M-MS: a Multi-Modal Synchrony Dataset to Explore Dyadic Interaction in ASD

Gabriele Calabrò[1,2], Andrea Bizzego[1], Stefano Cainelli[1], Cesare Furlanello[2], and Paola Venuti[1](✉)

[1] Dept. of Psychology and Cognitive Science - University of Trento, Rovereto, Italy
paola.venuti@unitn.it
[2] Fondazione Bruno Kessler, Trento, Italy

Abstract. Empathy, reciprocity and turn-taking are critical therapeutic targets in conditions of social impairment such as Autism Spectrum Disorder (ASD). These aspects are related to each other, converging into the construct of synchrony, which includes emotional, behavioural and, possibly, physiological components. Therefore, being able to quantify the synchrony could impact the way therapists adapt and maximise the efficacy of the interventions. However, current methods are based on the observational coding of behavior which is time-consuming and usually only performed after the interaction is over. In this study we propose to apply Artificial Intelligence (AI) methods on physiological data in order to obtain a real-time and objective quantification of synchrony. In particular, we introduce the Multi-Modal Synchrony dataset (M-MS), which includes 3 sources of information - electrocardiographic signals, video recordings and behavioral coding - to support the study of synchrony in ASD. As a first AI application, we are currently developing an unsupervised model to extract a multivariate embedding of the physiological data. The multivariate embedding has to be compared with the behavioral synchrony label to create a map of physiological and behavioural synchrony. The application of AI in the treatment of ASD may become a new asset for the clinical practice, especially if the possibility of providing real-time feedback to the therapist is exploited.

Keywords: Autism · Dyadic Interaction · Synchrony · Psychophysiology · Deep Learning

1 Introduction

Autism spectrum disorder (ASD) is an increasingly prevalent neurodevelopmental disorder with severe impacts on quality of life. The symptoms have a precocious onset and affect multiple areas, social impairment in particular being a core feature [1].

A pivotal form of social interaction is the dyadic one, which represents both a developmental tool available to the child [2, 3] and a key therapeutic channel available for ASD therapy. This one-on-one interaction has been viewed from different perspectives, with the construct of synchrony being one of the most relevant [4, 5]. Described in terms of mutuality, reciprocity, rhythmicity, harmonious interaction, turn-taking and shared affect [6], synchrony reflects ASD deficits and has been used to study development trajectories in children with ASD [7]. Being able to quantify and promote synchrony is therefore important to track and improve the efficacy of the intervention and guide the activity of the therapist [8]. However, observational procedures are time consuming, have a certain degree of subjectivity and often can be performed only after the intervention is concluded, thus limiting the usability of such information.

Leveraging on Artificial Intelligence (AI) methods to achieve an objective synchrony evaluation process and provide this information to the therapist in real time can offer a new asset to the clinical practice. There is already promising evidence of successful applications of AI methods to study ASD such as the detection of the occurrence of stereotypical motor movements using wearable accelerometers [9, 10] and the identification of meltdown-related behaviors from video recordings [11]. An AI approach to quantify the synchrony during social interaction could solve the problems related to subjectivity and temporal requirements, enabling an objective real-time inference on the current state of the therapy.

However, accuracy and reliability of AI models rest on appropriate datasets which are meager in our scenario due to the requirement of being based on clinically valid information.

We therefore composed the Multi-Modal Synchrony dataset (M-MS), a new AI-ready dataset including physiological and behavioral data acquired in a naturalistic clinical setting during actual therapeutic activities involving children with ASD. Physiological data consist of electrocardiographic (ECG) signals, which are linked to the engagement in socio-cognitive tasks [12]. Behavioral data are acquired in the form of video recordings of the activities and synchrony labels based on observational coding of behavior performed by experts.

As a first AI application on this dataset, we selected an unsupervised deep learning method based on an auto-encoder architecture. This category of AI methods already proved to be a viable approach to automatically extract from physiological data a representation which is useful to detect the occurrence of events of interest [13]. For our purposes, this translates into the first step in exploring synchrony during social interaction in ASD using an automatically learned representation of physiological data.

