Embodied Tempo Tracking with a Virtual Quadruped

Alex Szorkovszky*, Frank Veenstra, Olivier Lartillot, Alexander Refsum Jensenius, Kyrre Glette
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo, Norway
*alexansz@ifi.uio.no

ABSTRACT

Dynamic attending theory posits that we entrain to time-structured events in a similar way to synchronizing oscillators. Hence, a tempo tracker based on oscillators may replicate humans' ability to rapidly and robustly identify musical tempi. We demonstrate this idea using virtual quadrupeds, whose gaits are controlled by oscillatory neural circuits known as central pattern generators (CPGs). The quadruped CPGs were first optimized for flexible gait frequency and direction, and then an additional recurrent layer was optimized for entrainment to isochronous pulses. Using excerpts of musical pieces, we find that the motion of these agents can rapidly entrain to simple rhythms. Performance was found to be partially predicted by pulse entropy, a measure of the sample's rhythmic complexity. Notably, in addition to having wide tempo ranges, the best performing agents can also entrain to rhythms that are periodic but not quantized on a grid. Our approach offers an embodied alternative to other dynamical systems-based approaches to entrainment, such as gradient-frequency arrays. Such agents could find use as participants in virtual musicking environments, or as real-world musical robots.

1. INTRODUCTION

Human rhythmic abilities owe a great deal to our capacity for sensory-motor synchronization. In part due to the tight coupling between the auditory and motor regions of the brain, we spontaneously feel an urge to move when listening to music, and to track changes in tempo [1–3]. This tendency for bodily and/or neural oscillations to match a stimulus is known as rhythmic entrainment, and is the subject of intense research in music cognition [4, 5].

Beat tracking and tempo estimation, as with many other kinds of computing, have recently seen advances inspired by neuromorphic models of human cognition. While classically this has been achieved by signal processing methods such as Fourier transforms or autocorrelation [6], new methods employ techniques such as recurrent and convolutional neural networks [7, 8]. The seminal work of Large et al. [9, 10] provided a neurobiological model for rhythmic entrainment in particular, based on dynamic attending theory. While this model superficially resembles a bank of resonators, another common beat tracking technique, the use of self-organizing dynamical systems more accurately captures the human ability to discern beats from complex signals, in particular where the beat frequency is missing from the spectrum.

In this paper, we extend this self-organizing approach to bodily entrainment by using evolved virtual quadruped robots. These agents are controlled by central pattern generators (CPGs), neural circuits that organize periodic actions in vertebrates. Notably, CPGs for locomotion often have highly flexible periods and gait patterns, to allow for a range of movement speeds. In previous work, it was demonstrated that optimizing for gait flexibility facilitates entrainment to external rhythms [11]. Here, we analyze in detail the entrainment capabilities of these agents. In particular, we examine tempo ranges, uneven pulses, complexity in the context of real musical samples, and transient behaviour for real-time applications.

2. VIRTUAL ROBOT MODEL

Figure 1 shows the simulated quadruped body and a schematic of its controller layout. The simulation, controller and optimization are detailed fully in [11]. The controller is implemented in Python, while the robot simulation is implemented in Unity. [1]

The CPG consists of 12 spiking neurons based on the Matsuoka model with mutual connections [12], akin to biological neurons with inhibiting and excitatory connections [13, 14]. The 12 neurons are arranged in four modules with identical parameters and weights. The limbs are connected via the interneurons (I), with weights satisfying lateral symmetry, while the neurons labelled A and B are used to drive the hip and knee joints of the legs. Measurements of the body tilt from the simulation are fed back to the motor (A/B) neurons to stabilize the motion.

Characteristics of the motion—namely frequency and gait type—are largely determined by a constant input applied to all neurons. This control parameter models the slow input from the brain stem that is known to modulate locomotive CPGs [14, 15].

