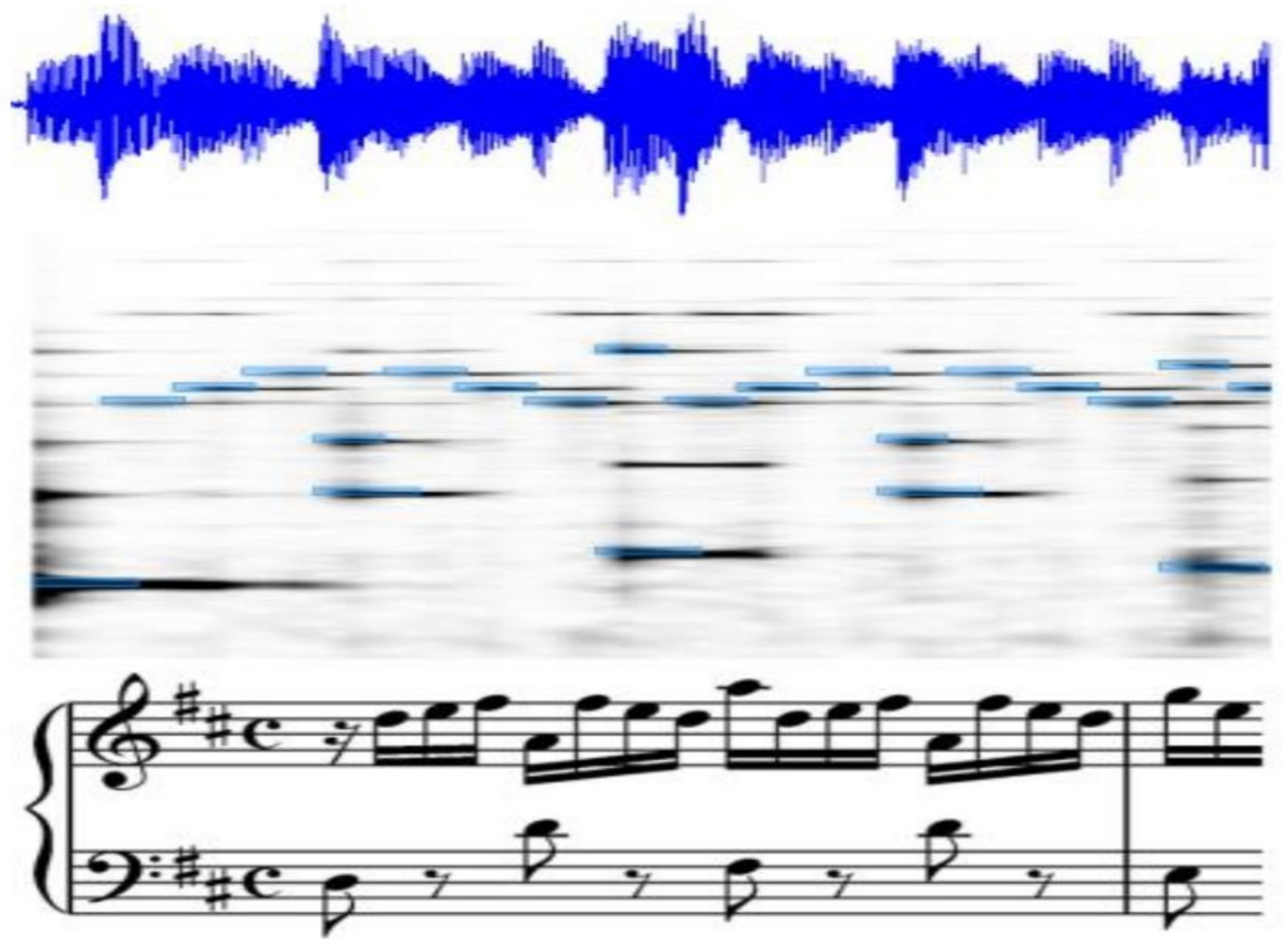


Fig 1: AMT process piano roll representation [2]

