

Echo State Network Ensemble for Human Motion Data Temporal Phasing: A Case Study on Tennis Forehands

Boris Bačić✉

School of Engineering, Computer and Mathematical Sciences, Auckland
University of Technology, Auckland, New Zealand
boris.bacic@aut.ac.nz

Abstract. Temporal phasing analysis is integral to ubiquitous/“smart” coaching devices and sport science. This study presents a novel approach to autonomous temporal phasing of human motion from captured tennis activity (3D data, 66 time-series). Compared to the optimised Echo State Network (ESN) model achieving 85 % classification accuracy, the ESN ensemble system demonstrates improved classification of 95 % and 100 % accurate phasing state transitions for previously unseen motions without requiring ball impact information. The ESN ensemble model is robust to low-sampling rates (50 Hz) and unbalanced data sets containing incomplete data time-series. The demonstrated achievements are applicable to exergames, augmented coaching and rehabilitation systems advancements by enabling automated qualitative analysis of motion data and generating feedback to aid motor skill and technique improvements.

Keywords: Computational Intelligence (CI) · Sport and rehabilitation · Biomechanics · Augmented Coaching Systems (ACS) · Data analytics · Human Motion Modelling and Analysis (HMMA)

1 Introduction

Over the recent years, there has been an increase in the popularity of exergames, and ubiquitous and wearable devices that promote an active lifestyle and health monitoring functions. Components such as Inertial Measurement Units (IMUs) and video sensors are found in mobile technology but also in augmented coaching devices that may be attached to the human body or sport equipment (e.g. www.xsens.com, www.shimmersensing.com, www.babolatplay.com, and www.zepp.com). In general, to develop augmented coaching system and technology (ACST) for a specific sport, the underlying technology should be able to capture, process and quantify specific movement parameters that happened over time in an on-line or off-line fashion. Whether the purpose is to analyse swing parameters or to track leg movement, it is common for developed solutions [1] to rely on proprietary sport-specific algorithms for *temporal phasing* analysis that could extract motion patterns around the impact vibration or sound (e.g. tempo and velocity at impact). However, there is little evidence about future developments of ACST that could provide intuitive feedback to aid sport

technique similar to a coach [2–5]. This paper supports the author’s vision that future ACST will combine multi-disciplinary approaches from sport science and qualitative analysis of human motion with data science and computational intelligence to provide feedback aimed at improving motor skill acquisition and technique associated with a specific motion pattern. This paper presents a novel modelling solution to temporal phasing analysis and indexing utilising captured real-life 3D motion data time-series that can learn from data and analyse previously unseen tennis forehands. The presented generic artefacts of achieved autonomous temporal phasing analysis are integral to various models of Qualitative Movement Diagnostics (QMD) in sport science, containing preparation, action zone and follow-through motion sequence associated with ball impact or projectile release, with kicking, throwing or striking action [6].

1.1 Related Work and Prior Studies

The Echo State Network (ESN), Liquid State Machine (LSM) and Spiking Neural Network (SNN) models are examples of the “third generation” of biologically inspired neural networks [7] for temporal and spatial problem areas that are also considered as alternatives to traditional algorithmic, neuro-fuzzy, heuristic- and feature extraction-based approaches [8, 9]. Jaeger’s ESN model [8] is based on recurrent neural networks which does not require signal-to-spike train conversions for its operation. For an ESN’s training task, it is common to perform adjustment the output layer, while leaving the input layer and reservoir unchanged. The presented work is linked to and expands the earlier studies in the context of Human Motion Modelling and Analysis (HMMA) and tennis [5, 10, 11].

Prior to modern development of the third generation of neural networks, the finite-state automata theory and finite-state machines (FSM) were introduced in the early days of computing science to address temporal and spatial problem areas with a distinct historical example known as the Turing machine. In this study, it is contended that a combination of modern and ‘old’ approaches can advance computer science, showing that the ESN ensemble model orchestration can be achieved by implementing sequential control logic to advance classifier performance and reduce the need for training data. The sequential logic of FSM eliminates the need for supervised learning from data as the domain expert knowledge required for orchestration can be expressed in the state transitions control mechanism. The expected benefits from such system configurations include potential improvements in classification tasks and robustness to unbalanced data sets; avoiding false positive event classifications; and preventing output state transitions that are not possible in real-life.

