Beyond the Beat *Towards Metre, Rhythm and Melody Modelling with Hybrid Oscillator Networks*

Andrew Lambert[1,2], Tillman Weyde[1], Newton Armstrong[2]

[1]Department of Computer Science, [2]Department of Music

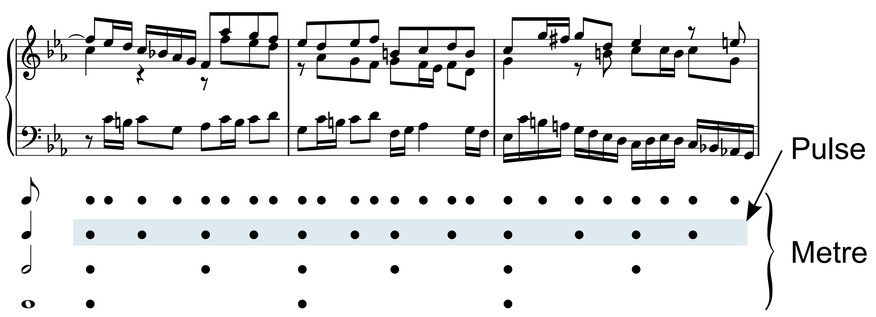
CITY UNIVERSITY LONDON
EST 1894

{andrew.lambert.1 | t.e.weyde | newton.armstrong.1}@city.ac.uk

1 - Introduction

Beat induction allows us to tap along to the beat of music, perceiving its **pulse**. Finding the pulse within a musical signal is a step towards achieving other music perception tasks, such as **metre** perception.

Metre
*The multi-layered divisions of time present in music, of which the referent layer is the **pulse**. Other layers in music divide the pulse into the smallest subdivisions of time, and extend it towards larger measures, phrases, periods, and even higher order forms.*



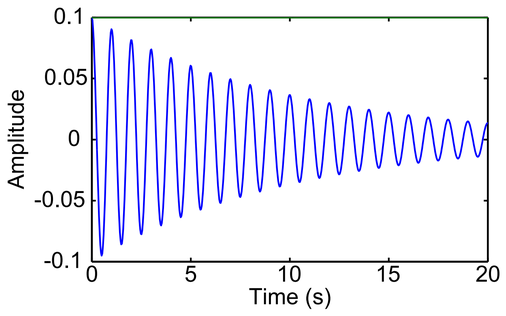
A single **beat** can occur at one or more metrical levels. The more levels on which beat occurs, the 'stronger' that beat is perceived, creating a beat hierarchy, or **metrical structure**[1].

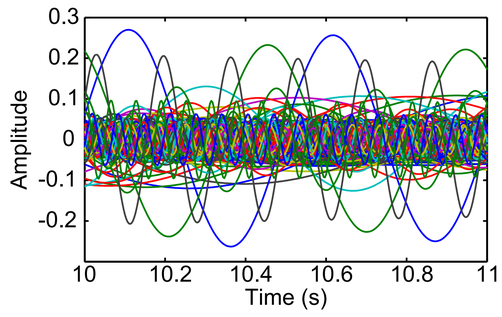
Taking a connectionist **machine learning** approach, this project's aim is to design a hybrid network which is able to learn metrical structures, generalising on a corpus of sequences to make predictions about future musical events.

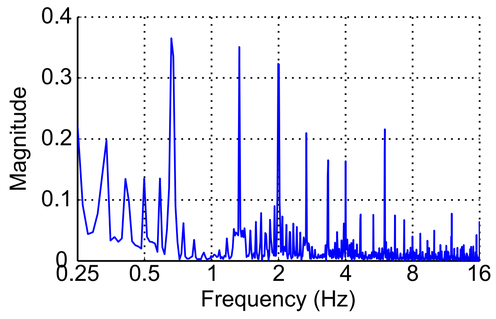
This is an investigation into machine models of melody and rhythm. We are investigating if a **music prediction** task produces better results when utilising a certain model of metrical structure.

2 - Models

Gradient Frequency Neural Network (GFNN) [2]







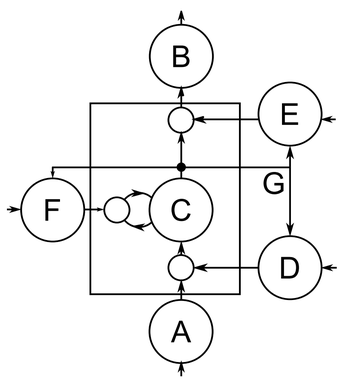
A network of **nonlinear oscillators**, distributed across a frequency spectrum. The oscillators entrain and resonate nonlinearly to periodicities in stimulus.

Resonances occur at integer ratios to the pulse and can be interpreted as a hierarchical **metrical structure**.

Long Short-Term Memory Network (LSTM) [3]

A Recurrent Neural Network (RNN) that overcomes the lack **global coherence** often found in other RNNs due to the lack of long-term memory.

A self-connected node known as the Constant Error Carousel (CEC) ensures constant error flow back through time.



A) Input
B) Output
C) CEC
D) Input gate
E) Output gate
F) Forget gate
G) Peepholes

The input and output gates control how information flows into and out of the CEC, and the forget gate controls when the CEC is reset. These gates are connected via ‘peepholes’.

3 - Experiments

We constructed **5 networks** and trained them on a **melody prediction** task: to predict the next sample in time-series music data.