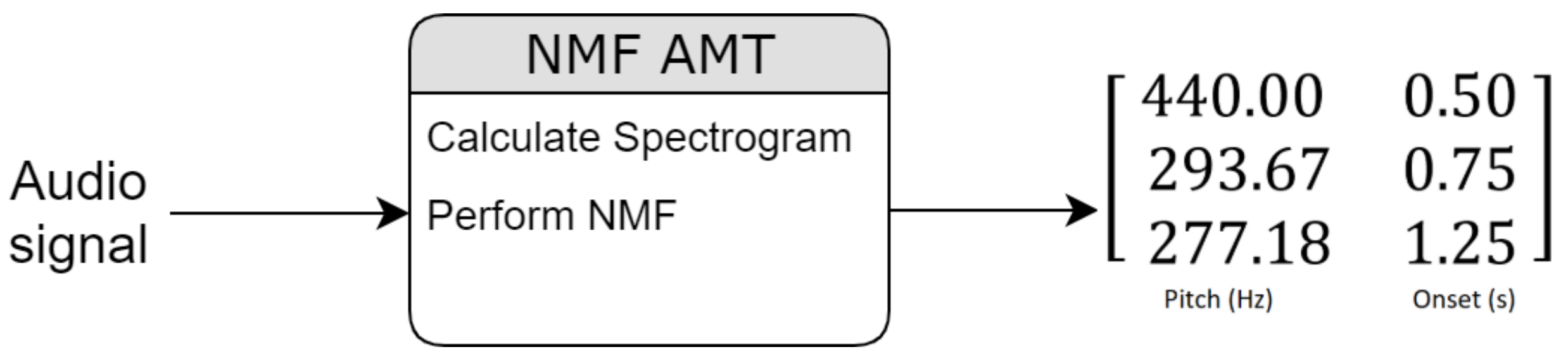
Key challenges

- Polyphonic mixtures – disentangling harmonics of consonant notes and diatonic harmonies
- Lack of ground truth transcriptions – annotation is extremely time consuming and there are a limited number of datasets

Methodology

NMF AMT model – monophonic

- Data – MAPS database piano recordings with ground truths
- Approach – NMF decomposition with peak picking algorithms



Audio signal

NMF AMT

Calculate Spectrogram

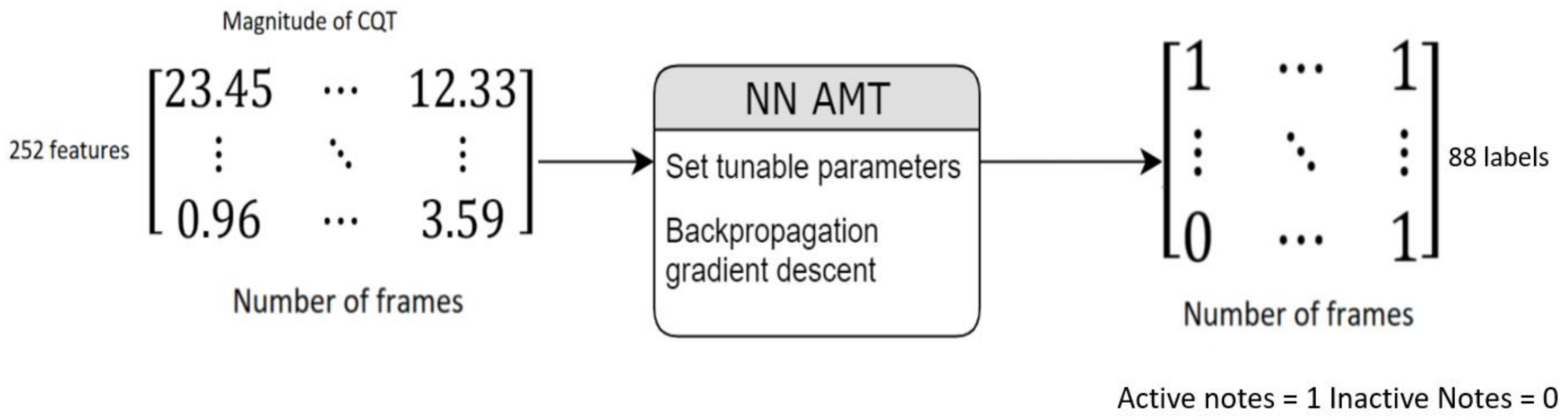
Perform NMF

$$\begin{array}{cc} \begin{bmatrix} 440.00 & 0.50 \\ 293.67 & 0.75 \\ 277.18 & 1.25 \end{bmatrix} \\ \text{Pitch (Hz)} \quad \text{Onset (s)} \end{array}$$

Fig 2: NMF model with input and output format

NN AMT model – polyphonic

- Data – Musicnet HDF5 file with audio and ground truths
- Approach – Multilabel classification with Constant Q-Transform input



Magnitude of CQT

$$\begin{array}{cc} \text{252 features} & \begin{bmatrix} 23.45 & \cdots & 12.33 \\ \vdots & \ddots & \vdots \\ 0.96 & \cdots & 3.59 \end{bmatrix} \end{array}$$

Number of frames

NN AMT

Set tunable parameters

Backpropagation gradient descent

$$\begin{array}{cc} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} & \text{88 labels} \end{array}$$