2 Material and Methods

The study was based on actual therapeutic activities carried out between the years 2017 and 2018 in Rovereto and Coredo (Trento, Italy). During all the activities, we simultaneously acquired the ECG signals both from the child and the therapist. The video recording of the therapy sessions became feasible for the activities carried out during the year 2018 thanks to technical improvements in the temporal alignment of the multi-modal data streams.

2.1 Participants

The study involved 19 male and 2 female children meeting the criteria for ASD diagnosis according to the diagnostic and statistical manual of mental disorders (DSM-5)[1] and aged 2-12 years old (average: 5.5 y, SD: 2.5 y). Although debated [14], the imbalance in the sample is in line with the long-standing different prevalence of ASD according to sex [15]. Each child interacted with the same male therapist, which had extensive experience in ASD treatment. The parents - or legal guardians - participated in a briefing about the experimental protocol and provided their informed consent to allow the participation of their child in the study.

2.2 Experimental Setting

We selected activities of music therapy due to its ability to effectively target synchrony aspects [16] regardless potential language impairments often present in children with ASD [1]. We gave priority to the clinical validity of the dataset aiming at the observation of actual therapy dynamics, thus we chose to not adopt a structured protocol. Therefore, the activities were performed without directives to the therapist in order to preserve the natural course and the therapeutic value of the intervention. In the future, this unrestricted approach allows its seamless extension to different therapeutic scenarios and to more ecological settings such as the child's home.

The child and the therapist interacted in a dedicated room containing music instruments and no common toys. Upon entering the room, the therapist allowed the child to get comfortable and acclimated to the setting. If the therapist considered the situation at risk of stressing the child, then no data were acquired. Otherwise, he proceeded to the placement of the sensors and carried out the activities while an operator managed the data acquisition platform.

2.3 Data Acquisition

During the activities, ECG signals were acquired both from the therapist and the child therefore providing 2 alignable time series. A critical factor in our experimental design was the adoption of sensorized garments specifically designed to reduce the risk of sensory overload which is a critical factor in ASD [17]. The prototypes were created in collaboration with a specialized company (ComfTech, Monza, Italy). The sensorized garments (see Figure 1) consisted of a vest integrated with 2 textile electrodes with a conductive surface area in direct contact with the skin and a textile connector transmitting the signal to an acquisition device operating at 128 Hz.

During the activities carried out in 2018, the video stream of the therapy was acquired using a camera operating at 25 frame per second with a resolution of 1440x1080 pixels. The alignment of the ECG signals to the video stream was manually achieved by creating a marker event on the acquisition device an a corresponding visual marker on the video stream.

Fig. 1. Example of a prototype sensorized garment designed for children with ASD. The two metal buttons present on the white jersey allow the connection of the textile electrodes with the sensing unit (not shown).

The videos were manually annotated by 2 trained experts applying a behavioural observation code appositely created to detect events relevant to synchrony. The range of annotated events spans from the presence of intentionality in the child's actions up to the full engagement in the interaction, allowing the quantification of the synchrony from the behavioural perspective.

In addition to behavioral and physiological data, we recorded the information about sex, age, DSM-5 diagnosis and severity of each participant. The data have been anonymized by substituting the names of the participants with alphanumeric codes and removing all the references to absolute timestamps.

2.4 Data Processing

To prepare the dataset for use with AI approaches we segmented the data into portions by windowing (length: 10 s; overlap: 9 s) and associated a label of synchrony derived from the behavioural coding of video recordings whenever available. The ECG signals of each portion were processed in order to automatically identify and remove portions presenting either movement artifacts or noise. The data processing procedure is presented in Figure 2.

The ECG signal processing pipeline was implemented in Python and based on the *pyphysio* library [18].

