



City Research Online

City St George's, University of London

Citation: Lambert, A. J., Weyde, T. & Armstrong, N. (2016). Adaptive Frequency Neural Networks for Dynamic Pulse and Metre Perception. In: Mandel, M. I., Devaney, J., Turnbull, D. & Tzanetakis, G. (Eds.), Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016, New York City, United States, August 7-11, 2016. (pp. 60-66). Dagstuhl: Schloss Dagstuhl LZI,. ISBN 978-0-692-75506-8

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/21655/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

ADAPTIVE FREQUENCY NEURAL NETWORKS FOR DYNAMIC PULSE AND METRE PERCEPTION

Andrew J. Lambert, Tillman Weyde, and Newton Armstrong

City University London

{andrew.lambert.1; t.e.veyde; newton.armstrong.1}@city.ac.uk

ABSTRACT

Beat induction, the means by which humans listen to music and perceive a steady pulse, is achieved via a perceptual and cognitive process. Computationally modelling this phenomenon is an open problem, especially when processing expressive shaping of the music such as tempo change. To meet this challenge we propose Adaptive Frequency Neural Networks (AFNNs), an extension of Gradient Frequency Neural Networks (GFNNs).

GFNNs are based on neurodynamic models and have been applied successfully to a range of difficult music perception problems including those with syncopated and polyrhythmic stimuli. AFNNs extend GFNNs by applying a Hebbian learning rule to the oscillator frequencies. Thus the frequencies in an AFNN adapt to the stimulus through an attraction to local areas of resonance, and allow for a great dimensionality reduction in the network.

Where previous work with GFNNs has focused on frequency and amplitude responses, we also consider phase information as critical for pulse perception. Evaluating the time-based output, we find significantly improved responses of AFNNs compared to GFNNs to stimuli with both steady and varying pulse frequencies. This leads us to believe that AFNNs could replace the linear filtering methods commonly used in beat tracking and tempo estimation systems, and lead to more accurate methods.

1. INTRODUCTION

Automatically processing an audio signal to determine pulse event onset times (beat tracking) is a mature field, but it is by no means a solved problem. Analysis of beat tracking failures has shown that beat trackers have great problems with varying tempo and expressive timing [5, 6].

The neuro-cognitive model of nonlinear resonance models the way the nervous system resonates to auditory rhythms by representing a population of neurons as a canonical nonlinear oscillator [15]. A Gradient Frequency Neural Network (GFNN) is an oscillating neural network

model based on nonlinear resonance. The network consists of a number of canonical oscillators distributed across a frequency range. The term ‘gradient’ is used to refer to this frequency distribution, and should not be confused with derivative-based learning methods in Machine Learning. GFNNs have been shown to predict beat induction behaviour from humans [16, 17]. The resonant response of the network adds rhythm-harmonic frequency information to the signal, and the GFNN’s entrainment properties allow each oscillator to phase shift, resulting in deviations from their natural frequencies. This makes GFNNs good candidates for modelling the perception of temporal dynamics in music.

Previous work on utilising GFNNs in an MIR context has shown promising results for computationally difficult rhythms such as syncopated rhythms where the pulse frequency may be completely absent from the signal’s spectrum [17, 23], and polyrhythms where there is more than one pulse candidate [2]. However, these studies have placed a focus on the frequencies contained in the GFNN’s output, often reporting the results in the form of a magnitude spectrum, and thus omitting phase information. We believe that when dealing with pulse and metre perception, phase is an integral part as it constitutes the difference between entraining to on-beats, off-beats, or something in-between. In the literature, the evaluation of GFNNs’ pulse finding predictions in terms of phase has, to our knowledge, never been attempted.

Our previous work has used GFNNs as part of a machine learning signal processing chain to perform rhythm and melody prediction. An expressive rhythm prediction experiment showed comparable accuracy to the state-of-the-art beat trackers. However, we also found that GFNNs can sometimes become noisy, especially when the pulse frequency fluctuates [12, 13].

This paper presents a novel variation on the GFNN, which we have named an Adaptive Frequency Neural Network (AFNN). In an AFNN, an additional Hebbian learning rule is applied to the oscillator frequencies in the network. Hebbian learning is a correlation-based learning observed in neural networks [9]. The frequencies adapt to the stimulus through an attraction to local areas of resonance. A secondary elasticity rule attracts the oscillator frequencies back to their original values. These two new interacting adaptive rules allow for a great reduction in the density of the network, minimising interference whilst also maintaining a frequency spread across the gradient.