1.2 Temporal Phasing Analysis: Sport Science and Tennis Backgrounds

Temporal phasing analysis is integral part of coaching practice, rehabilitation and sport science with the early observational model [12] introduced in 1984. Temporal phasing combined with computational intelligence is related to the areas of: surveillance, exergames and ACST design [5]. In today’s tennis, swings are executed with a variety

of diverse paces, rhythms, stances and individual preferences (e.g. backswing and swing preparation timings). In sport science, there is a known phenomenon of *experts' disagreements* on observed motion [6]. While the descriptive rules are more sport-specific and decision boundaries may vary between experts, phasing analysis is common to sport science and integral to a number of models of qualitative analysis of human motion. Typical examples of temporal phasing analysis in tennis are: (a) the use of replays during media coverage, during (b) coaching sessions for analysis and feedback; and (c) in real-time scenarios, where a coach can also rely on his/her temporal phasing expertise while observing a player's motion and communicating short keywords as *attention cuing*. In this paper it is contended that temporal phasing will be implemented in the near-future generations of ACST and exergames which will provide feedback to aid motor skill learning and technique acquisition. For the experimental design in this paper, it is important to notice that there is no ground truth or agreed measure for determining the exact start of the forehand swing event (as opposed to vibration or sound around the impact with the ball). Personalisation and the diversity of forehand swings may be considered as one of the challenges in coaching practice. The expert's decision in this paper was that the preparation phase starts at the racquet transitions from backward to forward movement.

2 Experimental Setup: Data Collection and Visualisation of Temporal Phasing for Supervised Machine Learning

The data set was acquired in a laboratory setting using a nine fixed-location camera motion capture system (SMART-e 900 eMotion/BTS) sampling at 50 Hz, and a set of 22 retro-reflective markers attached to anatomical landmarks on player's body. The relatively small data set (21 forehand samples) was considered of sufficient size for the purpose of this study without warranting the need for additional synthetic data. Temporal phasing of all data frames into four categories was achieved visually using 3D animated stick figure (Fig. 1). Temporal phasing output classes (0, 1, 2, 3) are labelled as ('non-event', 'preparation', 'action', 'follow-through').

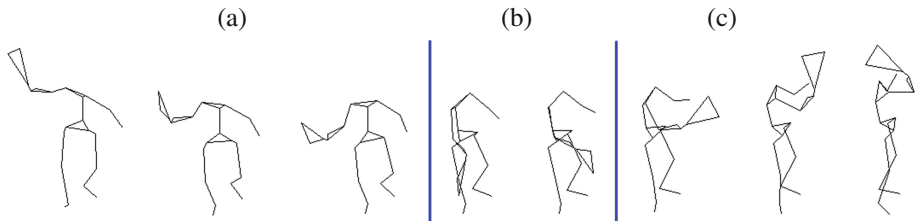


Fig. 1. Temporal phasing analysis of a forehand swing event. Distinct temporal phases for output classes (1, 2, 3) cover: (a) *preparation*, (b) *action* zone around estimated intended impact, and (c) *follow-through* post-impact recovery. The sequence of 3D stick figures is not shown at equal time distance.

3 Data Analysis and Modelling

The motion data set includes a variety of ‘good’ and ‘bad’ tennis swings, executed with diverse swing speed and from diverse stances. All tennis swings were performed by a certified tennis coach (the author of the paper) and reviewed by three other coaches. The motion data set does not include ball impact information (e.g. impact vibration pattern). The output class distribution (Fig. 2) is considered as unbalanced.

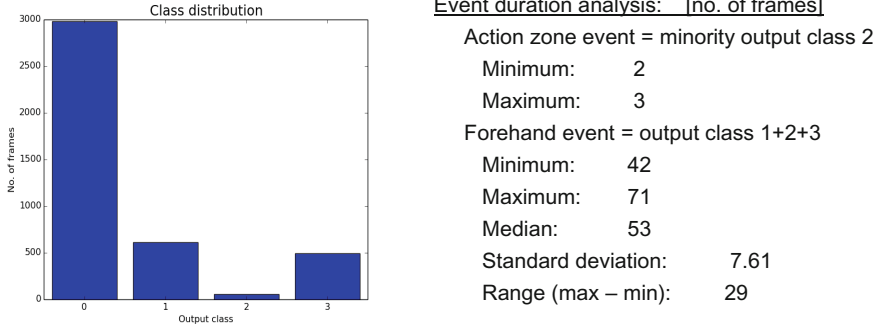


Fig. 2. Uneven class distribution where minority class typically lasts for two (or three frames for a slower swing) and majority class duration lasts for over 800 samples.