Neuron parameters, interconnection weights and feedback coefficients were all optimized using a multi-objective evolutionary algorithm (MOEA) [16]. Rather than optimiz-

[1] The code used and datasets generated for this study, as well as an example video, can be found at:
https://github.com/aszorko/COROBOREES/tree/Paper3

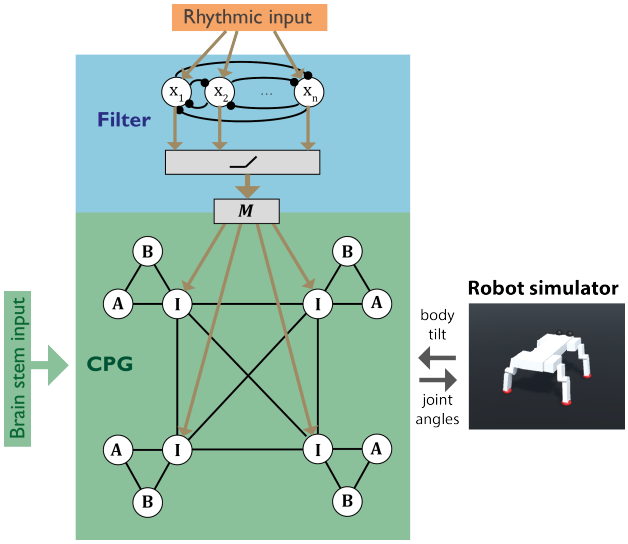

Figure 1. Controller schematic and simulated quadruped. Circles denote modified Matsuoka neurons (n in the filter component, and 12 in the central pattern generator). Arrows denote one-way connections (either excitatory or inhibitory), lines ending in circles denote inhibitory connections, and regular lines denote mutual connections that may be excitatory or inhibitory. The modules are connected via n rectifying linear units, followed by an n by 4 weight matrix M. A/B: motor neurons; I: interneurons. For this study, $n = 6$.

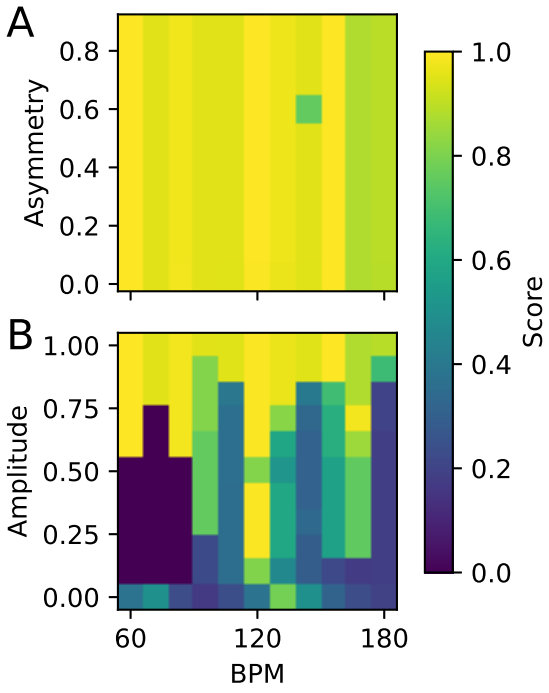

Figure 2. Entrainment score vs pulsed input tempo and (A) asymmetry, and (B) amplitude, for the top-performing agent.

ing a single controller for a single outcome, this method evolves a population of controllers so that each member optimizes a different combination of outcomes from a set, thus encouraging diversity of solutions. This was used to target flexibility of motion with respect to the brain stem input and another parameter controlling the body's centre of mass.

External input was then fed through a recurrent layer comprising of 6 fully connected neurons (the "filter"). This layer was optimized using a MOEA on a subset of CPGs by inputting isochronous pulses at three different tempi, with every fourth pulse missing, and with a small amount of noise in the pulse timings. Pulses were low-pass filtered delta functions, with a decay rate optimized separately for each agent. The objective (or "fitness") for the filter optimization was closeness of the inter-pulse interval and gait period for each tempo. The gait period in [11] and this study was determined by the location of the maximum sum of autocorrelation functions over all four limbs.

Of the two morphologies in the aforementioned paper, we use the short-legged variety for the present analysis due to its more stable behaviour. The 18 final controllers were used in the present work. These agents used either walking, trotting or bounding gaits to move.

The time constant of the CPG was 7.5 ms, while communication with the Unity simulation occurred every 100 ms for the isochronous pulse analysis (where the pulses are introduced in Python), and every 15 ms for the audio sample analysis (where the samples are stored in the Unity simulation). All simulations ran for 24 seconds and were run five times, with the median used as the final measurement.

