CEUR-WS.org/Vol-3499/paper8.pdf

Novices make more noise! But how can we listen to it?

Jan Schneider[1], Khaleel Asyraaf Mat Sanusi[2] , Bibeg Limbu[3] and Nghia Duong-Trung[4]

[1] DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, Germany
[2] Cologne Game Lab, TH Köln, Cologne, Germany
[3] University of Duisburg-Essen, Duisburg, Germany
[4] German Research Center for Artificial Intelligence (DFKI), Berlin, Germany

Abstract
This paper presents an extension to the approach described in [13] which was designed to help distinguish expert and novice performance easily by observing the sensor data without having to understand nor apply models to the sensor signal. The method consisted of plotting the sensor data and identifying irregularities in novice data and regularities in expert data. In this paper, we solidify the thesis that, with the help of sensors, expert performances are smoother, contain fewer irregularities, and have consistently uniform patterns than novice performances. We do so using the extended methodology on the same data set from the previous five cases in [13], namely running, bachata dance, salsa dance, tennis swings, and football penalty kicks, pointing out this assertion.

Keywords
Multimodal Learning Analytics, Sensors, Equity, Signal Interpretation, Diversity, and Inclusion

1. Introduction

As smart devices with a plethora of built-in sensors have become ubiquitous, they are becoming more prevalent in "Human Learning". Such smart devices, along with their sensing technologies, aid in the collection of important data and the provision of feedback to learners in the cognitive, affective, and psychomotor domains of learning [1,2,3]. Furthermore, sensing technologies can also be used to study, record/model expert performance, and use it to develop expertise [4,5].

However, in most learning scenarios, the stream of data captured by one sensor is insufficient to meaningfully comprehend learning. For example, in public speaking, the voice, words, gestures, and posture of the presenter should be congruent. Therefore, to train public speaking effectively, multiple modalities, and therefore multiple sensors, need to be used to capture the learning performance. This compounds the complexity of interpreting sensor data which is already complicated for a single sensor. Di Mitri et al., [6] propose a model to make sense of the multimodal data through machine learning and use the machine learning predictions to provide feedback to learners. This model has already been used to predict different Table Tennis strokes [3], identify task-switching performance based on physiological markers [7], develop learning applications to train cardiopulmonary resuscitation [8], etc.

However, there are recurrent challenges associated with developing a multimodal learning solution using the model [9]. For example, the model does not provide an out-of-the-box solution that is easy to implement. Developing a multimodal learning solution continues to be time-

MILeS 23: The Third International Workshop on Multimodal Immersive Learning Systems, September 4, 2023, Aveiro, Portugal
j.schneider@dipf.de (J. Schneider); ks@colognegamelab.de (K. A. Mat Sanusi); bibeg.limbu@uni-due.de (B. Limbu); nghia_trung.duong@dfki.de, (N. Duong-Trung)
0000-0001-8578-6409 (J. Schneider); 0000-0001-6766-4416: (K. A. Mat Sanusi); 0000-0002-1269-6864 (B. Limbu); 0000-0002-7402-4166 (N. Duong-Trung)
© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

consuming, tedious, and difficult to get enough accurate annotated recordings to train machine learning models capable of making useful predictions using multimodal data, despite following pragmatic approaches [10], using customizable tools to collect [11] and annotate multimodal data [12].

In this paper, we present an extension of the preliminary study [13] where we tested a completely different approach that might help to quickly and simply distinguish expertise levels based on sensor data. In [13], we hypothesized that experts display consistent and uniform differences from novices in their performance as a consequence of their repeated practice and extended experience. To test this hypothesis we plotted the sensor recordings of expert performance and the novice performance in various psycho-motor domains. The plots displayed recognizable regularities/patterns, compared to the chaos in the novice plots, which is in line with the findings of [14]. While the study in [13] showed visible differences, automatically

2. Method

To test our hypothesis in [13], we recorded expert and novice performances in different tasks using accelerometers in smartphones. The tasks that we recorded were the basic Bachata steps, basic Salsa steps, tennis swings, football penalty kicks, and running.

For the **running** case, we recorded the expert performance of a competitive amateur runner with more than two decades of regular running experience. The novice performance was captured from a participant who runs occasionally and has participated in a few races. To maintain consistency in recording, both participants held the smartphone in their left hand while running on a treadmill at a speed of 12 km/h for one minute.

In the case of **Bachata** steps, an expert teacher performed the fundamental steps, while a novice, who had no prior experience in Bachata, learned the basic steps shortly before the recording. During the recording, both participants placed smartphones in their back left pockets and danced to a slow Bachata song. This procedure was replicated for the basic **Salsa** steps, with the exception that the novice struggled to follow the music, resulting in separate recordings for the novice and expert steps without musical accompaniment.