© Andrew J. Lambert, Tillman Weyde, and Newton Armstrong. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Andrew J. Lambert, Tillman Weyde, and Newton Armstrong. “Adaptive Frequency Neural Networks for Dynamic Pulse and Metre Perception”, 17th International Society for Music Information Retrieval Conference, 2016.

The results of an experiment with GFNNs and AFNNs are also presented, partially reproducing the results from Velasco and Large’s last major MIR application of a GFNN [23], and Large et al.’s more recent neuroscientific contribution [17]. However, we place greater evaluation focus on phase accuracy. We have found that AFNNs can produce a better response to stimuli with both steady and varying pulses.

The structure of this paper is as follows: Section 2 provides a brief overview of the beat-tracking literature and the GFNN model, Section 3 introduces a phase based evaluation method, Section 4 introduces our new AFNN model, Section 5 details the experiments we have conducted and shares the results, and finally Section 6 provides some conclusions and points to future work.

2. BACKGROUND

2.1 Pulse and Metre

Lerdahl and Jackendoff’s *Generative Theory of Tonal Music* [19] was one of the first formal theories to put forward the notion of hierarchical structures in music which are not present in the music itself, but perceived and constructed by the listener. One such hierarchy is *metrical structure*, which are layers of beats existing in a hierarchically layered relationship with the rhythm. Each metrical level is associated with its own period, which divides the previous level’s period into a certain number of parts.

Humans often choose a common, comfortable metrical level to tap along to, which is known as a *preference rule* in the theory. This common metrical level is commonly referred to as ‘the beat’, but this is a problematic term since a beat can also refer to a singular rhythmic event or a metrically inferred event. To avoid that ambiguity, we use the term ‘pulse’ [4].

2.2 Beat Tracking

Discovering the pulse within audio or symbolic data is known as *beat tracking* and has a long history of research dating back to 1990 [1]. There have been many varied approaches to beat tracking over the years, and here we focus on systems relevant to the proposed model. Some early work by Large used a single nonlinear oscillator to track beats in performed piano music [14]. Scheirer used linear comb filters [22], which operate on similar principles to Large and Kolen’s early work on nonlinear resonance [18]. A comb filter’s state is able to represent the rhythmic content directly, and can track tempo changes by only considering one metrical level. Klapuri et al.’s system builds on Scheirer’s design by also using comb filters, and extends the model to three metrical levels [10]. More recently, Böck et al. [3] used resonating feed backward comb filters with a particular type of Recurrent Neural Network called a Long Short-Term Memory Network (LSTM) to achieve a state-of-the-art beat tracking result in the MIR Evaluation eXchange (MIREX)¹.

¹ <http://www.music-ir.org/mirex/>

2.3 Nonlinear Resonance

Jones [7] proposed a psychological entrainment theory to address how humans are able to attend temporal events. Jones posited that rhythmic patterns such as music and speech potentially entrain a hierarchy of oscillations, forming an *attentional rhythm*. Thus, entrainment assumes an organisational role for temporal patterns and offers a prediction for future events, by extending the entrained period into the future.

Large then extended this theory with the notion of *non-linear resonance* [15]. Musical structures occur at similar time scales to fundamental modes of brain dynamics, and cause the nervous system to resonate to the rhythmic patterns. Certain aspects of this resonance process can be described with the well-developed theories of neurodynamics, such as oscillation patterns in neural populations. Through the use of neurodynamics, Large moves between physiological and psychological levels of modelling, and directly links neural activity with music. Several musical phenomena can all arise as patterns of nervous system activation, including perceptions of pitch and timbre, feelings of stability and dissonance, and pulse and metre perception.

The model’s basis is the canonical model of Hopf normal form oscillators, which was derived as a model oscillating neural populations [16]. Eqn (1) shows the differential equation that defines the canonical model, which is a Hopf normal form oscillator with its higher order terms fully expanded:

$$\begin{aligned} \frac{dz}{dt} = & z(\alpha + i\omega + (\beta_1 + i\delta_1)|z|^2 + \frac{(\beta_2 + i\delta_2)\varepsilon|z|^4}{1 - \varepsilon|z|^2}) \\ & + kP(\varepsilon, x(t))A(\varepsilon, \bar{z}) + \sum_{i \neq j} c_{ij} \frac{z_j}{1 - \sqrt{\varepsilon}z_j} \cdot \frac{1}{1 - \sqrt{\varepsilon}\bar{z}_i}, \end{aligned} \quad (1)$$

