# Abstracting Steady Qualitative Descriptions over Time from Noisy, High-Frequency Data

Silvia Miksch[1], Andreas Seyfang[1], Werner Horn[2,3], Christian Popow[4]

[1] Institute of Software Technology, University of Technology, Vienna,
{silvia, seyfang}@ifs.tuwien.ac.at
[2] Department of Medical Cybernetics and Artificial Intelligence, University of Vienna
werner@ai.univie.ac.at
[3] Austrian Research Institute for Artificial Intelligence, Vienna
[4] Department of Pediatrics, University of Vienna
popow@akh-wien.ac.at

**Abstract.** On-line monitoring at neonatal intensive care units produces high volumes of data. Numerous devices generate data at high frequency (one data set every second). Both, the high volume and the quite high error-rate of the data make it essential to reach at higher levels of description from such raw data. These abstractions should improve the medical decision making. We will present a time-oriented data-abstraction method to derive steady qualitative descriptions from oscillating high-frequency data. The method contains tunable parameters to guide the sensibility of the abstraction process. The benefits and limitations of the different parameter settings will be discussed.

## 1 Introduction

Our application domain is the treatment planning for premature infants at neonatal intensive care units (NICUs). Many neonates need artificial ventilation for various reasons. Compared to the treatment of adults, mechanical ventilation of newborn infants is a highly sophisticated task because of the immature structure of their lungs. While medical knowledge has greatly increased over the past years [5], the integration of the data produced by today's monitoring devices into the therapy-planning process still remains an unsolved problem.

Monitoring mechanically ventilated neonates is a clinical, high-frequency domain. Various of devices yield a rather high volume of measured data – at a typical rate of one value per second – which is often faulty. Each measured data shows only a snapshot of a single aspect of the patient's situation in a particular moment.

To a physician these snapshots alone are of limited use. What she needs is an overview over a certain period of time and over various parameters which together give a more detailed and comprehensible picture of the patient's condition. Often she thinks in terms like "X is higher than normal for five minutes". Nevertheless some monitoring devices in current use show only the values measured in the previous seconds or even only the very last one. This leads to a strong need for facilities to visualize raw data as well as their abstractions.

While the sensors send possibly wrong numbers at high precision that represent a parameters value at a certain point of time, human users distinguish but a few different states like very high or medium low in context with a interval of time during which such a proposition holds. To close this gap we developed an algorithm to obtain maximum intervals during which a parameter stays constant. As a vehicle to reach this we introduce a statistically motivated representation of quantitative values, called a *spread*, which shows both position and uncertainty of a value at each point of time.

The ultimate goal of the algorithm is to present the information gathered from various monitoring devices as concise as possible to the physicians in order to reduce their information overload and improve the quality of care.

Currently we are acquiring and analyzing five types of input from various sources. The ECG gives the heart rate rather reliably. The pulse oximetry gives both arterial hemoglobin saturation of oxygen in the blood ($S_aO_2$) and pulse rate. Small movements of the patient result in a high volume of erroneous oscillations of these values. Transcutaneous electrodes measure the partial pressure of oxygen ($P_{tc}O_2$) and carbon dioxide ($P_{tc}CO_2$). We are analyzing data off-line. It is envisioned to employ the findings obtained thereby in on-line monitoring and alarming in the future.

In section 2 we show why related approaches do not cover our problem specification. Section 3 features the three steps of our algorithm which are elimination of data errors, clarifying the curve, and qualifying the curve. In section 4 we discuss the parameters involved in the process. In section 5 we discuss application and further direction of our work.

## 2 State of the Art

Temporal dimensions are a very important aspect in the medical domain, particularly when dealing with the interpretation of continuously assessed data. The most common methods are time-series analysis [1], control theory, probabilistic or fuzzy classifiers. These approaches have a lot of shortcomings, which lead to applying knowledge-based techniques to derive qualitative values or patterns of current and past situations of a patient, called *temporal data abstraction*. Several significant and encouraging approaches have been developed in the past years.