To record the expert performance of **tennis** swings, we attached a smartphone to the upper arm of an amateur tennis player's dominant hand. This player had been practicing the sport for over two decades. The expert then executed ten forehand swings, ten backhand swings, and ten tennis serves. For the novice performance, we recorded the same player executing the swings using their non-dominant hand, with the smartphone attached to the upper arm of the non-dominant hand.

To capture the expert **football** performance, we affixed a smartphone to the lower leg of an amateur player with over two decades of experience in the sport. In contrast, for the novice performance, the same player used their non-dominant lower leg for the recordings. Four penalty kicks were executed for both scenarios. To ensure precise technique execution, a visual representation of a goal post was marked on a wall, serving as a target during the recordings.

The recorded data was saved on .csv files, which stored the X, Y, and Z accelerometer values obtained from the smartphone. In [13], we analyzed the level of noise in the recordings through the following procedure. Initially, we trimmed the .csv files, utilizing data plotting to identify the activity's start and end points and extracting only the relevant data points. For activities like dancing steps, running, and badminton drills, we further trimmed the files to 1000 frames to facilitate an objective comparison of the plots. However, for tennis swings and football kicks, we opted to trim the recordings to 1500 frames, tailored to the specific requirements of these activities. After to trimming the .csv files, we generated plots for each accelerometer axis in [13]. Time was represented on the x-axis, while the accelerometer values were plotted on the y-axis. By observing the irregularities (noise) in the plotted data points, we found it straightforward for a human to discern whether the recordings belonged to a novice or an expert. For visual references, please refer to Figures 1, 2, and 3.

However, formalizing the level of noise of a signal without knowing the expected function is a hard problem. Thus, we explored different options for the analysis. First, we compressed the plotted data using a **. PNG algorithm** to see whether the algorithm could automatically recognize regularities and hence have a higher rate of compression, as shown in [15]. Second, by looking at the plotted data in [13] we hypothesized that the **standard deviation** of the novices' plots would be greater than the one of the experts. Finally, by looking at the plots we saw that especially for rhythmic movements (e.g. running) the graph was smoother, therefore we examined the aggregated values of the **first derivative** of the recordings expecting the aggregated values from novice performances to be greater than the ones of experts.

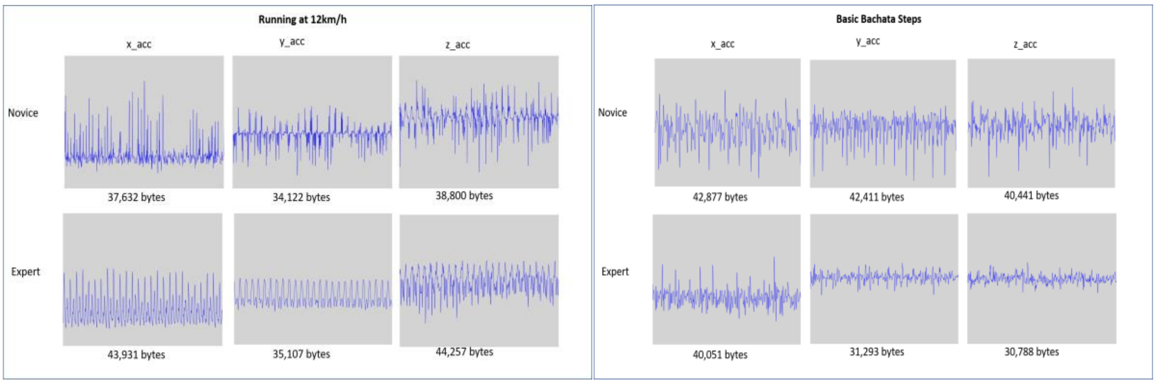

Figure 1. Left Plots of accelerometer data for running at 12km/h. The Y-Axis represents the sensor values and the X-Axis represents the time. Right Bachata Steps.

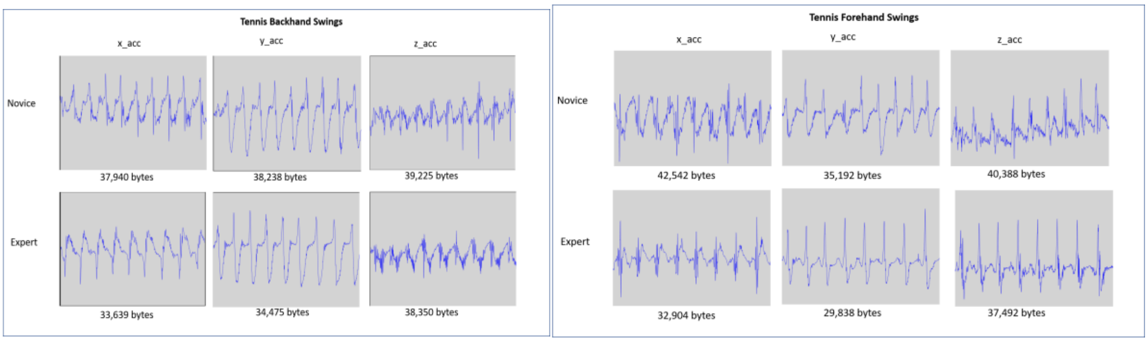

Figure 2. Left: Plots of accelerometer data for Tennis Backhand Swing. The Y-Axis represents the sensor values and the X-Axis represents the time. Right:

