Eurographics Conference on Visualization (EuroVis) 2016
K.-L. Ma, G. Santucci, and J. van Wijk
(Guest Editors)

Volume 35 (2016), Number 3

PhysioEx: Visual Analysis of Physiological Event Streams

Rishikesan Kamaleswaran[1,2], Christopher Collins[1], Andrew James[2,3], and Carolyn McGregor[1]

[1]University of Ontario Institute of Technology, Canada
[2]The Hospital for Sick Children, Canada
[3]Department of Paediatrics, The University of Toronto, Canada

Abstract

In this work, we introduce a novel visualization technique, the Temporal Intensity Map, which visually integrates data values over time to reveal the frequency, duration, and timing of significant features in streaming data. We combine the Temporal Intensity Map with several coordinated visualizations of detected events in data streams to create PhysioEx, a visual dashboard for multiple heterogeneous data streams. We have applied PhysioEx in a design study in the field of neonatal medicine, to support clinical researchers exploring physiologic data streams. We evaluated our method through consultations with domain experts. Results show that our tool provides deep insight capabilities, supports hypothesis generation, and can be well integrated into the workflow of clinical researchers.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces J.3 [Life and Medical Sciences]: Medical information systems—

1. Introduction

Identifying patterns in multidimensional streaming data is a challenging and common problem across many domains, from intelligence analysis to market research. It is also true in the neonatal intensive care unit (NICU), where clinical researchers need to identify complex patterns in a single patient that extend across multiple high frequency physiologic data streams. N-of-1, also known as the subject-of-one methodology is used by researchers to perform in-depth, multi-faceted analysis of real-world observations [LPD*11]. This method of analysis often results in the early generation of a hypotheses that can be more formally tested. However prior to that formal test, clinical researchers must perform an in-depth study of cases, to extract salient features that support an early hypothesis. In this paper we introduce a visual representation for in-depth analysis of multi-dimensional temporal data streams.

Physiologic streams represent a subset of complex data streams, because they change frequently over time as infants grow and mature, and new normal ranges and values are established week-by-week. Clinical researchers elicit knowledge from those multidimensional physiologic streams by isolating features and analysing behaviours that may predict the onset of clinical conditions. Conducting an analysis is a complex undertaking and currently requires significant manual siphoning of raw physiologic traces and other relevant clinical information. To address some of these challenges, we have developed a novel visualization technique, the Temporal Intensity Map, which reveals critical information about the frequency, duration and trajectory of streaming events generated by real-time event stream algorithms. A novel event-stream algorithm was developed by Thommandram and colleagues [TPE*13] that produces event features and classifications in real-time. The visualizations utilize these output to highlight salient temporal features that may assist the user in generating hypotheses about physiologic behaviour. We also contribute a unique representation of the bubble chart, named the Sequence Graph for identifying high level periodic patterns. Finally, we present methods of highlighting three salient temporal properties called the temporal tri-event parameters that include frequency, duration, and trajectory.

The Temporal Intensity Map contains three features, first, a non-linear binning method, represented on the vertical axis, that is based on both density estimation and logarithmic clustering to discretise the non-parametric distribution. Second, a time axis to isolate regions of temporal interest. Finally, we employ alpha blending and hue to control for severity and frequency of an event. The final feature rapidly conveys information about trending effects, such as the total percentage away from baseline, or duration of a critical event. We combine the Temporal Intensity Map with several other coordinated visualizations to create PhysioEx, illustrated in Figure 1, as a collection of multi-dimensional temporal representations that supports interactive coordinated brushing, zooming, filtering, and selecting high-level events to expose raw data. Currently there is no platform to support researchers studying behaviours in neonatal spells leading up to critical clinical conditions, such as infection.

In a preliminary study of domain experts using PhysioEx, participants detected correlations between low-level event features and high-level event classifications, identified salient features in the physiologic data streams that illustrate the infant's cardiorespira-

tory health, and deliberated over the presence of infection by carefully studying physiological trends. These findings are valuable in the face of a current lack of tools available to perform deep insight analysis of physiological data. The research contributions of this paper are as follows:

- The Temporal Intensity Map (TIM) visualization technique for frequency, duration and trajectory of events.
- The PhysioEx dashboard of coordinated views including TIMs, sequence graph, linear graph, and streams graph.
- A case study of PhysioEx with NICU clinical researchers.

In the remainder of this paper we will provide background of the specific problem domain, followed by related work, design requirements, design of PhysioEx, preliminary user study, discussions and ending with conclusion.

2. Problem Characterization

About 10% of the world's babies are born premature [BCO*12]. In the developed world, premature babies are usually admitted to the NICU. Babies within the NICU have continuous monitoring of their heart rate, breathing, and SpO_2 levels to detect any abnormal shifts. Neonatal sepsis, a form of nosocomial infection, is a life threatening condition that is difficult to detect and for which early detection significantly improves mortality [Fai13]. Apnoea is condition that is defined as a pause in breathing for 20 seconds or more [MMC86]. The term *neonatal spells* is commonly used in NICUs for cardiorespiratory events that may include pauses in breathing, fall in heart rate, or fall in blood oxygen saturation [TPE*13]. An increase in frequency of spells may be associated with neonatal sepsis. A research study by Moorman et al. [MDF*11] reported a potential association between reduced heart rate variability and increased bradycardia in the hours prior to the clinical suspicion of neonatal sepsis. Other studies have also linked the presence of sepsis with heart rate characteristics, especially reduced heart rate variability and bradycardia [FMLD10, GOB*03].

Premature infants experience a reduction in the amount of red blood cells shortly after birth, and frequent blood draws only exasperates this condition resulting in potentially severe clinical symptoms for the newborn [Ket12]. This condition is more pronounced in smaller and very premature infants. Therefore clinicians seek to minimize the number of blood draws taken from the infant for laboratory tests unless it is required.

We present PhysioEx as a tool enabling the end-user to explore neonatal spells event classifications produced by the real-time data stream algorithm around the time of suspicion of neonatal sepsis. By exposing novel neonatal spells event classification information, juxtaposed with the relatively aligned time of suspicion of neonatal sepsis, we provide clinical researchers with an expressive tool to support their analysis and hypothesis generation.

3. Related Work

A number of prior works have produced novel techniques for representing temporal big data, relying on techniques such as visualizing progressive analytics [GZA06, FMK12], hierarchical clustering [EF10], alpha blending [KBK11], and applying animations to compact visual objects [LJH13]. Over-plotting effects of time series data are a common problem and hence novel methods have been developed to reveal patterns [AMM*08]. GScope [TMK03] uses heatmaps to display biological microarray data, a domain which sees frequent use of this visualization method. Those heatmaps were generated using a hierarchical clustering method that highlights up-regulated or down-regulated genes. Temporal streams have also been visualized through linear or metaphoric representations. For instance, CareCruiser allows experts to observe changes in physiologic data following an intervention using visual highlights and interactive brushing of line graphs [GAK*11]. Meanwhile, Huron et al. [HVF13], use a sedimentation metaphor to visualize dynamic data streams. However, neither of these representations support all three of the temporal tri-event parameters.