Number of frames

Active notes = 1 Inactive Notes = 0

Fig 3: NN model with input and output format

Results/Findings

NMF AMT model – monophonic

- Overall accuracy achieved of 1.0 (100%)
- Peak picking algorithm threshold value and the number of harmonics were the most influential parameters

NN AMT model – polyphonic

- F-measure achieved of 0.75 (75%)
- Initial approach used a binary entropy loss function with an automatically inferred accuracy which did not take into account the class imbalance

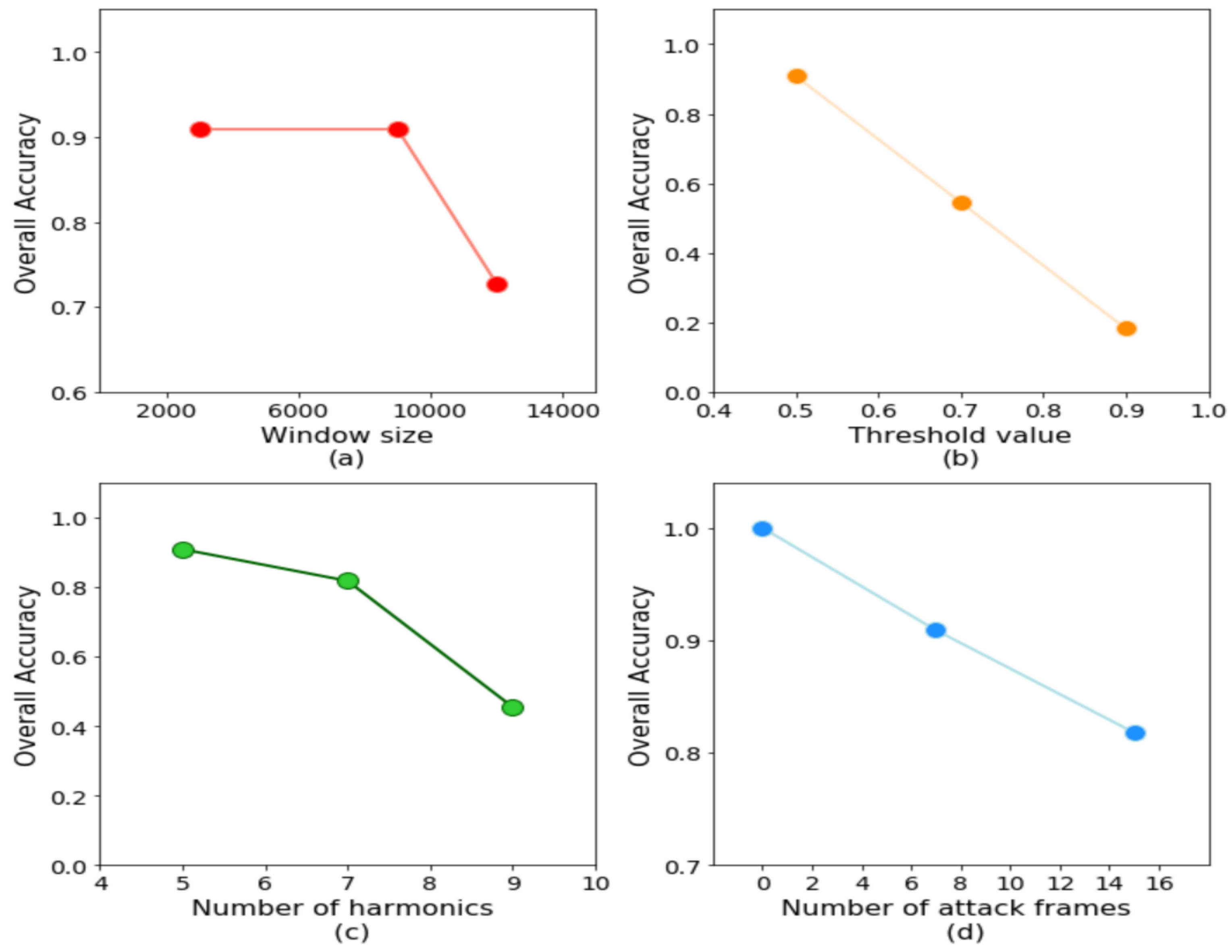


Fig 4 : NMF model performance

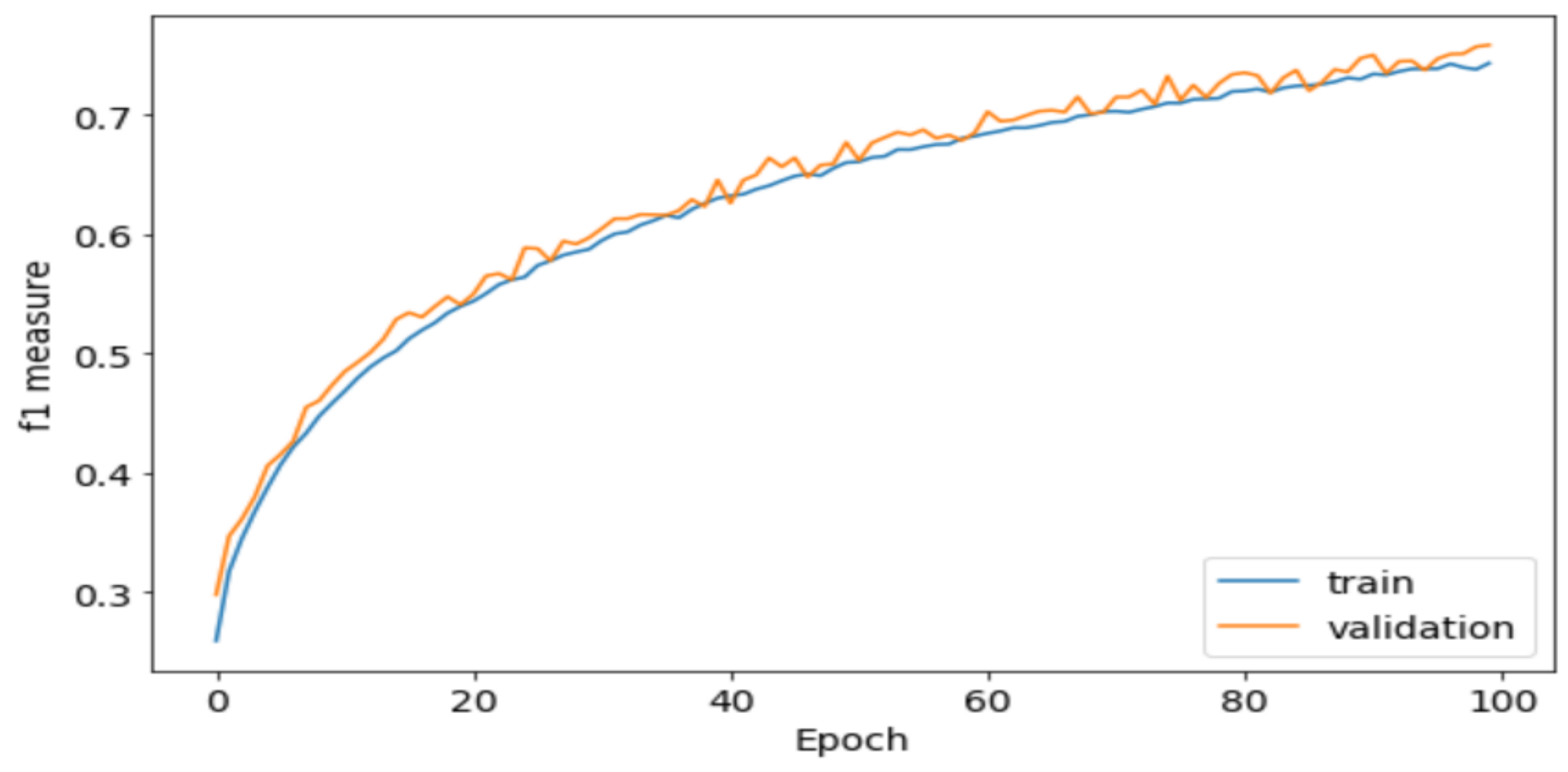


Fig 5 : NN model performance

Discussion

- NMF model had an accuracy of 1.0 in a simple monophonic transcription task but needs extension to polyphonic cases
- NN model used a weighted loss function and appropriate scoring metric which led to a stable model performance curve
- The NN model outperformed a similar model in a paper by Li et al [3] in which a regular binary entropy loss function is used

Conclusions

- NMF model has fast implementation but needs appropriate parameter tuning dependent on contextual factors
- NN model takes a lot of time to train but performed well in polyphonic case. In the future, temporal dependencies may be included through the use of different architectures
- Future work may include incorporation of prior musical knowledge such as instrument, recording environment and lead sheet information

References

[1] M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis”, IEEE J. Sel. Topics Signal Process, vol. 5, no. 6, pp. 1088–110, 2011.

[2] E. Benetos, “Automatic music transcription”, Tutorial presented at National University of Singapore, University of London, January 2019

[3] L. Li, I. Ni, and L. Yang, “Music transcription using deep learning”, 2019.

ENGN4200 Individual Project Thesis Presentation
25 May 2020

ANU CECS