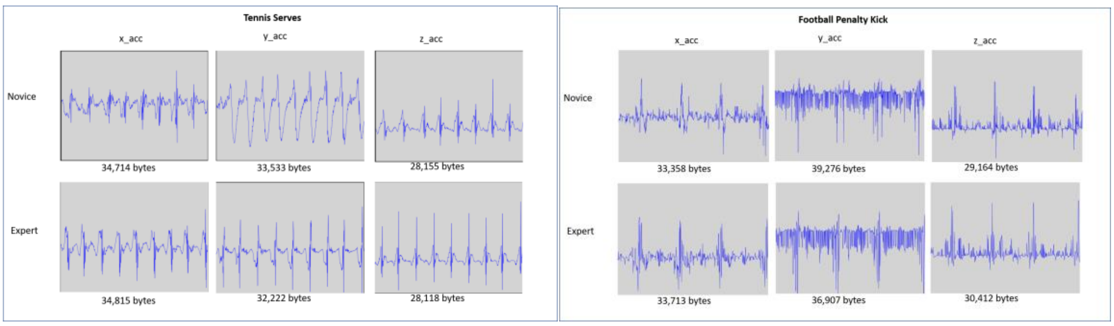

Figure 3. Left: Plots of accelerometer data for Tennis Serve. The Y-Axis represents the sensor values and the X-Axis represents the time. Right: Plots of accelerometer data for Tennis Serve. The Y-Axis represents the sensor values and the X-Axis represents the time.

3. Results

Table 1 shows the results from our three analyses: the file size of the plots, the standard deviation, and the aggregated values of the first derivative.

	PNG file size in KB		Standard Deviation		First Derivative	
	Expert	Novice	Expert	Novice	Expert	Novice
Running	47,0	36,7	10,32	8,91	3,78	5,27
	45,6	33,3	7,38	6,39	1,71	4,45
	47,0	37,8	5,25	4,78	2,21	3,07
Bachata	43,7	41,8	1,60	2,46	0,73	0,86
	43,1	41,6	1,16	2,92	0,51	1,32
	45,8	39,4	0,93	2,18	0,48	1,03
Salsa	43,5	44,5	1,50	2,44	0,66	0,77
	34,4	42,2	1,40	2,11	0,68	0,92
	44,0	33,7	1,12	1,24	0,59	0,63
TennisB	32,8	37,0	3,55	4,19	0,49	0,70
	33,6	37,3	4,68	4,91	0,39	0,45
	37,4	38,3	1,77	2,10	0,60	0,56
TennisF	32,1	36,5	2,39	2,75	0,53	0,61
	29,1	28,3	3,95	2,95	0,39	0,35
	36,6	36,4	2,65	2,40	0,55	0,48
TennisS	33,9	33,9	4,34	3,77	0,85	0,82
	31,4	32,7	4,69	5,43	0,84	0,75
	27,4	27,4	5,75	5,48	1,31	1,05
Football Kicks	36,3	32,5	3,74	3,25	5,34	4,34
	41,1	38,3	6,62	4,77	6,67	5,91
	32,4	28,4	4,87	3,75	8,06	7,83

By examining the size of the files, we found that there is no distinguishing trend differentiating the expert and novice performance. Similarly, when looking at the standard deviation we observe no trend. This can further be explained by looking at the distribution of the recorded values (See Figures 4 to 15).

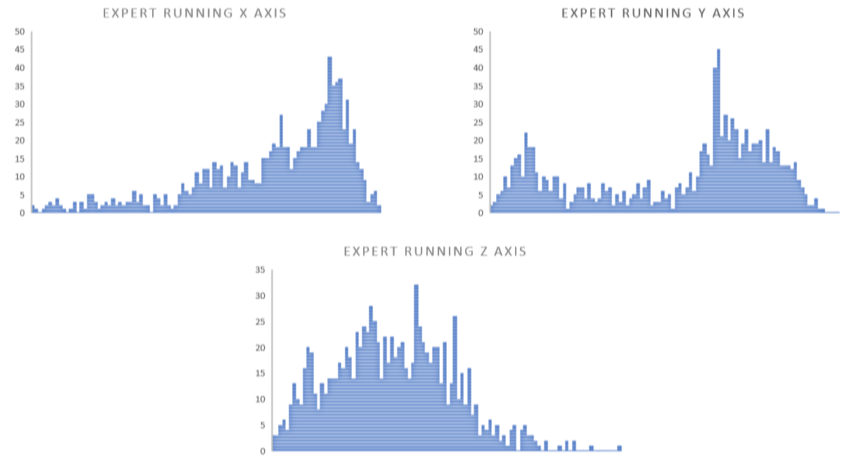

Figure 4. Running expert data.