The frequency tri-event parameter of streaming data has been investigated in prior work to highlight dense and active regions using kernel density estimation (KDE). The use of that non-parametric density estimation is a popular method for performing cluster analysis in high dimensional spaces [HK03, Sil86]. However datasets with heavy-tails can result in the loss of important information [Ver12]. Numerous methods exist to quantize outliers appearing in the heavy-tail. They are largely classified as adaptive bandwidth estimation and most popular methods include 'balloon estimators' [T*93]. To the best of our knowledge, the KDE method has not been applied to visualize salient clusters appearing in physiologic data. Our TIM visualization, using an adaptive bandwidth KDE, along with other coordinated displays, is designed to help researchers perceive and explore event features and event classifications in order to generate new hypotheses about clinical conditions.

There are also unique challenges that involve the temporal flow of information, that emphasise the trajectory tri-event parameter [CEH*09]. One prominent work, ThemeRiver [HHN00] builds on the metaphor of a flowing river to create smoothed stacked charts of text themes over time, now generally known as stream graphs. This technique, by virtue of its ability to highlight temporal variations, has inspired the development of several modern visualization designs, including FluxFlow [ZCW*14], RoseRiver [CLWW14], and LifeFlow [WGP*11]. In the network security domain, Fischer and Keim apply the NVisAware visual analytic tool, as part of the NStreamAware architecture for providing situational analysis insight [FS14]. The NVisAware system in particular allows the analyst to interact with dynamic feature data streams generated by an event stream algorithm. Xie et al., provide a algorithmic method of highlight salient regions where patterns exist, and hiding nonvariant event streams [XWR10]. They were able to reduce the time required to detect abnormal events.

4. Background and Task Analysis

PhysioEx makes use of Artemis, an online analytic platform for physiologic data streams which detects and classifies physiologic events (PE) in real-time [McG13]. PEs are events generated by the algorithm to detail temporal patterns of the data stream. The neonatal spells algorithm was executed against the retrospectively stored raw data for a total of 47 patients, which generated PEs that were saved to a database in real-time and used in this work. The study was approved by the Research Ethics Board at our institution, and

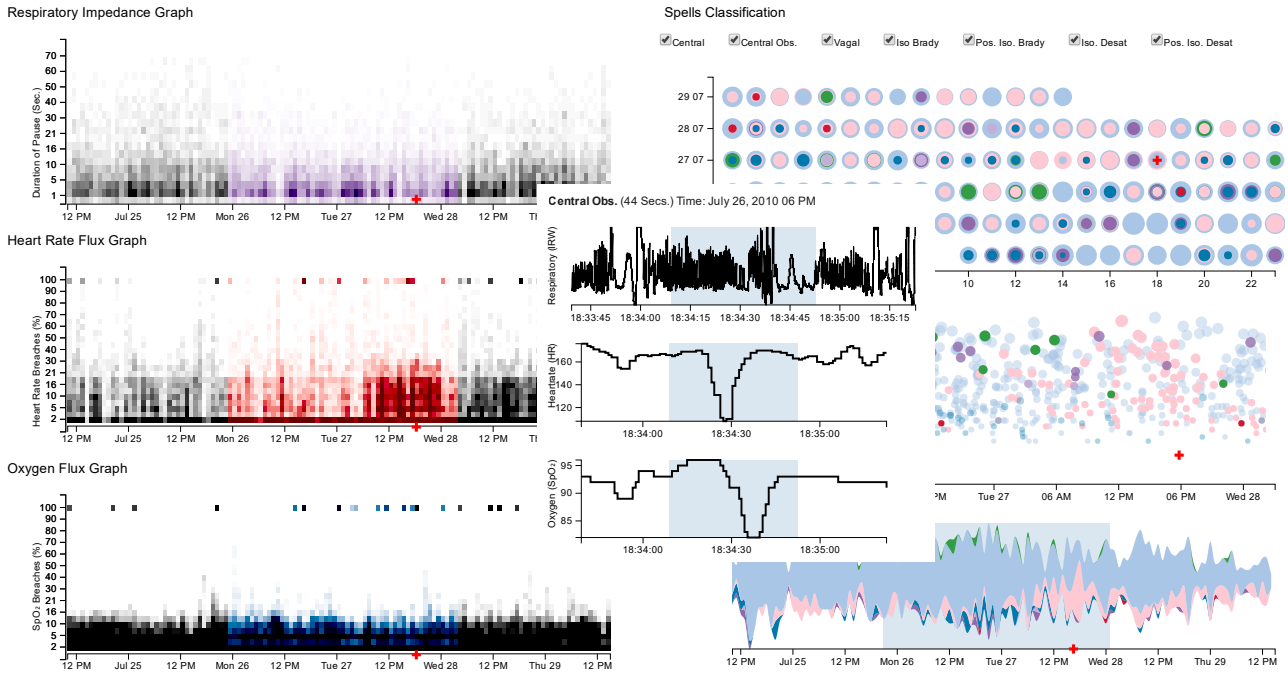


Figure 1: PhysioEx is a visual analysis tool for event stream analysis of multiple streams. Several Temporal Intensity Maps (left), in the coordinated dashboard reveal the duration, frequency, and intensity of physiologic data over time, alongside a selected raw data display (middle), and three visualizations (right, top to bottom): a sequence, linear, and stream graph.

all patient data was de-identified. One of the goals of that study was to determine whether neonatal sepsis was present at the time of suspicion of infection at the bedside.

To better understand the requirements of clinical researchers, we asked three domain experts to describe specific tasks they currently perform to predict physiological behaviours prior to the point of suspicion of infection (PSI). The common tasks were:

T1 *Identify the PSI.* The researcher uses the PSI as an anchor for subsequent analysis.

T2 *Identify PEs in the respiratory physiologic signal before PSI.* PEs having breathing pauses greater than 20 seconds were noted and associated with neighbouring clusters.

T3 *Analyze PEs across heart rate and SpO_2 data streams.* Heart rate signals and blood oxygen saturation signals are analysed to determine downwards shifts before the PSI.

T4 *Identify abnormal PEs.* Abnormal PEs are flagged and sometimes investigated to verify algorithm accuracy.