The pipeline is composed of the following steps:

1. Pre-processing: the ECG signals were resampled by interpolation at 2048 Hz, then filtered with a low-pass infinite impulse response filter (pass-band: <50 Hz, stop-band: >55 Hz) to remove high-frequency noise and with a high-pass filter (pass-band: >0.5 Hz, stop-band: <0.05 Hz) to remove low-frequency drifts

2. Beat detection: the R peaks in the ECG signals were detected to compute the Inter Beat Intervals (IBI); detection errors were automatically corrected and the IBI series were resampled by interpolation at 4 Hz
3. Segmentation: ECG and IBI signals were segmented into overlapping consecutive portions (length: 10 s, overlap: 9 s)

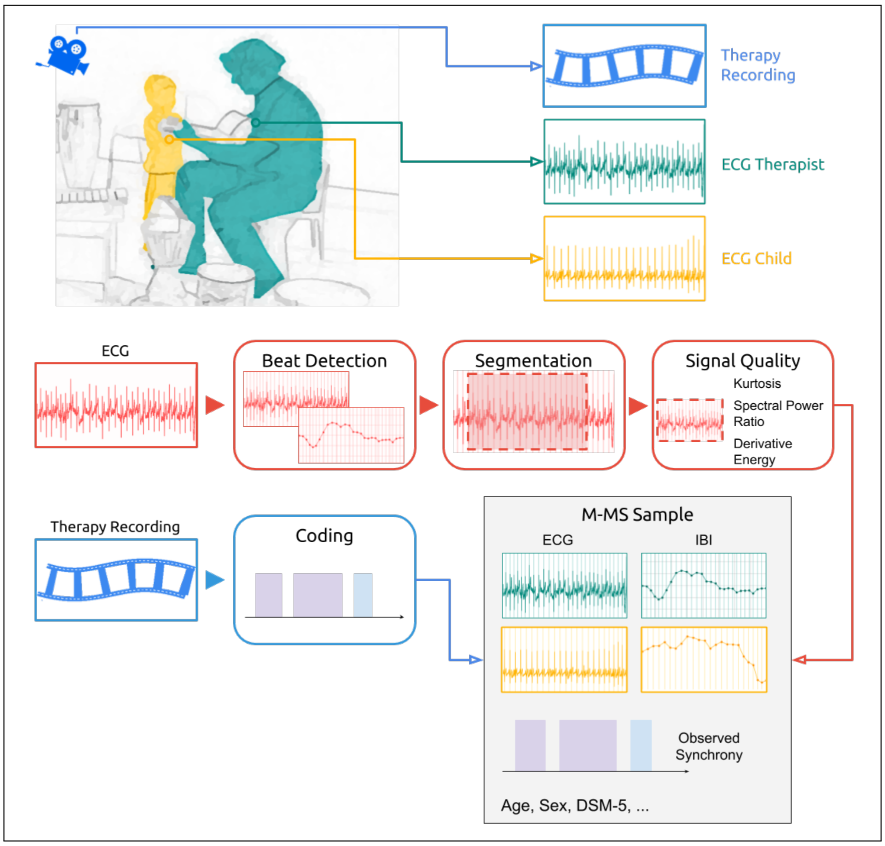

Fig. 2. Overview of the experimental setup (top) and data processing pipeline (bottom). Each sample of the M-MS dataset is composed of: the ECG and IBI segments from both the therapist and the child, the annotation of the observed synchrony and the session metadata.

The quality of the ECG data in each portion was assessed using 3 signal quality indicators (SQI) [19, 20]: kurtosis (K), spectral power ratio (SPR) and the mean squared of the absolute values of the signal derivative (DE) were computed on the ECG signals of both child and therapist. If at least one of the signals failed to meet the SQI cutoff criteria ($3 \leq K \leq 100$, $0.4 \leq SPR \leq 0.8$, $3 \leq DE \leq 12$) the portion was discarded. Each accepted portion provided the ECG and IBI data of both the child and the therapist for a total of 4 time aligned signals of 10 seconds duration each. These 4 signals were saved in a new data point folder with an additional reference file containing relevant information such as sampling frequency, starting and ending time and, when available, the observational coding of behavior. From these data we extracted a binary (good/poor) behavioral synchrony score used as label for the data point.

Of the 21 child-therapist dyads participating in the study, 6 were discarded due to the lack of data points meeting the SQI cutoff criteria.

2.5 Model

The implementation of an AI model with close to the expert performance on this dataset would be a step forward in supporting the daily practice of therapists and caregivers and improving the clinical training.

As a first study on this dataset, we developed an unsupervised AI model based on an auto-encoder architecture, which embeds the ECG signals of the dyad into a unique multi-variate representation. Unsupervised deep learning methods have been successfully used to create representations of physiological data [21]. In particular, convolutional auto-encoder based architectures showed good performance in structuring a representation useful to detect the occurrence of events of interest [22].