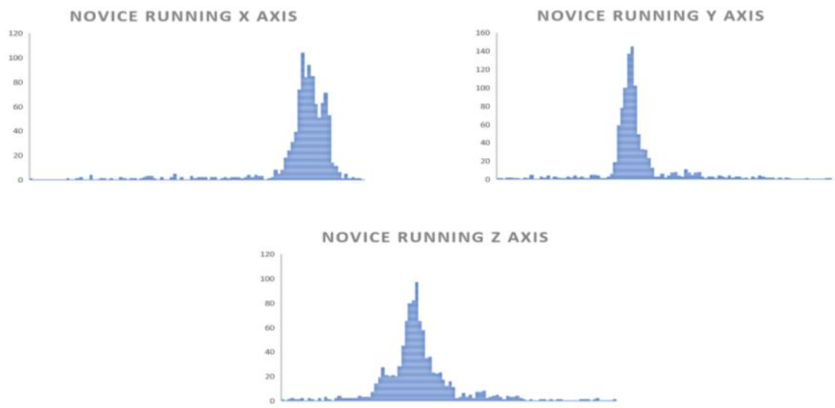

Figure 5. Running novice data.

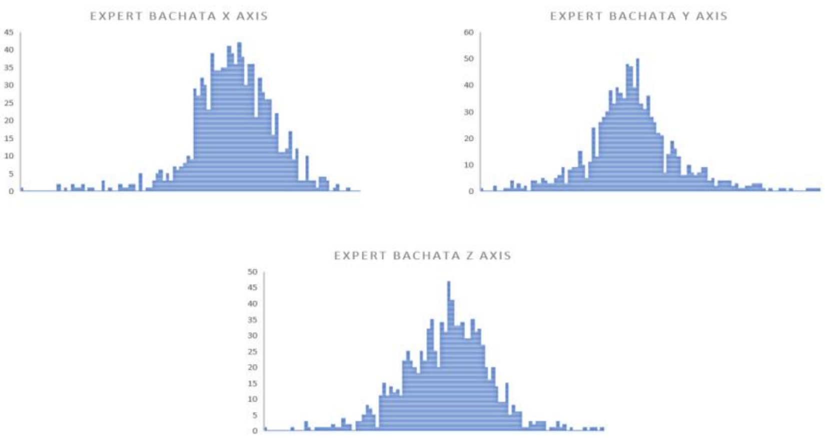

Figure 6: Bachata expert data.

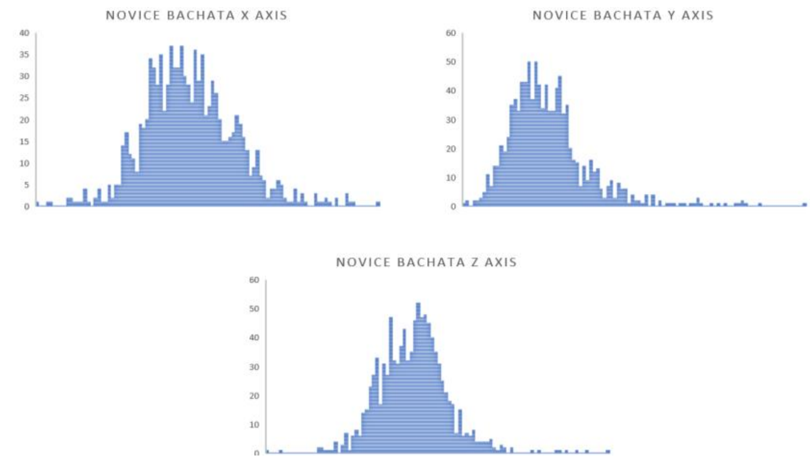

Figure 7. Bachata novice data.