3. ENTRAINMENT SCORE

The entrainment score in this paper was designed to measure the proximity of the measured gait period T to the input period T_{in}, allowing for halving, doubling or quadrupling of the period:

$$\text{score} = \begin{cases} \left(1 + \frac{|r-[r]|}{\epsilon}\right)^{-1}, & \text{if } -1.5 < r < 2.5 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where

$$r = \log_2(T/T_{in}) \,, \quad (2)$$

$[]$ denotes rounding to the nearest integer, and ϵ was set to 0.1 for this study.

An agent with good entrainment capability can be defined in two ways. Firstly, it could entrain to a wide range of pulse frequencies with the same level of brain stem drive (Definition 1). Secondly, it could have a wide range of intrinsic gait frequencies as the brain stem drive is modified, that are all able to be entrained to the same pulse tempo (Definition 2).

We sort the 18 robots according to Definition 1, using isochronous pulses ranging from 60 to 180 beats per minute with an amplitude of 1 (i.e. the same amplitude used during evolution), and taking the mean of the score according to Equation 1.

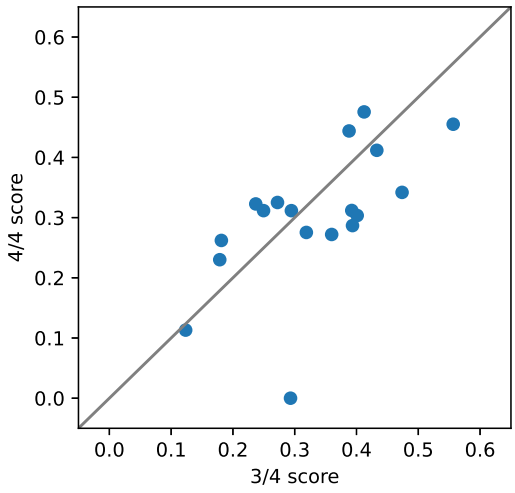

Figure 3. Comparison of 3/4 and 4/4 performance for each of the 18 robots. The score is the median over the range 60 to 180 bpm, with an amplitude of 1. The solid diagonal line has slope 1, indicating which meter had better performance for each robot.

4. ENTRAINMENT TO ISOCHRONOUS AND UNEVEN PULSES

To test robustness of the entrainment, we swept pulse amplitude and asymmetry in addition to tempo. Asymmetry was created by moving every second pulse a fraction of a beat, between 0.1 and 0.9. Apart from an asymmetry of 0.5, creating alternating dotted quarter notes and eighth notes, these asymmetries did not conform to binary metrical hierarchies.

Figure 2 shows the score for the best performing agent as a function of beats per minute against (A) asymmetry (amplitude $= 1$) and (B) amplitude (asymmetry $= 0$). While the entrainment degraded for lower amplitudes, it was in general able to entrain to asymmetric patterns. Performance was reduced slightly at higher tempos, likely due to the period approaching the CPG time-step interval, which is the limit of precision in the output period.

5. EFFECT OF METER

It is also worth asking whether the quadruped agents are sensitive to a rhythm's meter. To answer this, we tested performance against two repeated rhythms of comparable complexity, one in 3/4 (♩ ♩. 𝅘𝅥𝅯) and one in 4/4 (♩ ♩. ♩). Both rhythms contain three onsets per measure, and for each rhythm only one inter-pulse interval per measure corresponds to the beat (quarter note) duration.

As shown in Figure 3, a majority of agents had similar performance for each rhythm, particularly when compared to the variability between agents. The non-parametric Mann-Whitney U test did not find a significant difference between the two sets of scores (p=0.35).

6. RESPONSE TIME

The measurements in the preceding sections were made using autocorrelation functions over the second half of the simulation in order to determine the stable gait period. For real-time applications, however, the transient properties of the system with changing inputs are important. Hence, we use wavelet transforms to measure the time to entrain to a new input, the time to adjust to a change in tempo and the time to relax back to the agent's intrinsic tempo after the input is turned off.

The time series for the leg joint angles were convolved with a Morlet wavelet at the input period, with a resolution parameter $\sigma = 2$, and then a Gaussian filter was applied with a width of 0.5 s. As shown in Figure 4, the trotting gait adjusts to each of the input pulse tempi, and then back to its intrinsic tempo. A video of this agent responding to a similar sequence, two consecutive drum loops of different tempi, is available on this article's repository (see Footnote 1).