Haimowitz et al. [2] have developed the concept of trend templates ($TrenD_x$) to represent all the information available during an observation process. A trend template defines disorders as typical patterns of relevant parameters. These patterns consist of a partially ordered set of temporal intervals with uncertain endpoints. Trend templates are used to detect trends in time-stamped data.

The RÉSUMÉ project [11] performs temporal abstraction of time-stamped data without predefined trends. The system is based on a knowledge-based, temporal-abstraction method, which is decomposed into five sub-tasks: temporal context restriction, vertical temporal inference, horizontal temporal inference, temporal interpolation, and temporal pattern matching.

Larizza et al. [7] have developed methods to detect predefined courses in a time series. Complex abstraction allows to detect specific temporal relationships between intervals. The overall aim was to summarize the patient's behavior over a predefined time interval.

Keravnou [6] focuses on the periodicity of events derived from the patient history.

All these approaches are dealing with low-frequency data. Therefore, the problems of oscillating data, frequently shifting contexts, and different expectations of the development of parameters are not covered.

Two promising approaches for high-frequency data are the "Time Series Workbench" [4], which approximates data curves through a series of line-segments, and the temporal data abstraction module in the VIE-VENT system [8], which focuses on high-frequency domain of artificial ventilation of newborn infants. Its abstraction module consists of five different methods to arrive at unified, context-sensitive qualitative descriptions: context-sensitive transformation of quantitative data points into qualitative values (context-sensitive schemata for data point transformation), smoothing of data oscillating near thresholds, smoothing of schemata for data point transformation, context-sensitive adjustment of qualitative values, and transformation of interval data (context-sensitive and expectation-guided schemata for trend-curve fitting). VIE-VENT's smoothing and abstraction methods are a very good starting point. However, these methods are quite ad-hoc and do not cover in-depth analysis of the data curve over a longer period of time.

## 3 The Temporal Abstraction Method

The temporal abstraction method obtains intervals, in which a qualitative value stays steady, from oscillating raw data. In the struggle for smooth, steady curves one is confronted with two types of disturbances: *errors*, *noise*, and *physiological variations*.

Most *errors* can clearly be distinguished from correct input data. The absolute values of erroneous data points or the difference to their neighbors are beyond well-defined limits. Reasons for errors comprise technical details like the automatic re-calibration of the transcutaneous electrodes every three to four hours as well as unfortunate circumstances like sensors being badly attached.

*Noise* consists of small rapid oscillations of the measured values that cannot be sorted out as errors. They have very different reasons which makes them hard to handle. Some of them are caused by technical details of measuring devices and can be considered as small errors. Others are medically explicable phenomena or symptoms (e.g. variability of the heart rates) which we subsume as *physiological variations* and which should not be suppressed by the abstraction process.

It is thus clear that all transformations of the curves need parameters that control the amount of abstraction or smoothing that is performed. These parameters need to be carefully adjusted to the issues of medical practice. In section 5 we will present some observations on this topic.

The following steps of processing and abstracting the data can be distinguished:

1. **Eliminating data errors**. Sometimes up to 40 % of the input data is obviously erroneous e.g. exceeding the limits of plausible values.
2. **Clarifying the curve.** Transform the still noisy data into a steady curve with some additional information about the distribution of the data along that curve.
3. **Qualifying the curve.** Abstract qualitative values, like "normal" or "high", from the quantitative data and join data points of equal values to time intervals (qualitative description).

The results of each step of processing is displayed to the physician in combinations of choice to give her a clear perception of the abstraction process. In the following we will detail these three steps.

### 3.1 Eliminating Data Errors

Many errors can be eliminated by defining rather strict maximum and minimum values for each type of input as well as maximum change rates. Furthermore, if two types of input ($S_aO_2$ and pulse) come from the same sensor and one of the two is invalid, it can be concluded that the other one is not valid either. If two sensors measure the same value (pulse and heart rate) and their inputs differ, then you can discard the less reliable one (or do some adaptation).