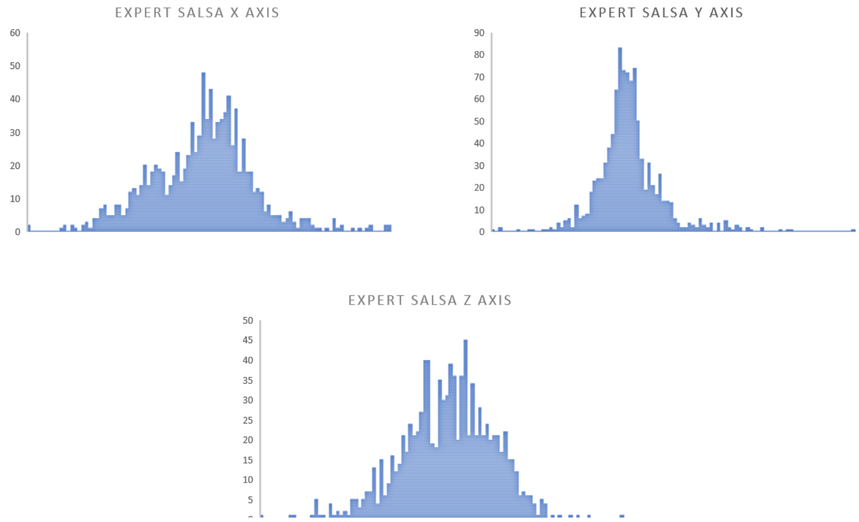

Figure 8. Salsa expert data.

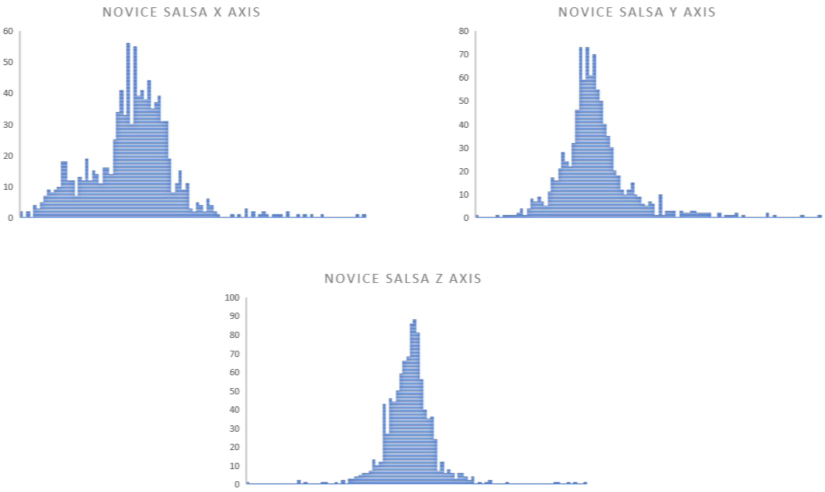

Figure 9. Salsa novice data.

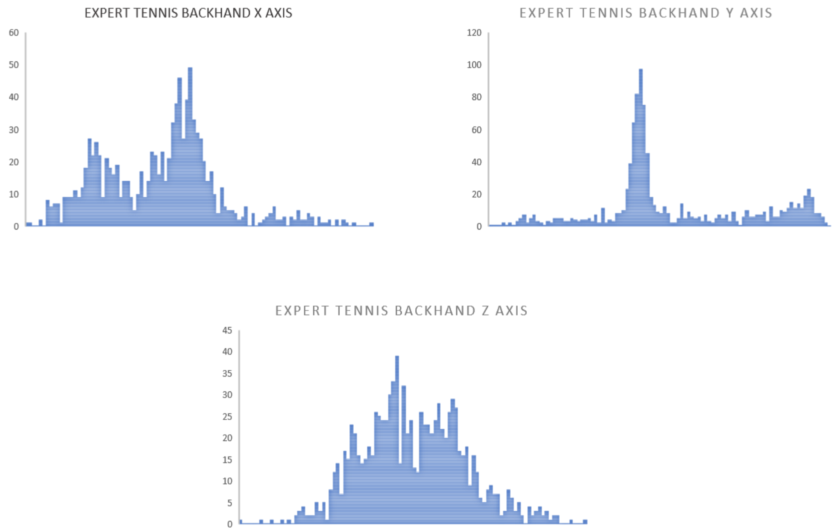

Figure 7. Tennis backhand data.

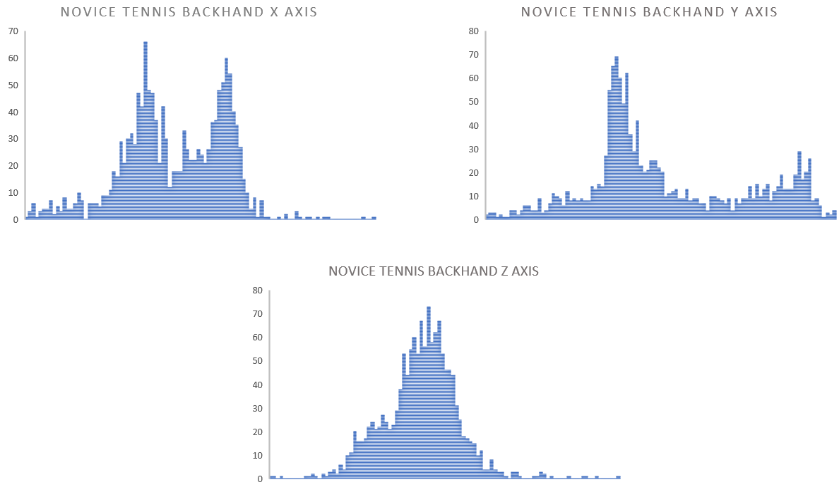

Figure 8. Tennis backhand novice data.