z is a complex valued output where the real and imaginary parts represent excitation and inhibition, \bar{z} is its complex conjugate, and ω is the driving frequency in radians per second. α is a linear damping parameter, and β_1, β_2 are amplitude compressing parameters, which increase stability in the model. δ_1, δ_2 are frequency detuning parameters, and ε controls the amount of nonlinearity in the system. $x(t)$ is a time-varying external stimulus, which is also coupled nonlinearly and consists of passive part, $P(\varepsilon, x(t))$, and an active part, $A(\varepsilon, \bar{z})$, controlled by a coupling parameter k . c_{ji} is a complex number representing phase and magnitude of a connection between the i^{th} and j^{th} oscillator (z_i, z_j). These connections can be strengthened through unsupervised Hebbian learning, or set to fixed values as in [23]. In our experiments presented here, we set c_{ij} to 0.

By varying the oscillator parameters, a wide range of behaviours not encountered in linear models can be induced (see [15]). In general, while the canonical model maintains an oscillation according to its parameters, it entrains to and resonates with an external stimulus nonlinearly. The α parameter acts as a bifurcation parameter:

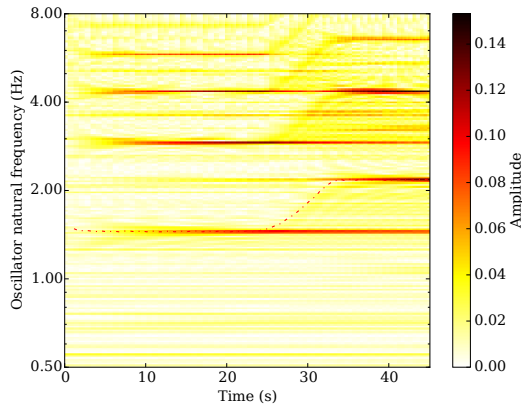


Figure 1. Amplitudes of oscillators over time. The dashed line shows stimulus frequency. The stimulus itself is shown in Figure 2. There is an accelerando after approximately 25s.

when $\alpha < 0$ the model behaves as a damped oscillator, and when $\alpha > 0$ the model oscillates spontaneously, obeying a limit-cycle. In this mode, the oscillator is able to maintain a long temporal memory of previous stimulation.

Canonical oscillators will resonate to an external stimulus that contains frequencies at integer ratio relationships to its natural frequency. This is known as mode-locking, an abstraction on phase-locking in which k cycles of oscillation are locked to m cycles of the stimulus. Phase-locking occurs when $k = m = 1$, but in mode-locking several harmonic ratios are common such as 2:1, 1:2, 3:1, 1:3, 3:2, and 2:3 and even higher order integer ratios are possible [17], which all add harmonic frequency information to a signal. This sets nonlinear resonance apart from many linear filtering methods such as the resonating comb filters used in [10] and Kalman filters [8].

2.4 Gradient Frequency Neural Networks

Connecting several canonical oscillators together with a connection matrix forms a *Gradient Frequency Neural Network* (GFNN) [16]. When the frequencies in a GFNN are distributed within a rhythmic range and stimulated with music, resonances can occur at integer ratios to the pulse.

Figure 1 shows the amplitude response of a GFNN to a rhythmic stimulus over time. Darker areas represent stronger resonances, indicating that that frequency is relevant to the rhythm. A hierarchical structure can be seen to emerge from around 8 seconds, in relation to the pulse which is just below 2Hz in this example. At around 24 seconds, a tempo change occurs, which can be seen by the changing resonances in the figure. These resonances can be interpreted as a perception of the hierarchical metrical structure.

Velasco and Large [23] connected two GFNNs together in a pulse detection experiment for syncopated rhythms. The two networks were modelling the sensory and motor cortices of the brain respectively. In the first network, the oscillators were set to a bifurcation point between damped

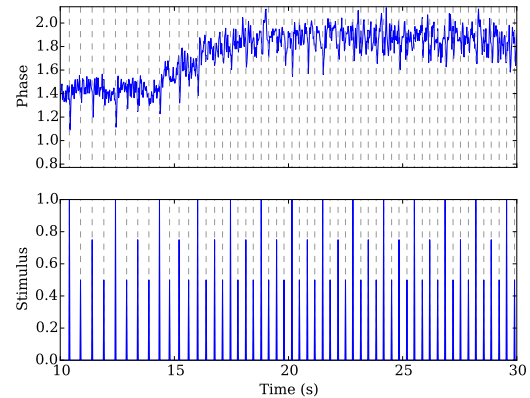


Figure 2. Weighted phase output, Φ , of the GFNN over time. The stimulus is the same as Figure 1.