To estimate the time taken to adjust to the input, the second derivative of the synchronization level was calculated, and the minimum was taken in the three second period after each transition time. According to this method, adjusting to the initial input took approximately 1.55 s, adjusting to the tempo change took 0.78 s, and relaxing back to the normal state took 0.89 s

7. ENTRAINMENT VS COMPLEXITY IN MUSICAL SAMPLES

Fifteen popular music excerpts, shown in Table 1, were taken from [17], in which urge to move in humans was measured as a function of rhythmic complexity. [2] Complexity was measured by pulse clarity [18], which is derived from the entropy of the autocorrelation function of the amplitude envelope. A lower entropy measure implies a clearer and more icochronous pulse. Here, we measure the entrainment performance as a function of both pulse clarity and a more recent measure of rhythmic complexity, metrical strength [19]. The latter extracts a hierarchy of metrical levels from the autocorrelation function, and increases with a more consistent pulse at any metrical level. Both values were measured from each audio sample using MIR toolbox version $1.8.1$ [20], using the *mirpulseclarity 'EntropyAutocor'* and *mirmetroid* functions respectively.

Samples were normalized to have equal average RMS levels. In the simulation, the RMS level was calculated every 30 ms within a window of the same length, and this was passed to the robot controller, without the low-pass filter that was used for isochronous pulses.

Since the samples have nearly uniform tempos (112 to 130 beats per minute), we used Definition 2 from Section 3 for entrainment performance. For the top five agents determined in Section 3 according to Definition 1, the brain

[2] In [17], two samples were edited to provide clearer pulses in the amplitude envelopes, otherwise hidden in spectral information. We use the unedited versions, as spectral information was not available to our agent. However, we leave out "South of Heaven" since in this sample the measurable onsets are no more frequent than 60 BPM. Pulse entropy values also differ slightly, as we trimmed the ends of the samples to allow seamless looping in Unity.