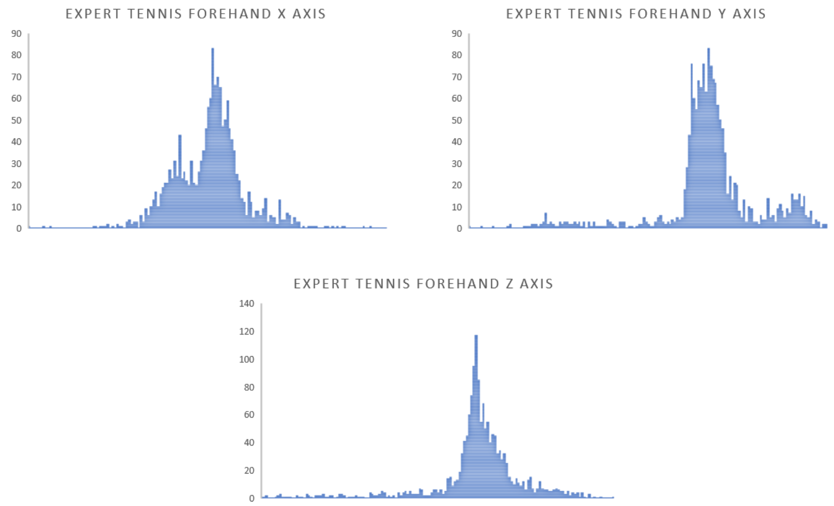

Figure 10. Tennis forehand expert data.

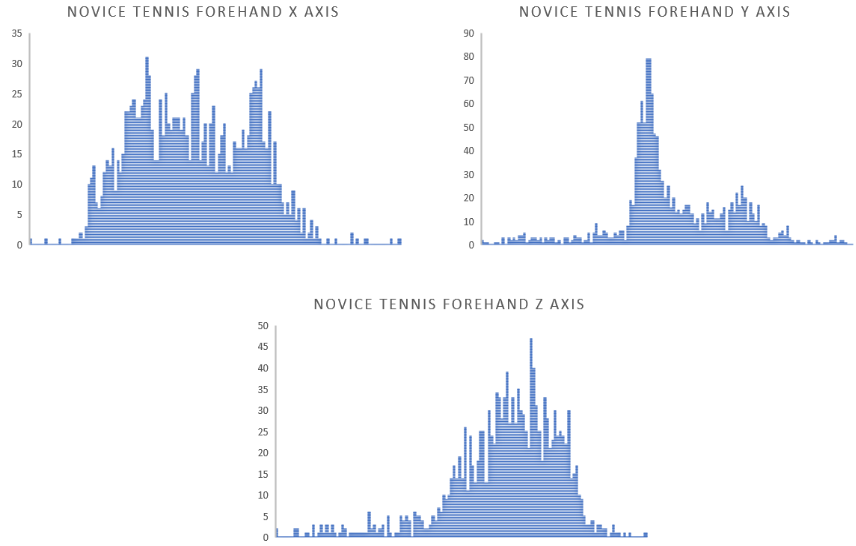

Figure 11. Tennis forehand novice data.

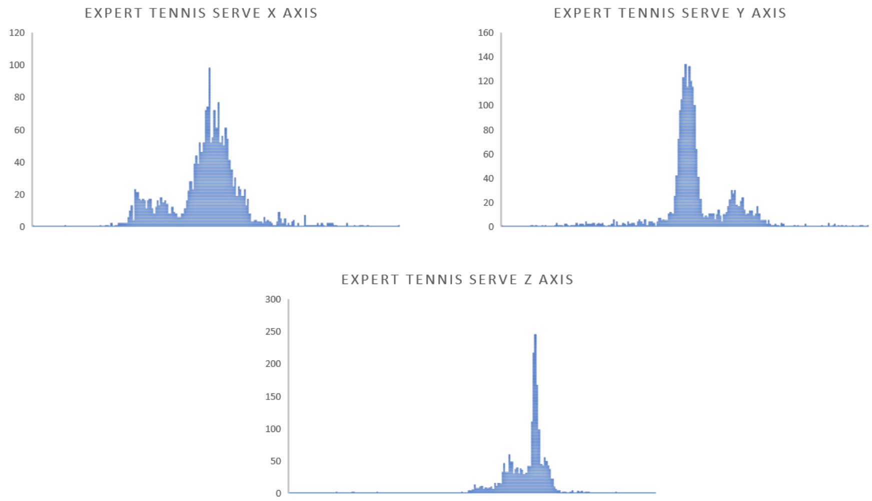

Figure 12. Tennis serves expert data.