and spontaneous oscillation ($\alpha = 0, \beta_1 = -1, \beta_2 = -0.25, \delta_1 = \delta_2 = 0$ and $\varepsilon = 1$). The second network was tuned to exhibit double limit cycle bifurcation behaviour ($\alpha = 0.3, \beta_1 = 1, \beta_2 = -1, \delta_1 = \delta_2 = 0$ and $\varepsilon = 1$), allowing for greater memory and threshold properties. The first network was stimulated by a rhythmic stimulus, and the second was driven by the first. Internal connections were set to integer ratio relationships such as 1:3 and 1:2, these connections were fixed and assumed to have been learned through a Hebbian process. The results showed that the predictions of the model confirm observations in human performance, implying that the brain may be adding frequency information to a signal to infer pulse and metre [17].

3. PHASE BASED EVALUATION

Thus far in the literature, evaluation of GFNNs has not considered phase information. The phase of oscillations is an important output of a GFNN; in relation to pulse it constitutes the difference between predicting at the correct pulse times, or in the worst-case predicting the off-beats. This is concerning in Velasco and Large's evaluation of pulse detection in syncopated beats, which by definition contain many off-beat events [23].

Phase and frequency are interlinked in that frequency can be expressed as a rate of phase change and indeed the canonical oscillators' entrainment properties are brought about by phase shifts. Since the state of a canonical oscillator is represented by a complex number, both amplitude and phase can be calculated instantaneously by taking the magnitude ($r = |z|$), and angle ($\varphi = \arg(z)$) respectively. We propose calculating the weighted phase output, Φ , of the GFNN as a whole, shown in (2).

$$\Phi = \sum_{i=0}^N r_i \varphi_i \quad (2)$$

Figure 2 shows the weighted phase output, Φ , over time. Even though the amplitude response to the same stimu-

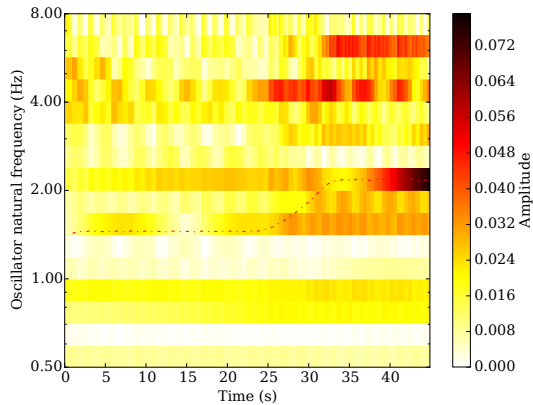


Figure 3. A low density (4opo) GFNN output. The dashed line shows stimulus frequency. Frequency information is not being captured as successfully, as can be observed by the low resonances.

lus shows a clear corresponding metrical hierarchy (see Figure 1), the phase response remains noisy. This is due to the high density of oscillators required in a GFNN. Velasco and Large used 289 oscillators per layer in their experiment, a density of 48 oscillators per octave (opo). These high densities are often used in GFNNs to capture a wide range of frequencies, but can cause interference in the network. The term ‘interference’ is used here to mean interacting signals amplifying or cancelling each other when summed. Since each oscillator can only entrain to a narrow range of frequencies, using a lower density not only increases the likelihood missing a relevant frequency, it also stops local frequency populations from reinforcing one another. An example of this can be seen in Figure 3, where frequency information is not being captured as successfully in a 4opo GFNN.

In our previous work, we have addressed this issue by using only the real part of the oscillator as a single mean-field output [11, 12]. This retained a meaningful representation of the oscillation, but ultimately removed important information. A selective filter could also be applied, by comparing each oscillator with the mean amplitude of the GFNN, and only retaining resonating oscillators. However, this is not an ideal solution to the interference problem as it requires an additional, non real-time processing step which cannot be easily incorporated into an online machine learning chain. Furthermore, new frequencies would not be selected until they begin to resonate above the selection threshold, meaning that new resonances in changing tempos may be missed.