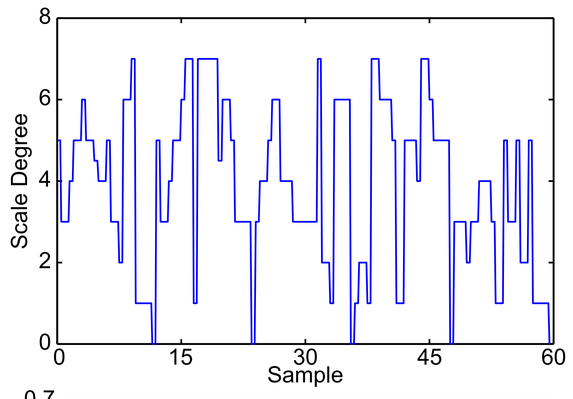
We used monophonic **symbolic** music data consisting of 100 German folk songs from the Essen Folksong Collection[4]. Pitch data was abstracted to relative **scale degrees**, accidentals were encoded by adding or subtracting 0.5 from the scale degree and rests were encoded as 0 values.

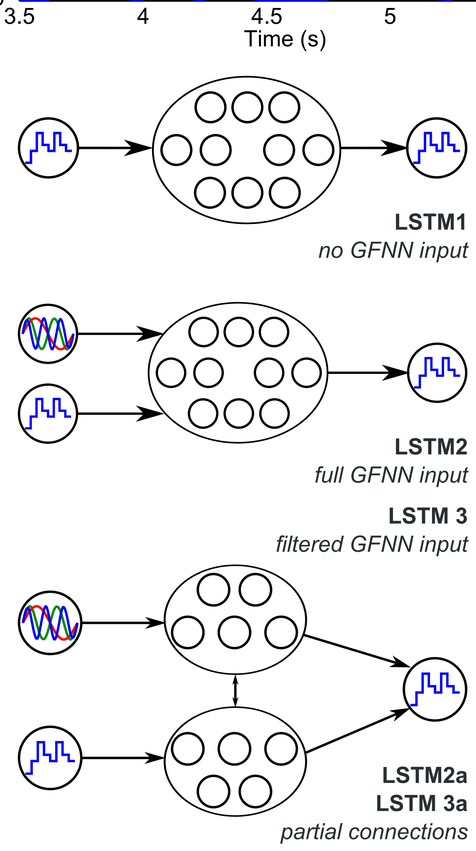
The GFNN consisted of **128 oscillators** stimulated by an **onset pattern**. Frequencies were logarithmically distributed from 0.25Hz to 16Hz.

All sequences in the corpus were synthesised at a tempo of 120bpm (2Hz), meaning that our metrical periodicities the GFNN ranged from a demisemiquaver (32nd note) to a breve (double whole note).

We **pre-filtered** the GFNN output on LSTM3 and LSTM3a by averaging the GFNN output over the corpus and finding the largest amplitude responses over the final 25% of the piece. Once these frequencies were found, they were fixed for all sequences. This reduced the dimensionality from 128 to 8.

The number of hidden LSTM blocks was fixed at 10 for all 5 networks. Training was done by backpropagation through time using RProp$^-$ and 4-fold cross validation.





4 - Results

Network	Sequence	Precision	Recall	F-measure
LSTM1	0.39842	0.91955	0.45575	0.60645
LSTM2	0.38229	0.93898	0.45729	0.61274
LSTM3	**0.49428**	0.9289	0.45214	0.60555
LSTM2a	0.38644	**0.95247**	**0.45953**	**0.61735**
LSTM3a	0.44366	0.92402	0.45849	0.60988

Mean results on the training folds

Network	Sequence	Precision	Recall	F-measure
LSTM1	0.39071	0.91962	0.45623	0.60599
LSTM2	0.32831	0.93313	0.45739	0.61157
LSTM3	**0.49273**	0.92689	0.45298	0.60582
LSTM2a	0.35777	**0.94818**	**0.46885**	**0.62507**
LSTM3a	0.4401	0.92421	0.46349	0.6142

Mean results on the validation folds

Sequence
a proportion of samples where the network output, rounded to the nearest half, matches the target value

Precision
a ratio of correctly predicted onsets to all predicted onsets

Recall
a ratio of correctly predicted onsets to ground truth onsets

F-measure
a harmonic mean of precision and recall

5 - Conclusions

The GFNN output, with its strong and weak nonlinear resonances at frequencies related to the pulse, can be interpreted as a **perception of metre**. Our results show that providing this data helped to improve melody prediction with an LSTM.

Sequence prediction was fairly poor for all networks, but LSTM3 achieved around 10% higher prediction accuracy than LSTM1. LSTM3 consistently outperformed LSTM1 in sequence modelling across training and test datasets and is a statistically significant result in both cases.

This provides some evidence that melody modelling benefits from the nonlinear resonance model. We hypothesise that this is due to the LSTM being able to make use of the relatively **long temporal resonance** in the GFNN output, and therefore model more coherent long-term structures.

By inputting metrical data, our system can be extended to work with real time data, as opposed to the metrically quantised data we are using here. This opens up the system for use with a multitude of different tempos and live performance applications.

6 - References

[1] F. Lerdahl and R. Jackendoff, “An overview of hierarchical structure in music,” Music Perception: An Interdisciplinary Journal, vol. 1, no. 2, pp. 229–252, Dec. 1983.
[2] E. W. Large, F. V. Almonte, and M. J. Velasco, “A canonical model for gradient frequency neural networks,” Physica D: Nonlinear Phenomena, vol. 239, no. 12, pp. 905–911, Jun. 2010.
[3] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with LSTM,” Neural Computation, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.
[4] H. Schaffrath. (1995) The essen folksong collection in kern format. [Online]. Available: http://www.esac-data.org/