Our pilot architecture (see Figure 3) consists of a convolutional auto-encoder taking as input the 2 IBI signals present in each data point and creating an embedding of 5 features that is then expanded to recover the original signals patterns. The values of the input IBI signals were normalized portionwise via rescaling between 0 and 1 based on maximum and minimum values. The architecture structure was implemented with *pyTorch*[23].

With this architecture we expect the embedding to include the relevant information about the physiological dynamics and mutual influence of the two input signals.

2.6 Model Training

The dataset was split into train and test partitions using a procedure akin to the leave-one-out cross-validation method. At each rotation, all data points from 1 child-therapist dyad were kept out and used as test partition to evaluate the performance on unseen data. For this reason, we call this approach leave-one-dyad-out cross-validation.

We performed the training of the network within 50 epochs using an optimizer based on the Adam algorithm [24] (learning rate: 0.0005) and a batch size of 16 samples. The training of the network was performed applying the mean squared errors (MSE) as loss function.

The performance of the network was then evaluated using the MSE on the test partition and comparing it to the MSE on the train partition in order to check for overfitting.

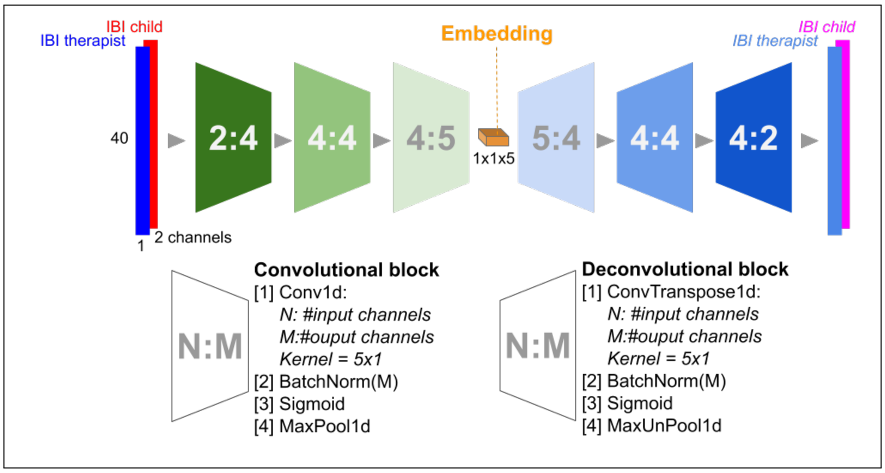

Fig. 3. Structure of the convolutional auto-encoder used to reconstruct the IBI segments of the therapist and the child. The structure is composed of two parts: the encoder (green convolutional blocks) and the decoder (blue de-convolutional blocks).

3 Results

Multi-modal data from a total of 41 sessions (578 minutes) were collected, with additional 22 sessions (396 minutes) with only the ECG signals. Two trained experts annotated the video recordings allowing the quantification of synchrony from the behavioural perspective. In addition to the physiological and behavioural data, we added the information about sex, age, DSM-5 diagnosis and severity of each participant. The names of the participants were substituted with random alphanumeric codes and all the references to absolute timestamps were removed in order to anonymize the data.

We processed a total of 53306 portions, resulting in a dataset composed of 37127 data points of high quality physiological data, 20134 of which also provide the video data and the behavioral synchrony label (see Figure 4).

The name of each data point incorporates the sequential numbering of the relative portion of raw data and the identifying number of the acquisition session. This makes the dataset modular: it is possible to use the information contained in the data point name to concatenate adjacent portions in order to reconstruct signals lasting more than 10 s, thus enabling the extraction of typical handcrafted time domain Heart Rate Variability (HRV) features.

Our convolutional auto-encoder architecture showed a reconstruction error of 5.4%. In particular, we evaluated the network performance using the leave-one-dyad-out cross-validation, obtaining a test MSE of 0.0535 and a train MSE of 0.0543.