4. ADAPTIVE FREQUENCY NEURAL NETWORK

The Adaptive Frequency Neural Network (AFNN) attempts to address both the interference within high density GFNNs, and improve the GFNNs ability to track changing frequencies, by introducing a Hebbian learning rule on the frequencies in the network. This rule is an adapted form of

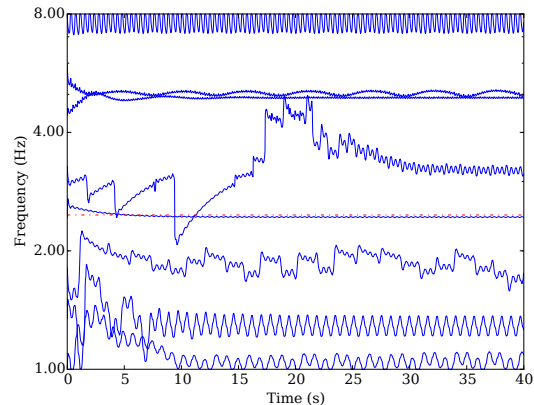


Figure 4. AFNN frequencies adapting to a sinusoidal stimulus. The dashed line shows stimulus frequency.

the general model introduced by Righetti et al. [21] shown in (3):

$$\frac{d\omega}{dt} = -\frac{\epsilon}{r}x(t)\sin(\varphi) \quad (3)$$

Their method depends on an external driving stimulus ($x(t)$) and the state of the oscillator (r, φ), driving the frequency (ω) toward the frequency of the stimulus. The frequency adaptation happens on a slower time scale than the rest of the system, and is influenced by the choice of ϵ , which can be thought of as a force scaling parameter. ϵ also scales with r , meaning that higher amplitudes are affected less by the rule.

This method differs from other adaptive models such as McAuley’s phase-resetting model [20] by maintaining a biological plausibility ascribed to Hebbian learning [9]. It is also a general method that has been proven to be valid for limit cycles of any form and in any dimension, including the Hopf oscillators which form the basis of GFNNs (see [21]).

We have adapted this rule to also include a linear elasticity, shown in (4).

$$\frac{d\omega}{dt} = -\frac{\epsilon_f}{r}x(t)\sin(\varphi) - \frac{\epsilon_h}{r}\left(\frac{\omega - \omega_0}{\omega_0}\right) \quad (4)$$

The elastic force is an implementation of Hooke’s Law, which describes a force that strengthens with displacement. We have introduced this rule to ensure the AFNN retains a spread of frequencies (and thus metrical structure) across the gradient. The force is relative to natural frequency, and can be scaled through the ϵ_h parameter. By balancing the adaptive (ϵ_f) and elastic (ϵ_h) parameters, the oscillator frequency is able to entrain to a greater range of frequencies, whilst also returning to its natural frequency (ω_0) when the stimulus is removed. Figure 4 shows the frequencies adapting over time in the AFNN under sinusoidal input.

The AFNN preserves the architecture of the GFNN; the main difference is the frequency learning procedure. Figure 5 shows the weighted phase output (Φ) of an AFNN

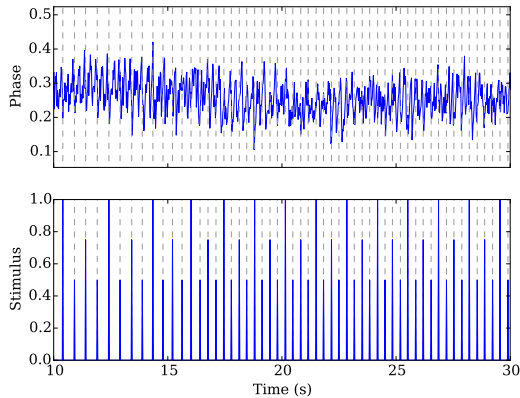


Figure 5. Weighted phase output, Φ , of the AFNN over time. Reduced interference can be seen compared with Figure 2.

stimulated with the same stimulus as in Figure 2. One can observe that a reduced level of interference is apparent.

5. EXPERIMENT

We have conducted a pulse detection experiment designed to test two aspects of the AFNN.

Firstly, we wanted to discover how the output of the AFNN compares with the GFNN presented in [23]. To this end, we are using similar oscillator parameters ($\alpha = 0, \beta_1 = \beta_2 = -1, \delta_1 = \delta_2 = 0$ and $\varepsilon = 1$). This is known as the ‘critical’ parameter regime, poised between damped and spontaneous oscillation. We are retaining their GFNN density of 48opo, but reducing the number of octaves to 4 (0.5-8Hz, logarithmically distributed), rather than the 6 octaves (0.25-16Hz) used in [23]. This equates to 193 oscillators in total. This reduction did not affect our results and is more in line with Large’s later GFNN ranges (see [17]).