T5 *Create mental temporal picture of underlying physiology.* Information gathered from all previous steps were used determine a hypothesis about the presence of infection.

Supporting these tasks is our design goal.

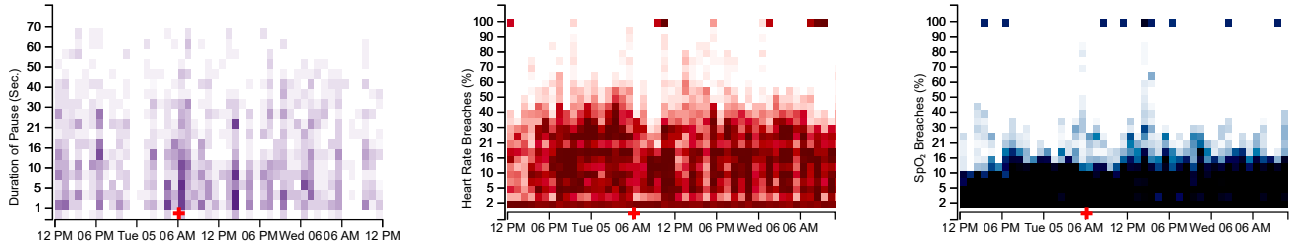
5. Design of PhysioEx

PhysioEx is illustrated in Figure 1, and consists of three groups of views: three TIM views; the sequence graph, linear graph, and streams graph; and three raw data views. The interface was developed using D3 [BOH11]. In this section we explain each component in detail.

The first group of views, namely the Respiratory Pause TIM, Heart Rate Flux TIM, and the SpO_2 TIM provide the user with the ability to rapidly analyse behaviours in event features stream. The second group of displays assist with analysing event classification data. A third view, when activated, provides the user with deeper contextualization by providing raw data that would be observed at the bedside. We mark the canvas with a red cross. This red cross indicates that a blood result was obtained after a physician suspected the infant of having infection. We do not show whether it was positive or negative to allow the researcher to use this position marker to conduct explanatory research for generating hypothesis about the onset of infection.

5.1. Temporal Intensity Map View

Each TIM provides users the ability to rapidly discern subtle behaviour in streaming data. We employ a novel use of the heatmap visual encoding, where positions along the vertical axis represents an aspect of an event's nonlinear *critical distance interval*, such as duration of breathing pause. It is termed a critical distance interval, because it helps determine the PE's severity. PEs are aggregated into critical distance interval bins as determined by the density estimation function. Hence, durations with smaller values are represented at the bottom of the graph while larger durations appear



(a) Respiratory Pause (b) Heart rate flux (c) Oxygen saturation (SpO_2) flux

Figure 2: Temporal Intensity Maps, compact visualizations for gaining rapid situational awareness of low-level behaviours in data streams.

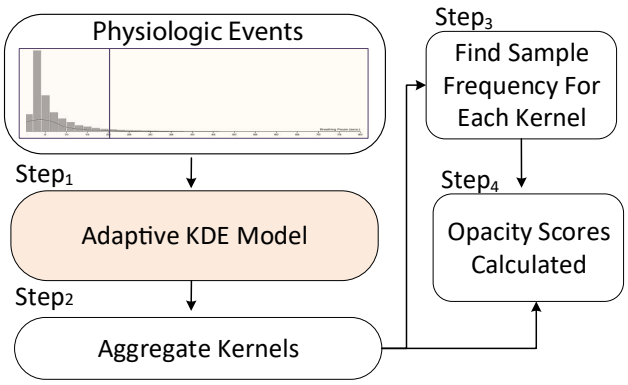


Figure 3: The four-step method of constructing the Temporal Intensity Map beginning with identifying kernels.

near the top. The horizontal axis represents temporal range of the dataset. A red cross is placed where a nominal clinical event (e.g. PSI) exists, to support task T1.

In order to support task T2 and T3, we contribute a combined adaptive bandwidth method of vertical binning, using the KDE generated probability density function (pdf) as illustrated in Figure 3. We began the process by calculating the KDE pdf for the entire dataset (Step 1). We utilized the scikit-learn to implement the density estimation [PVG*11]. The top-hat kernel form, an alternative to Gaussian, was selected as this kernel form involved less smoothing which produced more kernels. The width was also made narrow, and set to a value of 0.2. These two modelling decisions increased the likelihood of kernels identified in the heavy-tail of the distribution. All PEs were then aggregated into hourly sets (Step 2) and reduced to produce sample frequencies for each kernel (Step 3). The binning produces a two-dimensional array of PE critical distance interval sums ranging from 0 to N, where N is the furthest critical distance interval. The value of each element in the array are used to encode opacity.

The visual encoding of the TIM is a heatmap controlled for hue and opacity. The hue indicates the PE classification and is metaphoric: red for heart rate which evokes the colour of blood, and blue for desaturation of oxygen, due to blue-like colour of the skin when oxygen levels fall. The hue selection supports T2–T4, in which one must rapidly associate PE type. The opacity is controlled by the frequency value. The width is controlled by available space of the canvas, divided by the temporal range.

Where there are significant number of samples found in a particular kernel, the opacity score of each is reduced, and where the frequency is low the opacity is increased (Step 4). Thereby, events appearing in low-frequency kernels, such as in the heavy-tailed portions, are represented with increased visibility. These heavy-tail events, such as an extended breathing pause lasting several minutes are clinically significant and warrant increased visibility. Moreover, this method effectively addresses the requirement of highlighting outliers involved in the task T2 and T3, where a constant opacity score would have otherwise excluded them from view. The temporal trajectory of the health status is visually elicited from observations made on each distinct view generated by the encoding. As rectangles with varying hue are appended along the horizontal temporal axis, the user is able to visually glean information about ongoing changes in the physiologic signal. Finally, we considered the use of bar graphs as an alternative design, due to their familiarity. However, that encoding was not appropriate for illustrating all three temporal properties without creating visual clutter. Due to the nature of our dataset, the TIM encoding was more appropriate for identifying both frequency (dense areas) and duration (vertical dimension) along a temporal axis.

Figure 2 illustrates three uses of TIM, beginning with the respiratory pause map (Figure 2a), displaying data in the form of duration of breathing pauses between 0–80 seconds. In this dataset intermittent clusters of breathing pauses are seen throughout the entire duration. Breathing pause durations are also seen extending to pathological ranges above 21 seconds. The heart rate flux (Figure 2b) illustrates a measure between zero variability (0%) to high variability (100%) in heartrate. A sliding window sampling approach is used to compare the instantaneous heart rate every second against the average of the previous 30 seconds. The percent change is calculated and a block is added to the TIM at the appropriate height, if the heart rate reduced (bradycardia). In this chart, clinical researchers would be looking for repeated occurrences of severe bradycardia (high percentage change), or periods of low overall variability (high density low on the TIM). Both are indicative of pathological status.