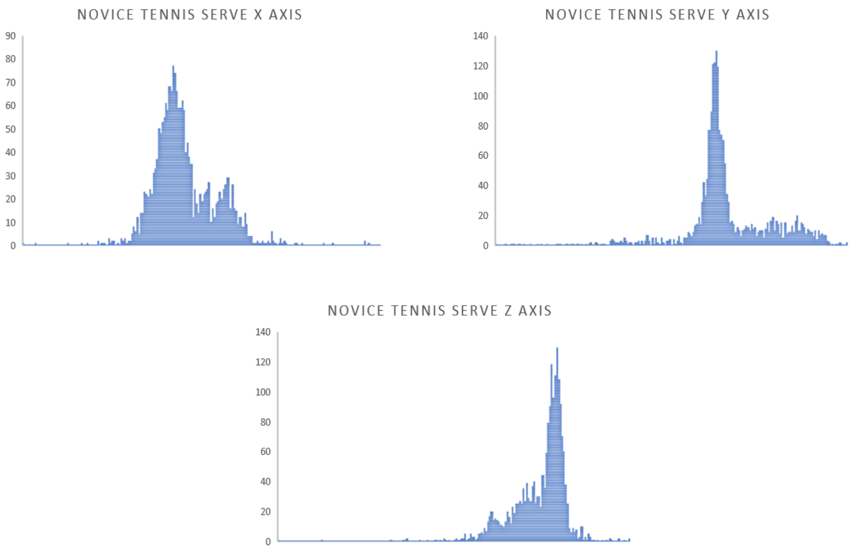

Figure 13. Tennis serves novice data.

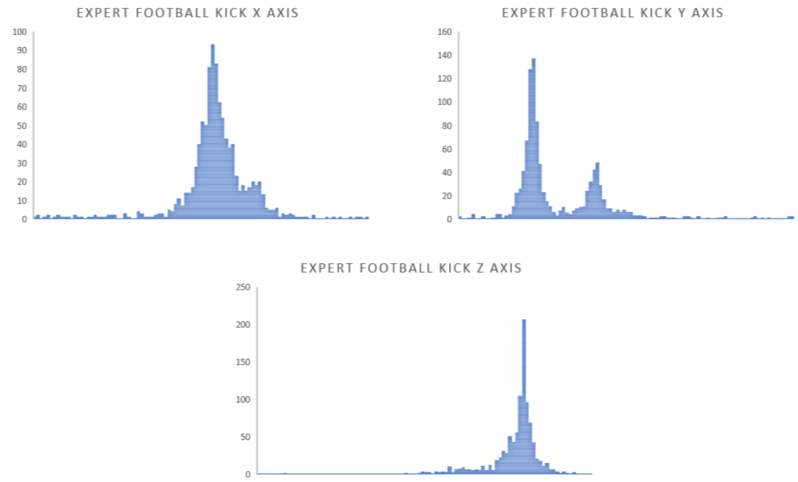

Figure 14. Football kicks expert data.

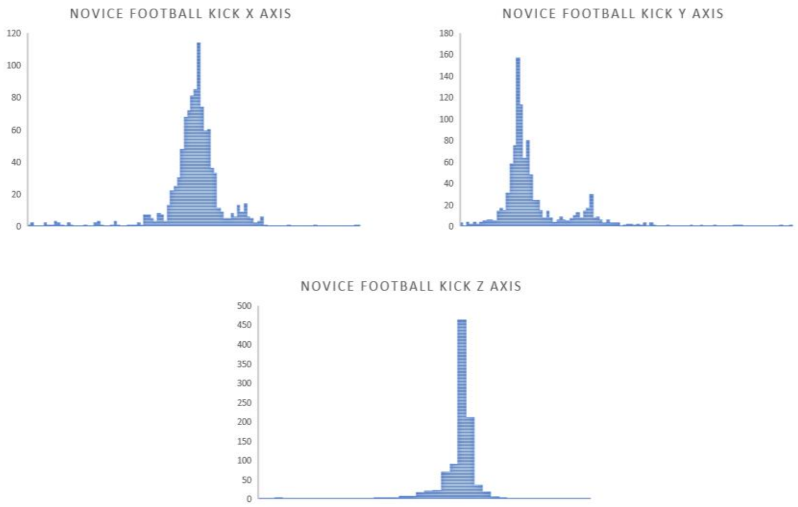

Figure 15. Football kicks novice data.