The AFNN uses the same oscillator parameters and distribution, but the density is reduced to 4opo, 16 oscillators in total. ϵ_f and ϵ_h were hand-tuned to the values of 1.0 and 0.3 respectively. For comparison with the AFNN, a low density GFNN is also included, with the same density as the AFNN but no adaptive frequencies.

We have selected two of the same rhythms used by Velasco and Large for use in this experiment, the first is an isochronous pulse and the second is the more difficult ‘son clave’ rhythm. We supplemented these with rhythms from the more recent Large et al. paper [17]. These rhythms are in varying levels of complexity (1-4), varied by manipulating the number of events falling on on-beats and off-beats. A level 1 rhythm contains one off-beat event, level 2 contains two off-beat events and so forth. For further information about these rhythms, see [17]. Two level 1 patterns, two level 2 patterns, two level 3 patterns, and four level 4 patterns were used.

The second purpose of the experiment was to test the AFNN and GFNN’s performance on dynamic pulses,

therefore we have included two additional stimulus rhythms: an accelerando and a ritardando.

We are additionally testing these rhythms at 20 different tempos selected randomly from a range 80-160bpm. None of the networks tested had any internal connections activated, fixed or otherwise ($c_{ij} = 0$). An experiment to study of the effect of connections is left for future work.

In summary, the experiment consisted of 5 stimulus categories, 20 tempos per category and 3 networks. There are two initial evaluations, one for comparison with previous work with GFNNs, and the second is testing dynamic pulses with accelerando and ritardando. The experiment used our own open-source *PyGFNN* python library, which contains GFNN and AFNN implementations².

5.1 Evaluation

As we have argued above (see Section 2.4), we believe that when predicting pulse, phase is an important aspect to take into account. Phase information in the time domain also contains frequency information, as frequency equates the rate of change in phase. Therefore our evaluation compares the weighted phase output (Φ) with a ground truth phase signal similar to an inverted beat-pointer model [24]. While a beat-pointer model linearly falls from 1 to 0 over the duration of one beat, our inverted signal rises from 0 to 1 to represent phase growing from 0 to 2π in an oscillation. The entrainment behaviour of the canonical oscillators will cause phase shifts in the network, therefore the phase output should align to the phase of the input.

To make a quantitative comparison we calculate the Pearson product-moment correlation coefficient (PCC) of the two signals. This gives a relative, linear, mean-free measure of how close the target and output signals match. A value of 1 represents a perfect correlation, whereas -1 indicates an anti-phase relationship. Since the AFNN and GFNN operate on more than one metrical level, even a small positive correlation would be indicative of a good frequency and phase response, as some of the signal represents other metrical levels.

5.2 Results

Figure 6 shows the results for the pulse detection experiment described above in the form of box plots.

We can observe from Figure 6a that the GFNN (A) is effective for tracking isochronous rhythms. The resonance has enough strength to dominate the interference from the other oscillators. The low density GFNN (B) performs significantly worse with little positive correlation and some negative correlation, showing the importance of having a dense GFNN. The outliers seen can be explained by the randomised tempo; sometimes by chance the tempo falls into an entrainment basin of one or more oscillators. Despite its low density, the AFNN (C) fairs as well as the GFNN, showing a matching correlation to the target signal, especially in the upper quartile and maximum bounds. Exploring more values for ϵ_f and ϵ_h may yield even better results here.