Figure 2b shows a region of reduced variability (three columns of lighter blocks) after 12 p.m. on Monday, and then a period of high variability with more density (darker red blocks) from 3 p.m. There is high oscillatory behaviour observed in this patient, poten-

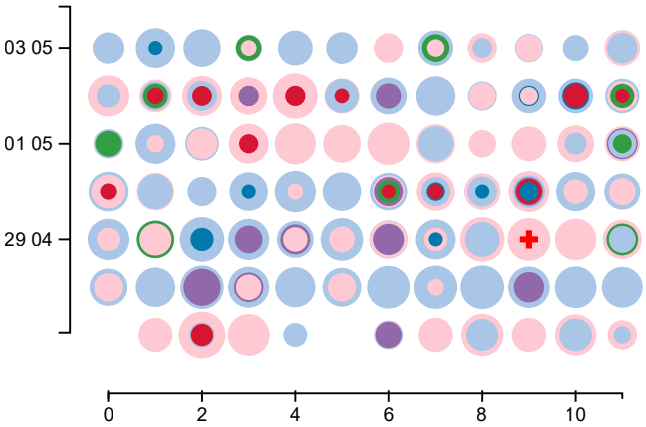


Figure 4: The Sequence Graph, illustrating a matrix of hours by days (truncated to 10 hours). Each bubble's radius encodes the total duration of episodes within that hour, and smaller bubbles are drawn on top.

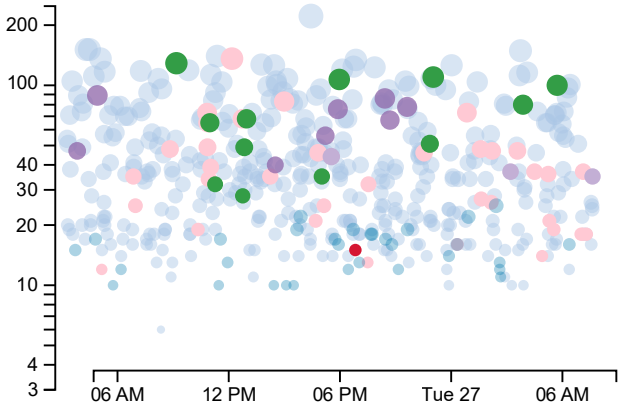


Figure 5: The Linear Graph shows a log-transformed duration of each event classification in a linear temporal view. Hue is used to indicate event classification, and event duration is double-encoded using size.

tially due to the influence of drugs or other systemic influences. Finally Figure 2c illustrates the oxygen flux. The data for this visualization is measured using the same metric as heart rate flux, however oxygen flux data is gathered each time a desaturation occurs in the SpO_2 signal. Observing Figure 2c, one sees a period of low variability initially, followed by a region of higher variability between 12 p.m. on the Tuesday and lasting 24 hours. Blocks at the 100% level in the flux TIMs likely indicate data errors (such as when a sensor disconnected) but are left in the chart as they may be clinically relevant and should be investigated. To differentiate zero data from missing data requires further research and improvements in data collection.

The researcher can use the interactive brushing functionality to highlight a region on any one of the TIM views, all other views are immediately updated to highlight that section. Figure 1 illustrates how each of the TIM views appear when a region is brushed. Here the researcher is interested in 48 hours prior and 24 hours post an infection event. Highlighting this region also triggers coordinated updates across the linear and the streams graph for more detailed analysis of event classifications.

5.2. Physiologic Event Classification Views

We developed three coordinated views to show PE classifications, coming from Artemis, including the sequence graph, linear graph and streams graph. We use similar hues with varying saturation to highlight complementary PE classifications of varying severity. For instance, an *isolated bradycardia* receives a more saturated pink than a *possible isolated bradycardia*. Oxygen desaturation events are blue.

5.2.1. Sequence Graph View

The first PE classification view found on the top right of 1 is the sequence graph (highlighted in Figure 4). This view supports T5, in which the user requires a rapid means of understanding temporal discontinuous event data. The advantage of this representation is that it reveals events occurring during the same hour across multiple days. This can be useful in associating the influence of routine events, such as bed-side interventions to changes in physiologic data. Each vertical position represents the same hour over multiple days. Specifically, the horizontal x-axis shows progression over 24 hours, and the vertical y-axis shows progression of events over days of the month. The axes can be configured to express seconds (x) by minutes (y), or days (x) by months (y), each producing a periodic view of high-level event classifications.

In order to control the size of circle in this view, we calculate the sample frequency for every hourly epoch. Less significant PE classifications receive a lower opacity, while more significant PE classifications have higher opacity. This allows the user to visually discern areas where greatest clinically significant PEs exist. The radius encodes for the log transform of the total duration in the hour (default view, Figure 4). The transformed values are then sorted in descending order and painted largest to smallest, producing a layered view. The fill hue is determined by the event classification type. The user can hover of the circle to reveal details of each inner circle. An alternative to this design was to use a stacked bar representation, which summarized the frequency of each event over the hour. However that representation does not convey periodic events that occurred over the same time-period spanning multiple days.

Figure 4 illustrates a vagal PE (green) at 1 a.m., followed by central apnoea PE (purple), at 3 a.m., 4 a.m., and 6 a.m. (horizontal) on 29th day of April (vertical). Possible isolated bradycardia (pink) and possible isolated bradycardia (light blue) are sustained over the next several hours. The researcher notices that a red cross, denoting a PSI, is visible at 9 a.m. that day. The researcher notes that till that period, the salient and clinically relevant PEs have become more prevalent by integrating the observed frequencies of vagal, central and possible isolated bradycardia and desaturation PEs.

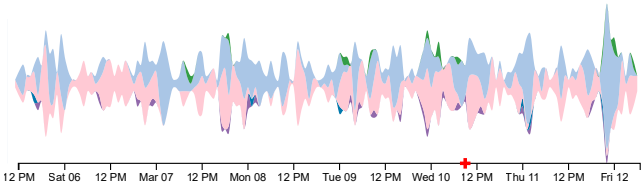


Figure 6: The Stream Graph illustrates the flow of event classification frequency over the analysis duration.