Preparing motion data for the optimised ESN model (Table 1) includes the pre-processing steps: (a) removal of the static marker time series (used for capture volume referencing and orientation purposes); (b) replacing missing values (NaN) with adjacent frames’ average values; and (c) linear normalisation within the interval $[-1, 1]$.

Given the properties of experimental data (temporal and spatial, unbalanced output class distribution, short duration of minority class, diversity of captured foreheads), the modelling task may be perceived as too complex. As a design decision in this study, it was decided to perform one grid search (Table 1) for parameter optimisation for all four data streams (leaving potential room for further improvements).

Preliminary findings (Fig. 3) show the optimised ESN model performing temporal phasing classification of 3D tennis motion data.

Table 1. Optimised parameters for leaky reservoir ESN model, utilising $\tanh(\cdot)$ for all neurons activation function

Parameter description	Optimal value	Increment	Range
Number of neurons	800	$x \pm 100$	[400, ..., 1000]
Leak rate	0.9	$x \pm 0.1$	[0.5, ..., 1]
Spectral radius	0.9	$x \pm 0.1$	[0.5, ..., 1]
Input scaling	0.0001	$x=10^{-y}, y+=1$	[0.00001, ..., 0.1]
Ridge parameter	0.0001	$x=10^{-y}, y+=1$	[0.00001, ..., 0.1]

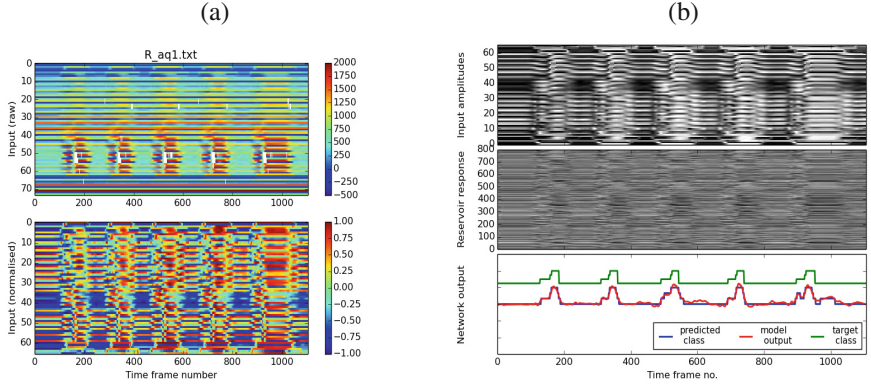


Fig. 3. Steps involved in processing and ESN classification of motion data time-series: (a) removal of static reference markers, input data pre-processing with (NaN) replacements and normalising to $[-1,1]$; and (b) a single ESN model for temporal phasing of forearm events.

Table 2. Error importance for temporal phasing from observed ESN model performance

Error importance	Description
Low	Start of preparation event
Low	End of follow-through
Moderate	Start of action zone
Moderate	End of action zone
High	False event detection
High	Impossible state transition.

The attribution of observed issues (Fig. 3) linked to a range of classification errors that warranted an ESN ensemble approach is summarised in Tables 2 and 3.

Based on visualisation of ESN 2 performance (similar to Fig. 3b), the developed ESN ensemble system (Fig. 4a and b) required only minor adjustments of the clipping threshold (CT) parameter in order to improve ESN 1 (CT = 0.55) and ESN 2 (CT = 0.4) readout signals conversion into their discrete outputs (Fig. 4a, c and d).

Table 3. Problem summary and error analysis that warranted an ensemble approach

Key issue	Error description
<ul style="list-style-type: none"> Unbalanced output class distribution Relatively low data sampling rate (50 Hz) Incomplete marker trajectories Varying duration of diverse forehands Predicted output class state transitions are in disagreement with domain expertise ESN dynamic behaviour producing ‘echoed’ and false state transitions Matching ESN readouts in the minority class 	<ul style="list-style-type: none"> Occasional appearance of missing fast moving marker trajectories as a broken time-series in captured motion data Event recognition associated with output class 2 The system produces states transitions among the output classes that appear implausible to domain experts (Fig. 4)

4 Classification Results and Discussion

All ESN models used the first 400 samples for model training (Fig. 4). Table 4 shows improved classification performance for the novel ESN ensemble system over a single ESN model.