² <https://github.com/andyr0id/PyGFNN>

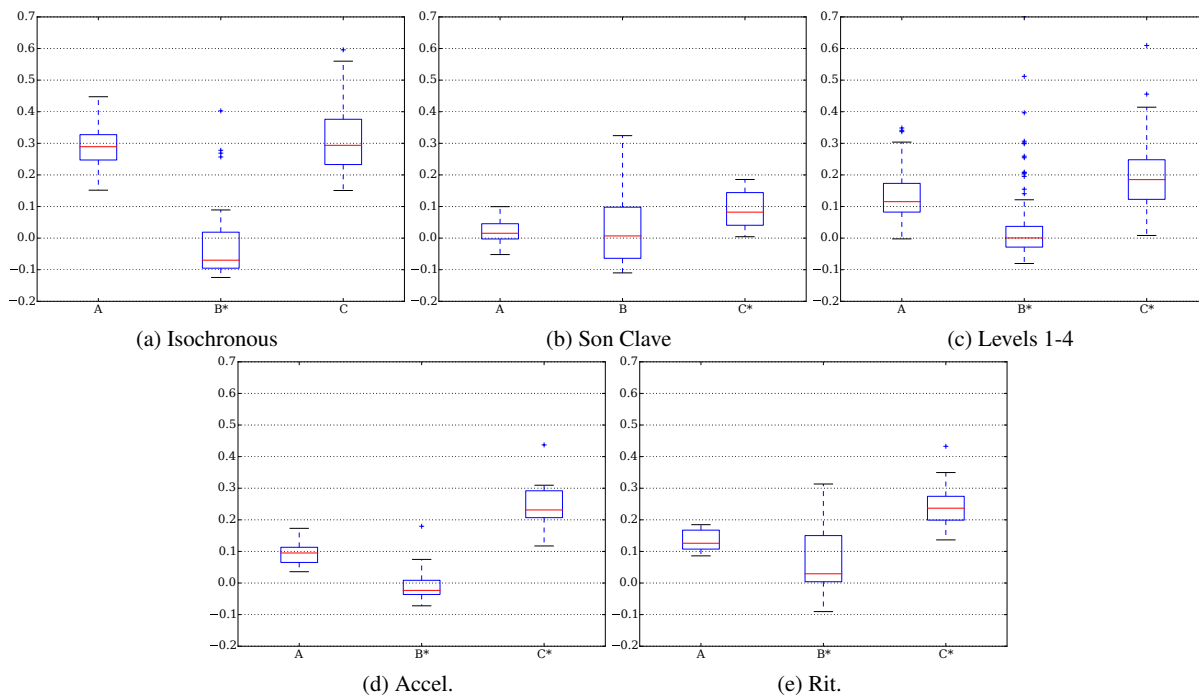


Figure 6. Box and Whisker plots of the PCC results. A) GFNN, B) Low density GFNN, C) AFNN. Boxes represent the first and third quartiles, the red band is the median, and whiskers represent maximum and minimum non-outliers. *Denotes significance in a Wilcoxon signed rank test ($p < 0.05$).

In the son clave results (Figure 6b) all networks perform poorly. A poorer result here was expected due to the difficulty of this rhythm. However, we can see a significant improvement in the AFNN, which may be due to the reduced interference in the network. In the Large et al. rhythms (Figure 6c) we notice the same pattern.

We can see from the Accelerando and Ritardando rhythms (Figure 6d and 6e) that Φ is poorly correlated, indicating the affect of the interference from other oscillators in the system. The AFNNs response shows a significant improvement, but still has low minimum values. This may be due to the fact that the adaptive rule depends on the amplitude of the oscillator, and therefore a frequency change may not be picked up straight away. Changing the oscillator model parameters to introduce more amplitude damping may help here. Nevertheless the AFNN model still performs significantly better than the GFNN, with a much lower oscillator density.

6. CONCLUSIONS

In this paper we proposed a novel Adaptive Frequency Neural Network model (AFNN) that extends GFNNs with a Hebbian learning rule to the oscillator frequencies, attracting them to local areas of resonance. Where previous work with GFNNs focused on frequency and amplitude responses, we evaluated the outputs on their weighted phase response, considering that phase information is critical for pulse detection tasks. We conducted an experiment partially reproducing Velasco and Large’s [23] and Large et al.’s [17] studies for comparison, adding two new

rhythm categories for dynamic pulses. When compared with GFNNs, we showed an improved response by AFNNs to rhythmic stimuli with both steady and varying pulse frequencies.

AFNNs allow for a great reduction in the density of the network, which can improve the way the model can be used in tandem with other machine learning models, such as neural networks or classifiers. Furthermore the system functions fully online for use in real time. In future we would like to explore this possibility by implementing a complete beat-tracking system with an AFNN at its core.

We have a lot of exploration to do with regards to the GFNN/AFNN parameters, including the testing values for the adaptive frequency rule, oscillator models and internal connectivity. The outcome of this exploration may improve the results presented here.

The mode-locking to high order integer ratios, nonlinear response, and internal connectivity set GFNNs apart from many linear filtering methods such as the resonating comb filters and Kalman filters used in many signal prediction tasks. Coupled with frequency adaptation we believe that that the AFNN model provides very interesting prospects for applications in MIR and further afield. In future we would like to explore this possibility by implementing a complete beat-tracking system with an AFNN at its core and perform an evaluation with more realistic MIR datasets.

7. ACKNOWLEDGEMENTS

Andrew J. Lambert is supported by a PhD studentship from City University London.