Finally, by looking at the aggregated value of the first derivative we see that all values from the expert recordings are smaller than the corresponding values for novices. We conducted a T-test to look for the significance of the values. The results from the T-test are $T(14)=-1.11$: $p=0.29$, showing a non-significant trend.

4. Discussion and Conclusion

In this paper, we explored how we can, potentially, automatically compare the performance of experts against novices from accelerometer data. We tried three different techniques that seemed intuitive. From these three techniques, only the comparison of aggregated values of the first derivative seems promising for rhythmic tasks (e.g. dancing, running, playing drums, etc.), however, we need more data to confirm this hypothesis. In the context of rhythmic tasks, these results may hint at smoother movements/regularities in the expert's performance. For the non-rhythmic movements, by looking at the plot data it is possible to see that the expert performance is more uniform. Therefore, it might be possible to distinguish between expert and novice performance by analyzing the variability of the main amplitude of the signal, which can be done by carefully segmenting the recordings and using a Fast Fourier Transformation (FFT) to obtain this amplitude.

In contrast to [14], the .png conversion method showed no trends in hinting at the noise levels and hence distinguishing between the experts' and the novices' performances. Similarly, our hypothesis that the standard deviation of expert data would be lower in comparison to the novices was proven wrong.

For future work, we plan to investigate whether the variance of the amplitude obtained with the FFT can provide some information to distinguish between novice and expert performance for non-rhythmic tasks, and collect more data for rhythmic and non-rhythmic tasks to get statistically significant results. Moreover, we can apply the same methodology but with multiple/sensors other than accelerometers.

References

[1] Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2015). Augmenting the senses: a review on sensor-based learning support. Sensors, 15(2), 4097-4133.

[2] Dikken, O., Limbu, B., Specht, M. (2022). Deliberate Practice of Handwriting: Supervision Under the Ghost of an Expert. In: Hilliger, I., Muñoz-Merino, P.J., De Laet, T., Ortega-Arranz, A., Farrell, T. (eds) Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption. EC-TEL 2022. Lecture Notes in Computer Science, vol 13450. Springer, Cham.

[3] Mat Sanusi, K. A., Mitri, D. D., Limbu, B., & Klemke, R. (2021). Table tennis tutor: Forehand strokes classification based on multimodal data and neural networks. Sensors, 21(9), 3121.

[4] Limbu, B. H., Jarodzka, H., Klemke, R., & Specht, M. (2018). Using sensors and augmented reality to train apprentices using recorded expert performance: A systematic literature review. Educational Research Review, 25, 1-22.

[5] Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets,P.(2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. Learning and Instruction, 25:62–70

[6] Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. Journal of Computer Assisted Learning, 34(4), 338-349.

[7] Buraha, T., Schneider, J., Mitri, D. D., & Schiffner, D. (2021, September). Analysis of the "D'oh!" Moments. Physiological Markers of Performance in Cognitive Switching Tasks. In European Conference on Technology Enhanced Learning (pp. 137-148). Springer, Cham.

[8] Di Mitri, D., Schneider, J., & Drachsler, H. (2022). Keep me in the loop: Real-time feedback with multimodal data. International Journal of Artificial Intelligence in Education, 32(4), 1093-1118.

[9] Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018, March). The Big Five: Addressing Recurrent Multimodal Learning Data Challenges. In CrossMMLA@ LAK.

[10] Di Mitri, D., Schneider, J., Specht, M. M., & Drachsler, H. J. (2019, August). Multimodal Pipeline: A generic approach for handling multimodal data for supporting learning. In First workshop on AI-based Multimodal Analytics for Understanding Human Learning in Real-world Educational Contexts.

[11] Schneider, J., Mitri, D. D., Limbu, B., & Drachsler, H. (2018, September). Multimodal learning hub: A tool for capturing customizable multimodal learning experiences. In European conference on technology enhanced learning (pp. 45-58). Springer, Cham.

[12] Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H. (2019, March). Read between the lines: An annotation tool for multimodal data for learning. In Proceedings of the 9th international conference on learning analytics & knowledge (pp. 51-60).

[13] Schneider, J., Mat Sanusi, K. A., Limbu, B. H., Schmitz, M., & Schiffner, D. (2023). Novices Make More Noise! The D&K Effect 2.0? In CrossMMLA 2023: Leveraging Multimodal Data for Generating Meaningful Feedback (Vol. 3439, pp. 43-48). (CEUR Workshop proceedings; Vol. 3439). CEUR Workshop.

[14] Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. Psychological Review, 4(1), 27.
[15] Zbili, M., & Rama, S. (2021). A quick and easy way to estimate entropy and mutual information for neuroscience. Frontiers in Neuroinformatics, 15, 596443.