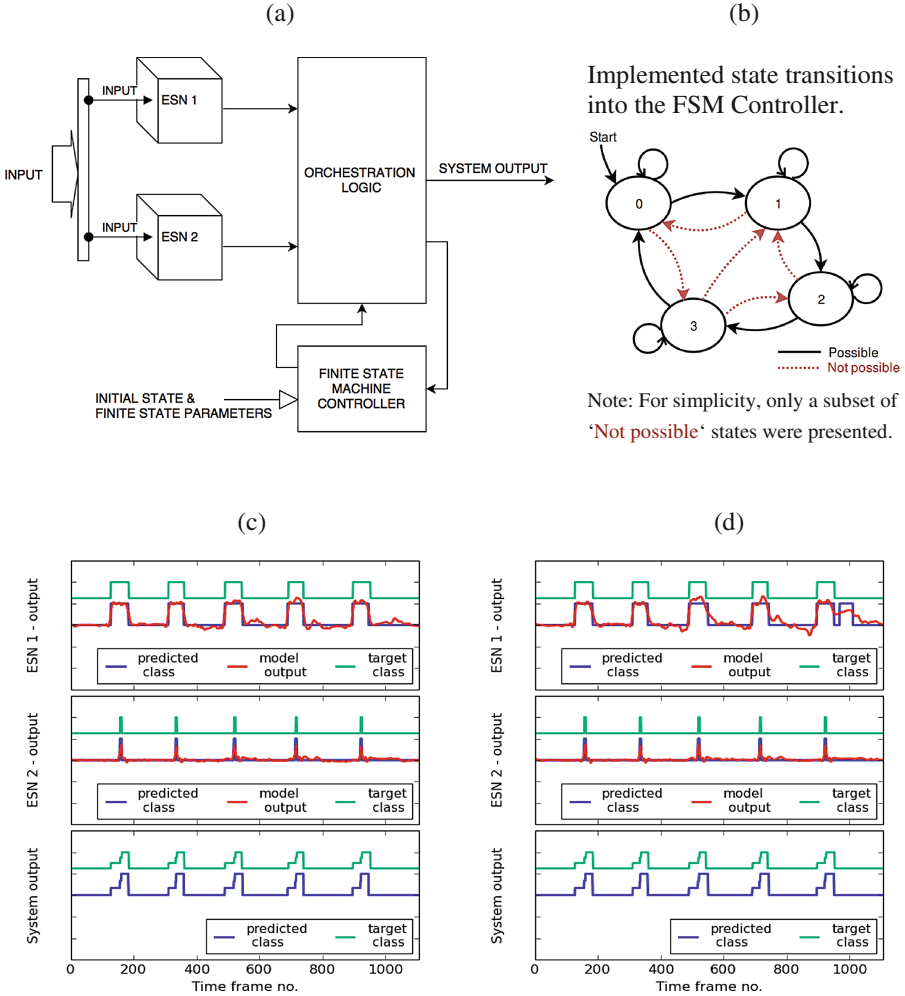

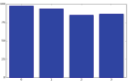


Fig. 4. Produced ESN ensemble system for autonomous temporal phasing of motion data (a); implemented state transition logic (b) into FSM controller required for orchestration of constituent ESN modules. Interim results of the system output visualisation show (c) typical system outputs and (d) robustness to occasional random classification errors (ESN1 - output, last false event) by FSM orchestration controller. ESN 2 event was extended (in training data) by 2 frames while the system output was reduced by two frames for improved accuracy at 50 Hz.

Table 4. Classification results for a single ESN and an ESN ensemble system over 12 repetitions

Experimental model	Data stream median accuracy [%]				Average class performance [%]	Average accuracy (Range)	Processing time (Range)
	1	2	3	4			
Single ESN	79.9	91.1	84.3	87.3		85.30% (14.14%)	1.46 s (0.22 s)
ESN ensemble	96.7	94.1	95.0	95.7		95.32% (3.81%)	3.20 s (0.7 s)

The achieved accuracy was above expectations, given the inclusion of all markers' trajectories, fast interpolation and normalisation, and a relatively small ESN model training portion (approx. 40 %). Regarding the validity of the exact classification of start and stop of event there are possible uncertainties in regards to expert's decision boundaries. However, it must be noted that the expectation is that, even with missing distal marker trajectories, tennis professionals should be able to identify the frame(s) with the most likely intended impact with the ball utilising the interactive functions of the developed animated 3D stick figure player [5]. The processing times were measured on a PC running Linux Ubuntu 14.04, Python 2.7 interpreter using single CPU processing with the hardware specification: CPU 4.2 GHz, RAM 16 GB/2.4 GHz, and a solid-state disk. The data set was considered similar to Microsoft's Kinect 2 in terms of occasional loss of marker time series information and relatively low sampling frequency.