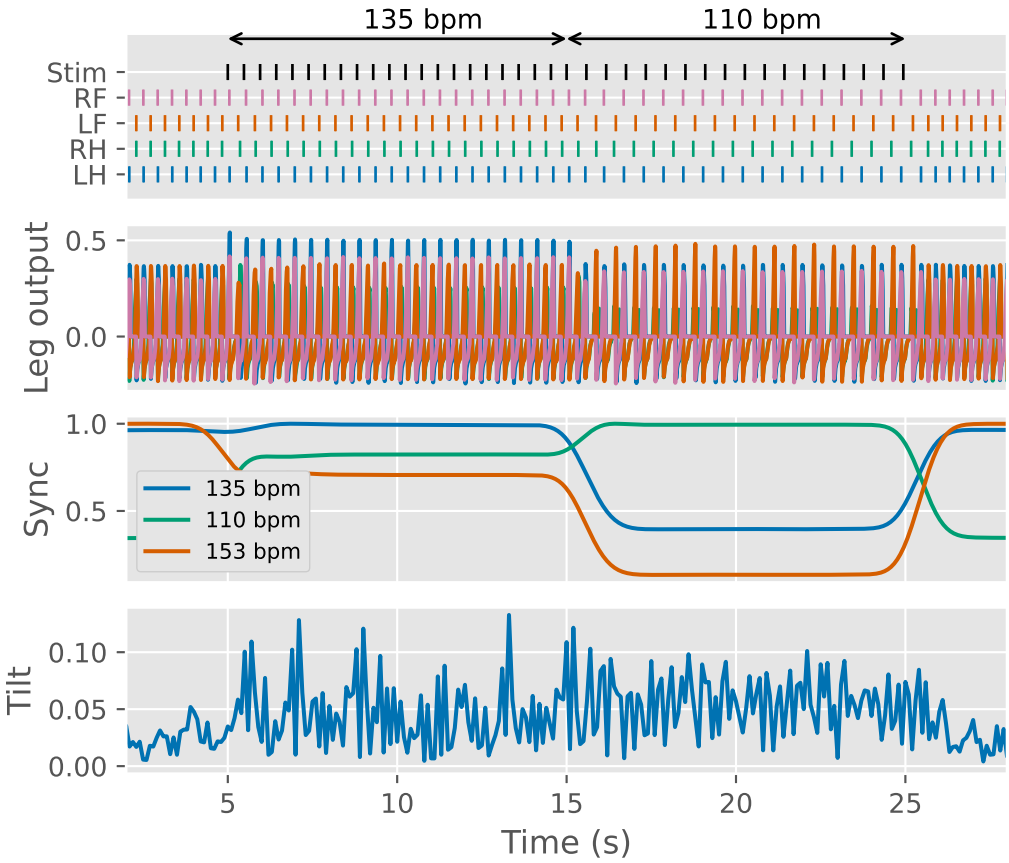

Figure 4. Transient response of the best performing agent (intrinsic period 153 bpm) to changing inputs. Pulsed input at 135 bpm was applied between 5 and 15 seconds, followed by 110 bpm between 15 and 25 seconds. Both inputs had amplitude 1.2. From top: Pulsed input and step timings (R/L=right/left limbs, F/H=front/hind limbs), leg angle output signals, wavelet convolution outputs (Sync) at 135 bpm and 110 bpm and the intrinsic 153 bpm, and total body tilt. The maximum sync (normalized to equal 1) corresponds to synchronization at that tempo.

stem drive was iterated from 0 to 1 in steps of 0.1, and the median entrainment scores to the musical excerpts were measured. A score was also measured for a no-input condition, to identify whether agents "accidentally" moved at the approximately correct tempo (120 bpm) through the whole range due to low frequency flexibility.

Figure 5 shows how the five best agents as defined above performed as a function of pulse entropy, compared to isochronous pulses at 120 bpm, and a no-input condition. All five agents had perfect scores for the isochronous input. A linear mixed-effects model with random slope and intercept (grouped by agent) was used to test whether either complexity measure predicted the decreasing performance. The EntropyAutocor measure of pulse clarity was found to be significant predictor of performance ($z = -3.3, P = 0.001$). Metrical strength was not significant at the $P < 0.05$ level ($z = -1.5, P = 0.14$).

The pulse clarity result indicates that while high complexity samples caused a score of either zero (generally indicating chaotic motion) or close to the no-input score, the lower-complexity inputs were more able to have the tempo reliably tracked by the agents. A notable exception was "Under Attack", which may be attributable to the mixed 3/4 and 4/4 meter.

8. DISCUSSION

We have demonstrated a proof of concept for virtual and physical agents with spontaneous entrainment of motor-sensory systems. A subset of our evolved virtual robots are able to track the tempo of simple rhythms, including real audio samples, in real time with fast adaptation. This can be attributed to highly non-linear dynamical networks that can adapt their endogenous oscillations to absorb periodic inputs. This approach is similar to the work of Large et al. [9, 10]. Notably, however, our oscillatory network is modelled on spinal circuits involved directly in motor control rather than higher-level cortical networks. This makes the translation to motion of an agent — in particular a robotic agent — natural while avoiding delays in the process. Hence, such agents may be useful for biomorphic visualization of musical rhythms.

The agents shown here can also provide a responsive real or virtual musical partner for the controlled study of interpersonal synchronization. Instantaneous response is crucial in musical contexts, however existing robotic partners that incorporate adaptive timing (e.g. [21]) do so using signal processing methods that require relatively long sampling times and computation. As a consequence, such sys-

Title	Artist	Excerpt	BPM	Genre	PC-E	MS	Avg. Score
Off The Wall	Michael Jackson	0:15-0:23	119	Pop/funk	0.525	0.604	0.699
Under Attack	ABBA	0:00-0:08	116	Pop	0.526	0.795	0.318
I Wanna Be Your Lover	Prince	0:00-0:08	117	Pop/funk	0.568	0.574	0.479
Voyager	Daft Punk	0:32-0:48	120	EDM	0.596	0.720	0.767
War of My Life	John Mayer	0:00-0:08	120	Rock	0.616	0.661	0.381
Sugar	Maroon 5	0:40-0:48	120	Rock	0.624	0.560	0.901
Get It Right	Aretha Franklin	0:11-0:19	121	Soul/funk	0.653	0.608	0.558
I Got the Feeling	James Brown	0:00-0:03	114	Soul/funk	0.706	0.587	0.303
For Whom the Bell Tolls	Metallica	0:57-1:05	120	Metal	0.711	0.567	0.619
What About Me	Snarky Puppy	0:15-0:23	127	Jazz/funk	0.731	0.505	0.297
Every Breaking Wave	U2	0:00-0:08	116	Rock	0.752	0.808	0.349
Getaway	Earth, Wind & Fire	0:00-0:09	112	Funk	0.753	0.486	0.589
Peggy	Orchards	0:00-0:08	130	Alt. pop	0.762	0.577	0.242
Pinzin Kinzin	Avishai Cohen Trio	2:01-2:05	116	Jazz/ Experimental	0.767	0.422	0.319
Smash	Avishai Cohen	0:08-0:16	115	Jazz/ Experimental	0.794	0.538	0.305