5.2.2. Linear Graph View

The linear graph, Figure 5, supports T4 and T5, in which the user identifies abnormal PEs as well as requiring detailed temporal view of all PEs over time. The y-axis represents a log transform of PE duration and x-axis the linear timeline view. PEs are plotted as circles where hue is determined by the classification type. The radius is double encoded with the log transform of the duration value. Reduced opacity is applied to PEs that are less important, while PEs with higher clinical significance maintain full opacity. Smaller bubbles are of low durations, while high duration events are larger and have more prominence at the top of the graph. A tooltip is available for additional information about each PE. Selecting a PE launches an overlay view of the associated raw data graphs. Figure 5 illustrates several prominent vagal apnoea (green) PE appearing before 12 p.m. and continuing till 6 a.m the following day. Intermittent central apnoea events (purple), along with possible isolated desaturations (pink) and possible isolated bradycardias (light blue), are observed throughout the night. Event classifications are rendered according to their frequency, and severity. Low severity events like possible isolated bradycardia and desaturation are rendered first, followed by the more significant PEs.

5.2.3. Streams Graph View

The third event classification view, illustrated in Figure 6, is the streams graph, revealing continuous event classification frequency over time, with the data summed to a count per hour and supports tasks T3–T5. Each stack is coloured with the event classification hues shared across all event classification views. A tooltip is available to explore details about the event classification. Brushing a stack causes all other stacks to fade, giving visual prominence to the hovered stack and reducing clutter. Figure 6 illustrates relatively high frequencies of possible isolated bradycardia (pink) lasting from the 12 p.m. mark, along with possible isolated desaturations (light blue), until 12 a.m.. Following that, possible isolated desaturation events diminish, only to return again in the late afternoon of the 6th. Between this range, there are also several other PE classifications identified, such as intermittent central apnoea episodes (purple), and vagal apnoea (green). An alternative to this design was to use line graphs, while commonly utilized in electronic medical records, the line graph encoding fared poorly when compared to the streams graph. The streams graph, through the use of filled area, allowed the user to rapidly elicit information about the most frequent event within one or more time windows.

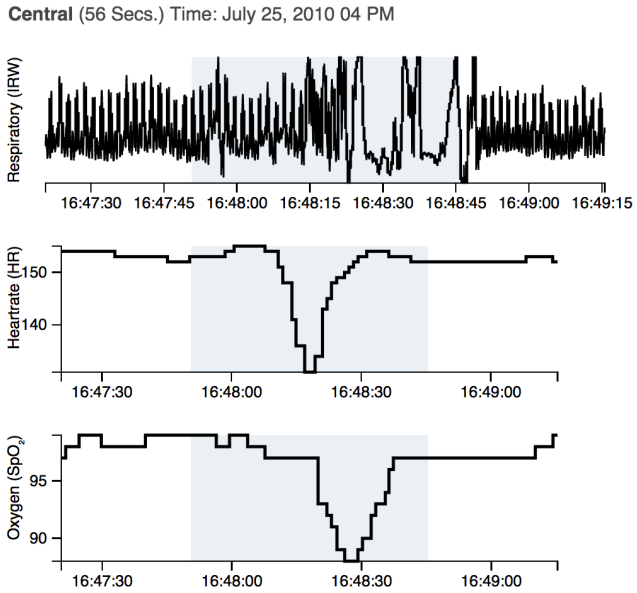


Figure 7: The Raw Data View displays sensor data using 3 line graphs. The highlighted region corresponds to a PE classification, and the white regions before and after the event are 30 second buffers for improved contextualization.

5.3. Raw Data View

The final user interface component, designed to primarily support T5, which serves as a critical step in confirming whether a patient is believed to be positive for sepsis, is the raw data display illustrated in Figure 7. In this view the respiratory pause graph is displayed at the top, followed by the heart rate trace, and finally the oxygen saturation graph at the bottom. This view is activated when the user performs a selection on one of the PE classifications in the linear graph view. In this view the analyst can immediately access low-level sensor data that lead up to the PE classification. This line-graph method is a familiar design for displaying sensor data. A background band is appended to the chart, representing the actual duration for the event classification.

5.4. Coordinated Analysis

When an analyst selects a portion of a graph using interactive brushing, all other graphs immediately update to highlight that section. For instance, in Figure 1, the highlighted region appears prominent in colour on each of the TIM displays, and is also highlighted the streams chart. The linear chart is zoomed in to show the selected time period in detail, from 6 a.m. of the 26th to 6 a.m. of the 28th. Due to its nature as a summary graph, the sequence graph maintains its view to provide high level details. In Figure 1, the analyst has also selected a central obstructive event on the linear chart, thereby revealing the low level raw physiologic waveform traces in an overlay.

6. Analysis Scenario: Neonatal Spells

To fully demonstrate the features of PhysioEx we synthesized the following analytic scenario based on observations and comments from semi-structured interviews held with four expert participants. The scenario corresponds to the PhysioEx view in Figure 1.

A neonatal researcher is conducting an analysis of a patient case to explore physiological patterns that led up to the red cross that denotes the point of suspicion of infection. Upon launching a patient dashboard he is drawn to the respiratory pause TIM. The researcher knows from experience that respiratory pauses lasting longer than 20 seconds are pathological and notices that there are very few regions in the map where longer pauses exist. This informs him of the infant's rather uneventful respiratory physiology. He then looks below to the Heart Rate Flux TIM to analyse the corresponding behaviour in this physiologic stream and notices that there are bigger swings in heart rate about 2–4 hours prior to the suspicion of infection, marked by one red cross on all plots. The Oxygen Flux TIM looks normal. The researcher then uses interactive brushing to highlight the two to four hour period. This action updates all TIMs by highlighting the area in focus with colour, and desaturates all other areas. At the same time, the linear map is also updated to show events that manifest during the highlighted interval.

The researcher looks at the linear graph to see that in this window there is a central apnoea episode that lasted for 107 seconds. To analyse the event further and to confirm the low-level sequences he selects the event to reveal raw physiological signals. The physiological signals look depressed, to confirm that this behaviour is not seen throughout other event classifications, he selects the neighbouring bubble. There again, the event classification shows abrupt breathing on the impedance respiratory waveform chart of the Raw Data view. He moves to looking at the streams graph to get an overview of all central apnoea event classifications that occurred over the entire timeline by hovering over the stream coloured in purple. Having seen other central events also occurring well in advance of the red cross, the researcher generates a hypothesis about the infection state of this infant. He believes this infant is not suspected to have infection.

7. Expert Evaluation

We conducted an expert evaluation to gain a better understanding of the utility of PhysioEx for clinical researchers. The primary condition in this study was the visualization technique, with two levels. PhysioEx was compared to a stacked bar view that is currently used to perform clinical research of neonatal spells behaviour [MJE*13]. Due to the difficulty in recruiting a large number of highly specialized domain experts, we adopt a primarily qualitative evaluation approach, engaging the available experts in real analysis tasks and both observing their experience and requesting their feedback, to build a holistic understanding of the potential for PhysioEx.