5 Conclusions, Recommendations and Future Work

The presented ESN ensemble approach for autonomous temporal phasing of forehand swings required a relatively small portion of training data (approximately 40 %) and domain expertise to achieve high classification accuracy (95 %) on a relatively small data set. The ESN ensemble model has also achieved 100 % on forehand detection accuracy and temporal phasing state transitions, eliminating the occasional random occurrence of false output states from the ESN constituent models. System configuration combining ESN ensemble handling and a finite-state machine (FSM) control mechanism for orchestration purposes showed improvements compared to a single optimised ESN model, in particular: (a) in avoiding false positive event classifications; and (b) output state transitions that are not possible in real life but could otherwise be incorrectly produced from temporal and spatial data. The produced artefacts (models, architecture, and FSM controller for ESN ensemble orchestration) contribute to computer science by combining old (FSM) with modern approaches (third generation of neural networks), enabling multi-disciplinary scientific advancements for similar temporal and spatial problem areas such as computational intelligence, data analytics and sport science (e.g. qualitative movement diagnostics). From the experimental

results, it is evident that FSM orchestration is faster than the ESN approach and requires no training data. Future work will be extended to diverse human motion data contexts.

Acknowledgements. The author wishes to express his appreciation to developers of Oger Toolbox (<http://organic.elis.ugent.be/organic/engine>) and Spyder IDE (<https://pythonhosted.org/spyder/>) utilised in this study. The tennis data was obtained in the Peharec polyclinic for physical therapy and rehabilitation, Pula (Croatia) in collaboration with Petar Bačić (biomechanics lab specialist and professional tennis coach). The author also wishes to express his sincere appreciation to Dr. Stefan Schliebs and Dr. Russel Pears for their valuable comments and insights.

References

1. Lightman, K.: Silicon gets sporty. *IEEE Spectr.* **55**, 48–53 (2016)
2. Bačić, B.: Bridging the gap between biomechanics and artificial intelligence. In: XXIV International Symposium on Biomechanics in Sports - ISBS 2006, Salzburg, Austria, pp. 371–374 (2006)
3. Bacic, B.: Evolving connectionist systems for adaptive sports coaching. *Neural Inf. Process. – Lett. Rev.* **12**, 53–62 (2008)
4. Bačić, B.: Connectionist methods for data analysis and modelling of human motion in sporting activities. Ph.D. thesis, School of Computer and Mathematical Sciences, AUT University, New Zealand, Auckland (2013)
5. Bačić, B.: Prototyping and user interface design for augmented coaching systems with MATLAB and Delphi: implementation of personal tennis coaching system. In: MATLAB Conference 2015, Auckland (2015)
6. Knudson, D.V.: Qualitative Diagnosis of Human Movement: Improving Performance in Sport and Exercise. Human Kinetics, Champaign (2013)
7. Maass, W.: Networks of spiking neurons: the third generation of neural network models. *Neural Netw.* **10**, 1659–1671 (1997)
8. Jaeger, H., Lukoševičius, M., Popovici, D., Siewert, U.: Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Netw.* **20**, 335–352 (2007)
9. Paugam-Moisy, H.: Spiking neuron networks a survey. In: IDIAP Research Institute (2006)
10. Bacic, B.: Echo state network for 3D motion pattern indexing: a case study on tennis forehands. In: Bräunl, T., et al. (eds.) PSIVT 2015. LNCS, vol. 9431, pp. 295–306. Springer, Heidelberg (2016). doi:[10.1007/978-3-319-29451-3_24](https://doi.org/10.1007/978-3-319-29451-3_24)
11. Bacic, B.: Extracting player’s stance information from 3D motion data: a case study in tennis groundstrokes. In: Mori, S., et al. (eds.) PSIVT 2015 Workshops. LNCS, vol. 9555, pp. 307–318. Springer, Heidelberg (2016). doi:[10.1007/978-3-319-30285-0_25](https://doi.org/10.1007/978-3-319-30285-0_25)
12. Gangstead, S.K., Beveridge, S.K.: The implementation and evaluation of a methodical approach to qualitative sports skill analysis instruction. *J. Teach. Phys. Educ.* **3**, 60–70 (1984)

Neural Information Processing

23rd International Conference, ICONIP 2016, Kyoto,

Japan, October 16–21, 2016, Proceedings, Part IV

Akira, H.; Seiichi, O.; Doya, K.; Kazushi, I.; Minho, L.;

Derong, L. (Eds.)

2016, XIX, 663 p. 254 illus., Softcover

ISBN: 978-3-319-46680-4