Table 1. Musical excerpts used in the study, from [17], and mean entrainment scores for the top five agents. PC-E: Pulse clarity - EntropyAutocor measure; MS: Metrical strength. Samples are ordered by increasing complexity according to pulse entropy. All excerpts were in 4/4 apart from "Under Attack" which was a mixture of 4/4 and 3/4.

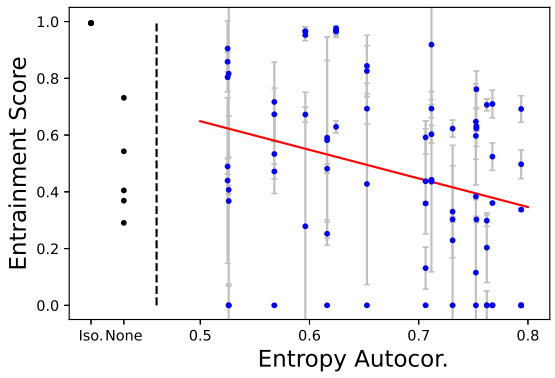

Figure 5. Entrainment vs pulse clarity *EntropyAutocor* measure for the top five agents. None: no input. Iso.: isochronous pulses at 120 bpm. Each blue dot is one agent's mean score for a single sample. Error bars indicate the mean range between maximum and minimum score over five iterations with the same agent, sample and brain stem drive. The solid line is the fit from the final linear mixed-effects model.

tems often rely on turn-taking and/or human reading of robotic gestures [22]. We expect rapidly responsive tempo tracking to greatly enhance the experience of human-robot musical interactions.

Our model may also help research in music cognition. As a biologically inspired and mechanistic sensorimotor synchronization model, it may be useful as a comparison in experiments such as tapping studies [4, 17]. An existing modelling approach for tapping experiments involves coupled harmonic oscillators [23], while integrate-and-fire type models are another potential candidate [24]. These approaches are well suited to isochronous rhythms, however more complex tasks require more nonlinear oscillators or neural models such as the one we present here.

We found that pulse clarity was a more suitable measure than metrical strength for rhythmic complexity in our case. This is unsurprising, as the agents were optimized using isochronous pulses, which maximize pulse clarity (i.e. minimise entropy in the autocorrelation). In addition, the oscillations of the CPG network are not expected to be susceptible solely to hierarchically organised rhythms, as shown in Section 4.

Our agents were optimized using evolutionary algorithms. Hence, performance is partially determined by how thoroughly fitness is measured. In particular, performance was unpredictable for input amplitudes lower than that used during evolution. For music with high dynamic range, this issue may be mitigated by pre-processing with a compression algorithm.

If used as a tempo tracker, the ouput tempo would need to be limited to a factor of two, e.g. 75-150 bpm, due to the tendency for frequency doubling or halving depending on the affordances of the agent; i.e. with such limits in place, a gait at 60bpm or 240bpm can be interpreted as 120bpm.

For precise prediction of beat timings and not just tempo estimation, a phase-correcting feedback would be necessary. This is because it takes time for perception to be translated into action in our open-loop framework. To this end, the integration of force feedback from the feet may be sufficient to synchronize timings [25].

Acknowledgments

This project has received funding from the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement 101030688,

and is partially supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262762. We would like to thank Connor Spiech for providing the audio samples, David Løberg Code for useful discussions, and the reviewers for helpful feedback.

9. REFERENCES

[1] B. Burger and P. Toiviainen, "Embodiment in electronic dance music: Effects of musical content and structure on body movement," *Musicae Scientiae*, vol. 24, no. 2, pp. 186–205, 2020.

[2] A. Zelechowska, V. G. Sanchez, and A. R. Jensenius, "Standstill to the'beat' differences in involuntary movement responses to simple and complex rhythms," in *Proceedings of the 15th International Audio Mostly Conference*, 2020, pp. 107–113.