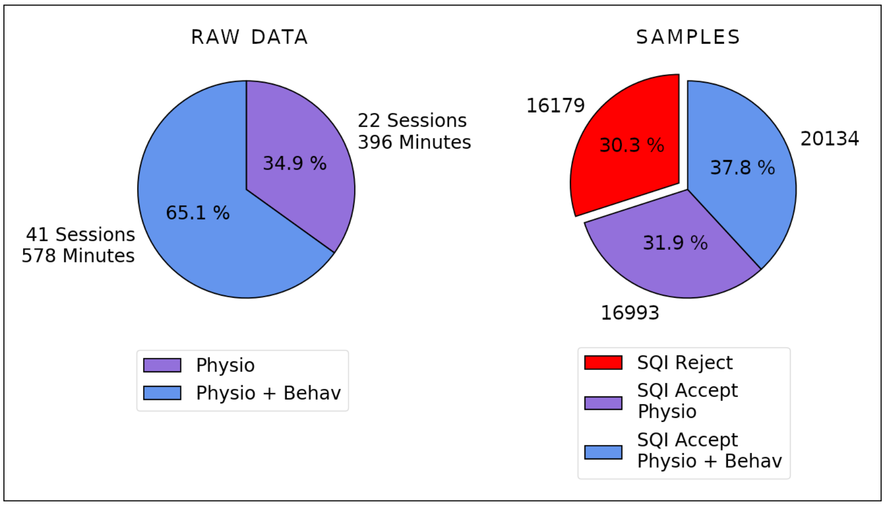

Fig. 4. Summary of the raw data acquired (left) and overview of the quality of the portions used to create the samples in the M-MS dataset (right).

4 Discussion

In this work we proposed the M-MS dataset, a new multi-modal dataset aiming at the exploration of physiological and behavioral synchrony during dyadic interaction between a child with ASD and a therapist. The information included are the ECG and IBI data of both the child and the therapist, the video stream of the interaction between them and a behavioral synchrony label based on expert judgement.

Due to the ecological approach in the selection of the experimental setting, the sample composition reflects the actual adhesions to the therapy programs and is in line with the gender imbalance in ASD prevalence present in literature. Nevertheless, the gender imbalance in the sample composition represents a limitation of the dataset which might be inadequate for gender-related studies.

The dataset creation was based on a sliding window procedure with 90% overlap. This form of data augmentation - which is a common practice in the AI field [25] - combined with the naming procedure of the data points makes the dataset modular: the researcher can manage the amount of overlapping information and reconstruct signals of the desired length, selectively enabling the extraction of typical handcrafted features such as HRV indices [26].

The M-MS dataset can be used to train supervised and unsupervised AI models in order to develop an objective and streamlined evaluation of synchrony which would enable an automated annotation of social interaction in ASD thus paving the way to real-time feedback to the therapist. From the clinical perspective, this dataset can additionally be used to study the association between the physiological and behavioural patterns and phenotypes in ASD.

As a first application of AI methods to the dataset, we implemented a convolutional auto-encoder representing the child-therapist IBI signals in an embedding of 5 features with a reconstruction error of 5.4%. This preliminary evidence entails further study aiming at an unsupervised model able to extract latent variables from physiological data which would match the behavioral synchrony evaluated by the experts.

5 Data Availability

The M-MS dataset is not publicly available due to the personal information included, but is available from the corresponding author upon reasonable request.