Participants: We engaged four domain experts with experience working with neonatal physiologic data on a day-to-day basis ranging from five to 35 years. Three of the experts were males and one was a female. All four experts report using the computer at least once a work day for analytic purposes. Both visualization techniques used in the study were unknown to all participants.

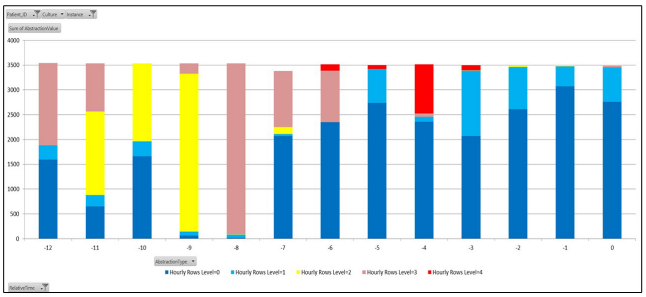


Figure 8: Stacked bar representation of cardiorespiratory behaviour prior to the suspicion of infection.

Dataset: The study dataset consists of 29 patients who were suspected of infection and for whom we had truth data about the presence of infection. Suspicion of infection was defined by the presence a blood draw for a laboratory test for infection. The results of the laboratory test provided the truth data for this study. The apnoea event classification algorithm was run over seven days' worth of data for each patient: 120 hours before and 48 hours after the time the blood culture results were received. Prior research suggests that neonatal sepsis may be detected in physiological data several days before current practices suspect it at the bedside. We decided to use this case study as it provides an exploratory means by which the domain experts can investigate and potentially arrive at novel findings.

Task and measures: The task of the domain experts was to use each visualization to determine whether the patient has an infection (sepsis) and state any additional insights they had about the data. We measured the accuracy of determination of infection and the time taken in analysis. In addition, we engaged participants in a semi-structured interview about their analysis process, preferences, and usability issues which arose. Screen and voice recording was used to allow for detailed analysis as well as easy transcription of the collected data.

7.1. Stacked Bar Graph

We compare PhysioEx against a stacked bar graph, illustrated in Figure 8, which has been used by clinical researchers to understand trends in neonatal spells (PEs relevant to the neonatal domain) preceding a point of suspicion of infection [MJE*13]. This view provides a higher level and non-interactive view on the physiologic data by counting event classifications every second and summarizing them on an hourly basis. In the stacked bar graph the *levels* refer to event classifications (in order): all normal, heart rate variance changes, respiratory pauses, oxygen desaturation, and blood pressure drop. The stacked bar view is aligned with the time of the suspicion of infection (red cross on PhysioEx) at the zeroth hour, then all the preceding events sorted by hour to the left. An analyst would start at the zeroth hour to analyse the spells behaviour. Figure 8 shows that beginning at the -12th hour up to the -4th hour there are sustained fluctuations in the infant's cardiorespiratory status. The infant seems to be improving as it approaches the zeroth mark (more times classified as normal). Note that there is missing

data in hour -7, perhaps due to infant movement or sensor malfunction.

7.2. Procedure

Each session began with a brief introduction to the study and a semi-structured interview to assess participant prior knowledge about the domain of neonatal care and spells. This was followed by a series of 7 training trials, 11 timed experimental trials, and a brief questionnaire eliciting feedback on the interface design, repeated for each visualization technique. Due to data availability, the same dataset of 7 samples (in random order) were used for the training trials in both conditions. For the experimental trials, two different datasets of 11 samples each was used, one for each condition. The ordering of technique and experimental datasets was counterbalanced. There was a total 2 techniques $\times$ 11 trials $\times$ 4 participants $=$ 88 trials. The analysis task was repeated for each training and experimental trial. Feedback about the correctness of determination of infection was provided during training. Participants were asked describe spells activity around the point of suspicion of infection and to state whether an infection was present. At the end of the experiment, a brief questionnaire was administered to collect participant preference between techniques. Experimental sessions lasted two hours and participants were able to take breaks as needed.

7.3. Results

In this section we report the results of the study comparing PhysioEx and a stacked bar of complex physiologic data. The accuracy of determination of sepsis was uniformly distributed and below 50%, for the dataset containing 7 sepsis and 22 non-sepsis patients for both conditions, thus we did not investigate further. We instead focus our analysis on the quality and depth of insights expressed during the analysis process, and the subjective feedback.

7.3.1. Identification of Physiological Behaviours

Experts provided a range of comments on the stacked bar method of representation while they investigated physiological status of that infant. Although the stacked bar provided a simple interface for identifying how much of the hour was attributed to one physiological measure, it did not provide additional and more salient information about the changes that took place within the hour. Experts found it difficult to discern the events that occurred uniformly in the hour, the associated severity, and distribution of durations. The time to analysis was however, rapid, with a majority of the analysis being completed within ten seconds.

Meanwhile PhysioEx allowed them to rapidly elicit physiological behaviour, frequency within an hour, the duration of all event classifications aggregated in an hour in addition to duration of single classifications. When asked to describe the physiological status of the infant, experts often spent several minutes describing the intricate behaviour, frequency, duration and sequences of events seen in TIMs and also on the sequence and streams graph. This was seen consistently, with analysis time ranging from 2-10 minutes per patient. One expert comments about the Respiratory TIM: "I see a burst of activity here, on this Friday starting before 11 p.m., and going through to about noon, then I see a trivial amount of activity about 24 hours later, and then I see another burst of activity starting about midnight starting about the 28th, which seems to be of the same intensity as the first burst I observed but has a longer duration. In the middle, I see very little variation." The stream graph was also noted to be a unique tool in the domain of physiological research. Experts had not encountered this representation and therefore required some time to adapt to it. One expert found that he was relying on it as a final 'truth' indicator, after having analysed all other representations.

7.3.2. Hypothesis generation

Using the stacked bar view, experts found it difficult to generate hypotheses unless there was a clear and distinguishable trend. Where events occurred without any clear trend, all experts stated difficulty with determining whether these events had any relationship with the point of interest at the zeroth hour. All experts described the colour scheme to be very favourable when determining patterns and trends. One expert mentioned "I'm looking for the stacks with a lot of yellow, the red is distracting for me, but the yellow is interesting". Another expert physician stated that "[the stacked bar] is too simple, it doesn't work for me".

Experts described PhysioEx as useful, and powerful when generating hypotheses, they also mentioned that the coordinated interactive brushing was most useful when they wanted to reaffirm incremental patterns. They found the coordinated brushing and highlighting across all TIMs provided the most benefit in terms of closely analysing neonatal spells preceding the infection suspicion point. The ability to select the event classifications to reveal low-level sensor data was appreciated by all experts and heavily utilized by one expert. Two experts were able to derive bed-side intervention information from the patterns exposed on the Respiratory Pause TIM. They revealed information about potential respiration modality of the infant. Some quotes received from experts include:

"Oh wow look at that... look at that... this is a baby that got intubated... a fully manually intubated baby. Well this child cannot apnoea... if you look at the respiratory pauses they are all so uniform."