[3] J. J. Cannon and A. D. Patel, "How beat perception co-opts motor neurophysiology," *Trends in Cognitive Sciences*, vol. 25, no. 2, pp. 137–150, 2021.

[4] B. H. Repp and Y.-H. Su, "Sensorimotor synchronization: a review of recent research (2006–2012)," *Psychonomic bulletin & review*, vol. 20, pp. 403–452, 2013.

[5] D. J. Levitin, J. A. Grahn, and J. London, "The psychology of music: Rhythm and movement," *Annual review of psychology*, vol. 69, pp. 51–75, 2018.

[6] M. Muller, D. P. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," *IEEE Journal of selected topics in signal processing*, vol. 5, no. 6, pp. 1088–1110, 2011.

[7] S. Böck, F. Krebs, and G. Widmer, "Accurate tempo estimation based on recurrent neural networks and resonating comb filters." in *ISMIR proceedings*, 2015, pp. 625–631.

[8] H. Schreiber and M. Müller, "A single-step approach to musical tempo estimation using a convolutional neural network." in *ISMIR proceedings*, 2018, pp. 98–105.

[9] E. W. Large, "Beat tracking with a nonlinear oscillator," in *Working Notes of the IJCAI-95 Workshop on Artificial Intelligence and Music*, vol. 24031, 1995.

[10] E. W. Large, J. A. Herrera, and M. J. Velasco, "Neural networks for beat perception in musical rhythm," *Frontiers in systems neuroscience*, vol. 9, p. 159, 2015.

[11] A. Szorkovszky, F. Veenstra, and K. Glette, "Central pattern generators evolved for real-time adaptation," *arXiv preprint arXiv:2210.08102*, 2022.

[12] K. Matsuoka, "Sustained oscillations generated by mutually inhibiting neurons with adaptation," *Biological cybernetics*, vol. 52, no. 6, pp. 367–376, 1985.

[13] B. Ermentrout and D. H. Terman, *Mathematical foundations of neuroscience*. Springer, 2010, vol. 35.

[14] S. M. Danner, N. A. Shevtsova, A. Frigon, and I. A. Rybak, "Computational modeling of spinal circuits controlling limb coordination and gaits in quadrupeds," *Elife*, vol. 6, p. e31050, 2017.

[15] A. J. Ijspeert, "Central pattern generators for locomotion control in animals and robots: a review," *Neural networks*, vol. 21, no. 4, pp. 642–653, 2008.

[16] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: solving problems with box constraints," *IEEE transactions on evolutionary computation*, vol. 18, no. 4, pp. 577–601, 2013.

[17] C. Spiech, M. Hope, G. S. Câmara, G. Sioros, T. Endestad, B. Laeng, and A. Danielsen, "Sensorimotor synchronization increases groove," *PsyArXiv*, 2022.

[18] O. Lartillot, T. Eerola, P. Toiviainen, and J. Fornari, "Multi-feature modeling of pulse clarity: Design, validation and optimization." in *International Conference on Music Information Retrieval*, 2008, pp. 521–526.

[19] O. Lartillot, D. Cereghetti, K. Eliard, W. J. Trost, M.-A. Rappaz, and D. Grandjean, "Estimating tempo and metrical features by tracking the whole metrical hierarchy," in *The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15, 2013*. University of Jyväskylä, Department of Music, 2013.

[20] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International conference on digital audio effects*, vol. 237. Bordeaux, 2007, p. 244.

[21] M. Krzyżaniak, "Musical robot swarms, timing, and equilibria," *Journal of New Music Research*, vol. 50, no. 3, pp. 279–297, 2021.

[22] G. Hoffman and G. Weinberg, "Interactive improvisation with a robotic marimba player," *Autonomous Robots*, vol. 31, pp. 133–153, 2011.

[23] A. P. Demos, H. Layeghi, M. M. Wanderley, and C. Palmer, "Staying together: A bidirectional delay–coupled approach to joint action," *Cognitive Science*, vol. 43, no. 8, p. e12766, 2019.

[24] K. Nymoen, A. Chandra, K. Glette, and J. Torresen, "Decentralized harmonic synchronization in mobile music systems," in *2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2014, pp. 1–6.

[25] D. Owaki, M. Goda, S. Miyazawa, and A. Ishiguro, "A minimal model describing hexapedal interlimb coordination: the tegotae-based approach," *Frontiers in neurorobotics*, vol. 11, p. 29, 2017.