8. REFERENCES

- [1] Paul E. Allen and Roger B. Dannenberg. Tracking musical beats in real time. In *Proceedings of the 1990 International Computer Music Conference*, pages 140–3, San Francisco, CA, 1990.
- [2] Vassilis Angelis, Simon Holland, Paul J. Upton, and Martin Clayton. Testing a Computational Model of Rhythm Perception Using Polyrhythmic Stimuli. *Journal of New Music Research*, 42(1):47–60, 2013.
- [3] Sebastian Böck, Florian Krebs, and Gerhard Widmer. Accurate tempo estimation based on recurrent neural networks and resonating comb filters. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, pages 625–31, Malaga, Spain, 2015.
- [4] Simon Grondin. *Psychology of Time*. Emerald Group Publishing, 2008.
- [5] Peter Grosche, Meinard Müller, and Craig Stuart Sapp. What Makes Beat Tracking Difficult? A Case Study on Chopin Mazurkas. In *Proceedings of the 11th International Society for Music Information Retrieval Conference, 2010*, pages 649–54, Utrecht, Netherlands, 2010.
- [6] Andre Holzapfel, Matthew E.P. Davies, José R. Zapata, João L. Oliveira, and Fabien Gouyon. Selective Sampling for Beat Tracking Evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9):2539–48, 2012.
- [7] Mari R. Jones. Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83(5):323–55, 1976.
- [8] Rudolph E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.
- [9] Richard Kempter, Wulfram Gerstner, and J Leo Van Hemmen. Hebbian learning and spiking neurons. *Physical Review E*, 59(4):4498–514, 1999.
- [10] Anssi P. Klapuri, Antti J. Eronen, and Jaakko T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):342–55, 2006.
- [11] Andrew Lambert, Tillman Weyde, and Newton Armstrong. Beyond the Beat: Towards Metre, Rhythm and Melody Modelling with Hybrid Oscillator Networks. In *Joint 40th International Computer Music Conference and 11th Sound & Music Computing conference*, Athens, Greece, 2014.
- [12] Andrew Lambert, Tillman Weyde, and Newton Armstrong. Studying the Effect of Metre Perception on Rhythm and Melody Modelling with LSTMs. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.
- [13] Andrew J. Lambert, Tillman Weyde, and Newton Armstrong. Perceiving and Predicting Expressive Rhythm with Recurrent Neural Networks. In *12th Sound & Music Computing Conference*, Maynooth, Ireland, 2015.
- [14] Edward W. Large. Beat tracking with a nonlinear oscillator. In *Working Notes of the IJCAI-95 Workshop on Artificial Intelligence and Music*, pages 24–31, 1995.
- [15] Edward W. Large. Neurodynamics of Music. In Mari R. Jones, Richard R. Fay, and Arthur N. Popper, editors, *Music Perception*, number 36 in Springer Handbook of Auditory Research, pages 201–231. Springer New York, 2010.
- [16] Edward W. Large, Felix V. Almonte, and Marc J. Velasco. A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*, 239(12):905–11, 2010.
- [17] Edward W. Large, Jorge A. Herrera, and Marc J. Velasco. Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, 9(159), 2015.
- [18] Edward W. Large and John F. Kolen. Resonance and the Perception of Musical Meter. *Connection Science*, 6(2-3):177–208, 1994.
- [19] Fred Lerdahl and Ray Jackendoff. *A generative theory of tonal music*. MIT press, Cambridge, Mass., 1983.
- [20] J. Devin McAuley. *Perception of time as phase: Toward an adaptive-oscillator model of rhythmic pattern processing*. PhD thesis, Indiana University Bloomington, 1995.
- [21] Ludovic Righetti, Jonas Buchli, and Auke J. Ijspeert. Dynamic hebbian learning in adaptive frequency oscillators. *Physica D: Nonlinear Phenomena*, 216(2):269–81, 2006.
- [22] Eric D. Scheirer. Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103(1):588–601, 1998.
- [23] Marc J. Velasco and Edward W. Large. Pulse Detection in Syncopated Rhythms using Neural Oscillators. In *12th International Society for Music Information Retrieval Conference*, pages 185–90, Miami, FL, 2011.
- [24] Nick Whiteley, Ali T. Cemgil, and Simon J. Godsill. Bayesian Modelling of Temporal Structure in Musical Audio. In *Proceedings of the 7th International Society for Music Information Retrieval Conference*, pages 29–34, Victoria, Canada, 2006.