"Look at the heart rate variability, it swings everywhere and then it comes back. [...] It looks like they had a ventilator mode change, maybe to biphasic, but they've also taken a culture at the same time, this is an odd practise, we tend to do things one at a time."

7.3.3. Satisfaction of Use

Domain experts who used both the stacked bar view and PhysioEx reported greater satisfaction with the simplicity of the former, but expressed concern over excessive simplicity and hiding of potentially useful data. When analysing trends on the stacked bar view experts found that while they were able to verbalize trends of high-level event classifications more easily, they were unable to provide detailed descriptions. The stacked bar, however, provided the domain experts with a familiar format. This familiarity factor contributed to reduced training times.

Experts were encouraged by additional details presented in PhysioEx. The TIM representations were favourably received by all. They paid keen attention to the behaviours expressed in heart rate

and SpO_2 TIMs, and stated that it was helpful when conceptualizing the infant's status over many hours. The sequence graph was used by three experts for determining sequences of events prior to the suspicion of infection, the fourth expert did not use the display at all. On the simplicity of PhysioEx, the responses were mixed. While domain experts greatly appreciated the increased level of detail, it also proved to be cognitively demanding task, requiring learning new interaction methods for selecting, filtering, and retrieving information about physiological signals. The experts attributed the cognitive load due to the overwhelming number of possible events that had prominence in almost all patients. Moreover, experts also noted the usefulness and utility of PhysioEx could be even further improved with the addition of contextual information, such as the infant's gestational age, gender, method of respiration, and other comorbidities.

7.3.4. General Comments

Experts provided numerous comments on the usability and potential utility of PhysioEx. Two experts, also physicians, mentioned that TIMs may contribute additional means of gaining insight on subtle physiological behaviours of the infant that are currently unavailable for bedside decision makers. Six coordinated views, as currently instantiated in PhysioEx, were found to be useful for research but likely too complex for use at the bedside. All experts using the TIMs representation were immediately cognizant of the data quality available for analysis. Data quality is an ongoing challenge in the neonatal intensive care environment. However, obtaining consistent and continuous data samples is very difficult, due to frequent disconnects from sensors.

8. Discussion and Future Work

We used an expert evaluation consisting of four domain experts analysing neonatal spells behaviour in an attempt to predict the likelihood of neonatal sepsis. Although the results of determination of sepsis in our dataset was inconclusive, our study revealed that PhysioEx deeply involved clinical researchers in the analytic pipeline. Experts using PhysioEx were able to verbalize subtle physiologic behaviour spanning numerous days and for numerous patients. Many of the insights discovered with PhysioEx were hidden by the stacked bar. While the time for using PhysioEx was much longer, this may be explained by the richer interface, interactivity, and novelty of the visualization. Rapid analysis is needed in bedside situations, but for retrospective research, such as analysing the relationship of physiologic measures, spells, and neonatal sepsis, depth of insight is more important than speed.

PhysioEx gave experts the first opportunity to interactively explore physiological event features and event classifications. To our knowledge, there are currently no other tools that provide interactive exploration of detailed physiological changes of low-level clinical events. However, introducing such a novel tool does have limitations. Some experts experienced fatigue after enduring a long training phase and then analyzing a total of 18 patients on PhysioEx. Contributing to the fatigue was the significant cognitive load imposed by using a novel tool to perform a difficult task. The TIM views provided experts with a simple and rapid method of appreciating physiologic behaviour. Most experts relied on the TIMs to base their decisions on whether the infant was experiencing normal or abnormal changes in physiology. Dense and low density regions were rapidly identified by all experts. This information was then augmented by the event classification display. Experts, especially practitioners, also used the TIMs to characterise the data quality for that particular patient. Since this is a commonly faced issue in NICUs, the ability to see drops in data quality gave more insight about the infant and their management.

The sequence graph was heavily utilized by some to track incremental hourly changes leading up to the point of suspicion. One expert commented that the bubble matrix provided a unique ability to recognize patterns that commonly occur at various times of the day. Events such as blood draw occurring in the afternoon, loss of data for short durations, and transfer of the infant to other units, were speculated. One physician was able to associate the periodic appearance of bradycardia during morning rounds, and expressed anecdotal evidence to suggest this relationship. While this information was provided to the experts, the ability for the experts to augment clinical expertise provides an opportunity as future work for automated annotation capabilities for PhysioEx. The automated annotation of events would further supplement researchers with much needed context to explore the event space in more detail.

PhysioEx was found to provide a greater advantage to explain neonatal spells behaviour than the current alternative. One expert physician with extensive involvement in neonatal spells research, had mentioned that they are now inclined to invest a day in training a neonatal fellow so they would be better able to describe physiological behaviour of spells. There are however, limitations with PhysioEx and our preliminary study. We only tested PhysioEx with four expert participants drawn from the larger clinical researcher population. Moreover, there are no established clinical links yet between neonatal spells and infection. Therefore, the experts participating in the study were not looking for known associations. Many experts noted that lack of contextual information (patient metadata) as a limitation of both techniques. We had developed PhysioEx to cater to exploring physiological data, however in future work incorporating clinical information would certainly be highly advantageous for supporting analytic activities. To address the cognitive overload from analysing several patients independently, in future work we intend to propose new techniques that assist in analysing population cohorts in a single view.

9. Conclusion

In this paper, we contribute a novel visualization technique, the Temporal Intensity Map, and introduced PhysioEx, a visual analytic tool for complex multidimensional sensor data exploration. We present a task analysis for designing visualization displays for the complex and heterogeneous sensor network environment in neonatal care and draw on this analysis to inspire design. Our preliminary study supports further investigation into PhysioEx as an important addition to the tools available for clinical researchers. In future work we aim to deploy PhysioEx to support additional use cases, such as exploring physiological behaviours for other clinical conditions. Moreover, we aim to integrate more contextual information such as clinical histories into PhysioEx for the development a more tightly integrated physiological clinical research system.