References

1. A. P. Association, et al., Diagnostic and statistical manual of mental disorders (DSM-5®), American Psychiatric Pub, 2013.
2. D. Cohen, The developmental being. modeling a probabilistic approach to child development and psychopathology, Neuropsychiatrie de l'Enfance et de l'Adolescence 5 (60) (2012) S25–S26.
3. A. Azhari, W. Leck, G. Gabrieli, A. Bizzego, P. Rigo, P. Setoh, M. H. Bornstein, G. Esposito, Parenting stress undermines mother-child brain-to-brain synchrony: A hyperscanning study, Scientific reports 9 (1) (2019) 1–9.
4. A. W. Harrist, R. M. Waugh, Dyadic synchrony: Its structure and function in children's development, Developmental Review 22 (4) (2002) 555–592.
5. A. Bizzego, A. Azhari, N. Campostrini, A. Truzzi, L. Y. Ng, G. Gabrieli, M. H. Bornstein, P. Setoh, G. Esposito, Strangers, friends, and lovers show different physiological synchrony in different emotional states, Behavioral Sciences 10 (1) (2020) 11.
6. C. Leclère, S. Viaux, M. Avril, C. Achard, M. Chetouani, S. Missonnier, D. Cohen, Why synchrony matters during mother-child interactions: A systematic review, PloS ONE 9 (12) (2014) e113571.
7. M. Siller, M. Sigman, The behaviors of parents of children with autism predict the subsequent development of their children's communication, Journal of Autism and Developmental Disorders 32 (2) (2002) 77–89.
8. P. Venuti, A. Bentenuto, S. Cainelli, I. Landi, F. Suvini, R. Tancredi, R. Igliozzi, F. Muratori, A joint behavioral and emotive analysis of synchrony in music therapy of children with autism spectrum disorders, Health Psychology Report 5 (2) (2017) 162–172.
9. N. M. Rad, S. M. Kia, C. Zarbo, T. van Laarhoven, G. Jurman, P. Venuti, E. Marchiori, C. Furlanello, Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders, Signal Processing 144 (2018) 180–191.
10. N. M. Rad, A. Bizzego, S. M. Kia, G. Jurman, P. Venuti, C. Furlanello, Convolutional neural network for stereotypical motor movement detection in autism, arXiv preprint arXiv:1511.01865.
11. V. S. P. Patnam, F. T. George, K. George, A. Verma, Deep learning based recognition of meltdown in autistic kids, in: 2017 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2017, pp. 391–396.

12. S. D. Palma, A. Tonacci, A. Narzisi, C. Domenici, G. Pioggia, F. Muratori, L. Billeci, Monitoring of autonomic response to sociocognitive tasks during treatment in children with autism spectrum disorders by wearable technologies: A feasibility study, Computers in Biology and Medicine 85 (2017) 143–152.
13. A. Supratak, L. Li, Y. Guo, Feature extraction with stacked autoencoders for epileptic seizure detection, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 4184–4187.
14. J. N. Constantino, Taking stock of critical clues to understanding sex differences in the prevalence and recurrence of autism, Autism 21 (6) (2017) 769–771.
15. D. M. Werling, D. H. Geschwind, Sex differences in autism spectrum disorders, Current Opinion in Neurology 26 (2) (2013) 146.
16. J. Kim, T. Wigram, C. Gold, Emotional, motivational and interpersonal responsiveness of children with autism in improvisational music therapy, Autism 13 (4) (2009) 389–409.
17. L. D. Wiggins, D. L. Robins, R. Bakeman, L. B. Adamson, Breif report: Sensory abnormalities as distinguishing symptoms of autism spectrum disorders in young children, Journal of Autism and Developmental Disorders 39 (7) (2009) 1087–1091.
18. A. Bizzego, A. Battisti, G. Gabrieli, G. Esposito, C. Furlanello, pyphysio: A physiological signal processing library for data science approaches in physiology, SoftwareX 10 (2019) 100287.
19. J. Behar, J. Oster, Q. Li, G. D. Clifford, ECG signal quality during arrhythmia and its application to false alarm reduction, IEEE Transactions on Biomedical Engineering 60 (6) (2013) 1660–1666.
20. Q. Li, R. G. Mark, G. D. Clifford, Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter, Physiological Measurement 29 (1) (2007) 15.
21. Y. Yuan, G. Xun, K. Jia, A. Zhang, A multi-view deep learning method for epileptic seizure detection using short-time fourier transform, in: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, ACM, 2017, pp. 213–222.
22. Y. Yuan, G. Xun, K. Jia, A. Zhang, A multi-view deep learning framework for EEG seizure detection, IEEE Journal of Biomedical and Health Informatics 23 (1) (2019) 83–94.
23. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., PyTorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, 2019, pp. 8024–8035.
24. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
25. L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, arXiv preprint arXiv:1712.04621.
26. C. A. Rickards, K. L. Ryan, V. A. Convertino, Characterization of common measures of heart period variability in healthy human subjects: implications for patient monitoring, Journal of Clinical Monitoring and Computing 24 (1) (2010) 61–70.