References

[AMM*08] AIGNER W., MIKSCH S., MULLER W., SCHUMANN H., TOMINSKI C.: Visual methods for analyzing time-oriented data. *IEEE Trans. on Visualization and Computer Graphics 14*, 1 (2008), 47–60. 2

[BCO*12] BLENCOWE H., COUSENS S., OESTERGAARD M. Z., CHOU D., MOLLER A.-B., NARWAL R., ADLER A., GARCIA C. V., ROHDE S., SAY L., ET AL.: National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet 379*, 9832 (2012), 2162–2172. 2

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D^3 data-driven documents. *IEEE Trans. on Visualization and Computer Graphics 17*, 12 (2011), 2301–2309. 3

[CEH*09] CHEN M., EBERT D., HAGEN H., LARAMEE R. S., VAN LIERE R., MA K.-L., RIBARSKY W., SCHEUERMANN G., SILVER D.: Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications 29*, 1 (2009), 12–19. 2

[CLWW14] CUI W., LIU S., WU Z., WEI H.: How hierarchical topics evolve in large text corpora. *IEEE Trans. on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2281–2290. doi:10.1109/TVCG.2014.2346433. 2

[EF10] ELMQVIST N., FEKETE J.-D.: Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Trans. on Visualization and Computer Graphics 16*, 3 (2010), 439–454. 2

[Fai13] FAIRCHILD K. D.: Predictive monitoring for early detection of sepsis in neonatal ICU patients. *Current Opinion in Pediatrics 25*, 2 (2013), 172–179. 2

[FMK12] FISCHER F., MANSMANN F., KEIM D. A.: Real-time visual analytics for event data streams. In *Proc. of the ACM Symp. on Applied Computing* (New York, New York, USA, 2012), ACM Press, p. 801. 2

[FMLD10] FLOWER A. A., MOORMAN J. R., LAKE D. E., DELOS J. B.: Periodic heart rate decelerations in premature infants. *Experimental Biology and Medicine 235*, 4 (2010), 531–538. 2

[FS14] FISCHER F., STOFFEL F.: NStreamAware: Real-time visual analytics for data streams (vast challenge 2014 mc3). In *Proc. of the IEEE Conf. on Visual Analytics Science and Technology (VAST)* (Oct 2014), pp. 373–374. doi:10.1109/VAST.2014.7042572. 2

[GAK*11] GSCHWANDTNER T., AIGNER W., KAISER K., MIKSCH S., SEYFANG A.: Carecruiser: Exploring and visualizing plans, events, and effects interactively. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE* (2011), IEEE, pp. 43–50. 2

[GOB*03] GRIFFIN M. P., O'SHEA T. M., BISSONETTE E. A., HARRELL F. E., LAKE D. E., MOORMAN J. R.: Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. *Pediatric Research 53*, 6 (2003), 920–926. 2

[GZA06] GOTZ D., ZHOU M. X., AGGARWAL V.: Interactive visual synthesis of analytic knowledge. In *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)* (2006), IEEE, pp. 51–58. 2

[HHN00] HAVRE S., HETZLER B., NOWELL L.: ThemeRiver: visualizing theme changes over time. In *Proc. of the IEEE Symp. on Information Visualization* (2000), IEEE Comput. Soc, pp. 115–123. doi:10.1109/INFVIS.2000.885098. 2

[HK03] HINNEBURG A., KEIM D. A.: A general approach to clustering in large databases with noise. *Knowledge and Information Systems 5*, 4 (2003), 387–415. 2

[HVF13] HURON S., VUILLEMOT R., FEKETE J.-D.: Visual sedimentation. *IEEE Trans. on Visualization and Computer Graphics 19*, 12 (2013), 2446–2455. 2

[KBK11] KRSTAJIC M., BERTINI E., KEIM D. A.: Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Trans. on Visualization and Computer Graphics 17*, 12 (2011), 2432–2439. 2

[Ket12] KETT J. C.: Anemia in infancy. *Blood 1599* (2012), 1678. 2

[LJH13] LIU Z., JIANG B., HEER J.: immens: Real-time visual querying of big data. *Computer Graphics Forum (Proc. EuroVis) 32*, 3 (June 2013), 421–430. doi:10.1111/cgf.12129. 2

[LPD*11] LILLIE E. O., PATAY B., DIAMANT J., ISSELL B., TOPOL E. J., SCHORK N. J.: The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized medicine 8*, 2 (2011), 161–173. 1

[McG13] MCGERGOR C.: Big data in neonatal intensive care. *IEEE Computer 46*, 6 (2013), 54–59. 2

[MDF*11] MOORMAN J. R., DELOS J. B., FLOWER A. A., CAO H., KOVATCHEV B. P., RICHMAN J. S., LAKE D. E.: Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring. *Physiological measurement 32*, 11 (2011), 1821. 2

[MJE*13] MCGREGOR C., JAMES A., EKLUND M., SOW D. M., EBLING M., BLOUNT M.: Real-time multidimensional temporal analysis of complex high volume physiological data streams in the neonatal intensive care unit. In *MedInfo* (2013), pp. 362–366. 7

[MMC86] MARTIN R. J., MILLER M. J., CARLO W. A.: Pathogenesis of apnea in preterm infants. *The Journal of Pediatrics 109*, 5 (1986), 733–741. 2

[PVG*11] PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., ET AL.: Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research 12* (2011), 2825–2830. 4

[Sil86] SILVERMAN B. W.: *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986. 2

[T*93] TURLACH B. A., ET AL.: *Bandwidth selection in kernel density estimation: A review*. Université catholique de Louvain, 1993. 2

[TMK03] TOYODA T., MOCHIZUKI Y., KONAGAYA A.: GSCope: a clipped fisheye viewer effective for highly complicated biomolecular network graphs. *Bioinformatics 19*, 3 (2003), 437–438. 2

[TPE*13] THOMMANDRAM A., PUGH J. E., EKLUND J. M., MCGREGOR C., JAMES A. G.: Classifying neonatal spells using real-time temporal analysis of physiological data streams: Algorithm development. In *Proc. of IEEE Point-of-Care Healthcare Technologies* (2013), IEEE, pp. 240–243. 1, 2

[Ver12] VERMEESCH P.: On the visualisation of detrital age distributions. *Chemical Geology 312* (2012), 190–194. 2

[WGP*11] WONGSUPHASAWAT K., GUERRA GÓMEZ J. A., PLAISANT C., WANG T. D., TAIEB-MAIMON M., SHNEIDERMAN B.: LifeFlow: visualizing an overview of event sequences. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI)* (2011), ACM, pp. 1747–1756. 2

[XWR10] XIE Z., WARD M. O., RUNDENSTEINER E. A.: Visual exploration of stream pattern changes using a data-driven framework. In *Advances in Visual Computing*. Springer, 2010, pp. 522–532. 2

[ZCW*14] ZHAO J., CAO N., WEN Z., SONG Y., LIN Y.-R., COLLINS C.: #FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Trans. on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 1773–1782. doi:10.1109/TVCG.2014.2346922. 2