Still a number of faulty data points – those which fall just inside the range of allowed values – and nearly all of the noise will be left after such processing. They must be handled with in the other steps of the method. See Horn et al. [3] for a thorough discussion of error detection and correction in the domain of clinical monitoring.

### 3.2 Clarifying the Curve

The algorithm presented in the following seeks to derive a smooth, easy comprehensible, and stable curve from noisy and error-prone data. For a selected interval of time, e.g. one minute, we derive an abstraction representing the values within this interval. Moving along the time axis we shift this interval (*time window*) to receive continuous abstractions of the curve.

So for example, if we consider a time window of one minute and a step width of one second, we do not calculate only one value per minute but for every second in the whole period of measurement we calculate an abstraction within that time window. For each time window a linear regression model is calculated as explained below. Figure 1 shows the abstraction within one time window and the moving of this window.

Given the fact that not all deviations of data points from the main line can be considered negligible although many certainly are, it is clear that any abstraction must not only provide the mean of the curve at a certain point of time but also

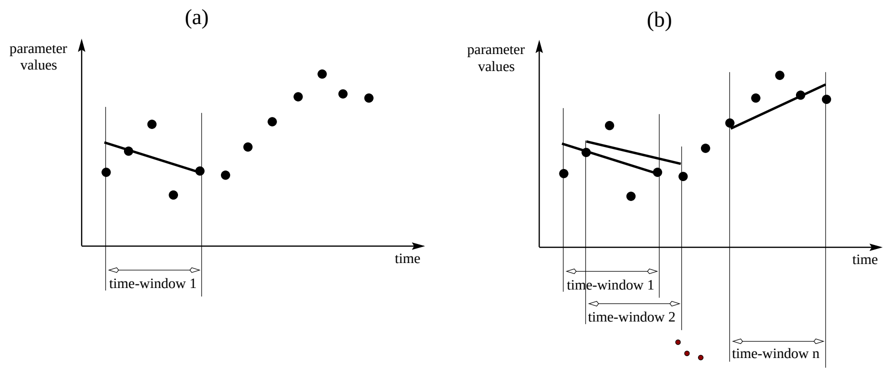


**Fig. 1.** The calculation of the linear regression is done for a time window of fixed size sliding over the entire curve in small steps. (a) shows a single time window and the line calculated from the data points within it. (b) shows a sequence of overlapping time windows and the resulting lines.

some measure for the certainty with which this abstraction can be done at that point. Such measures include standard deviation, standard error, quartiles, etc.

All these measures are only one-dimensional. Applying them on the x-coordinates of the data points would presume that the curve is horizontal. Since this rarely is the case, we first must find a "common line" of the data points in the considered interval. Only relative to that line we can define measures for the closeness of a point to the entity.

Among several candidates we chose the linear regression model as a well-proven technique for this task [1, 9]. We calculate overlapping lines in user defined steps which can be as small as a second. The length which is also user defined typically ranges from several seconds to one or two minutes. Figure 1 shows some of the lines in a close zoom.

The calculation yields not only the center of the distribution but also the inclination of the line optimally fitted through the data points (minimizing their squared deviations) and the standard deviation. The standard deviation is a very good measure of uncertainty unless some data points are missing (or removed by the error detection). Dividing the standard deviation by the square root of the number of data points used in the calculation gives the standard error. This value is preferable against the standard deviation since it grows with the decrease of valid data points reflecting thus growing uncertainty.

Plotting the standard deviation on the center of the distribution instead of the standard error gives a much wider band which exactly depicts the average distance of the data points to the line but is invariant to number of points involved in the calculation. To combine the advantages of both standard deviation and standard error, we multiply the standard error by the square root of the maximum number of data points possible within the interval of time considered and name it *adapted standard error*.

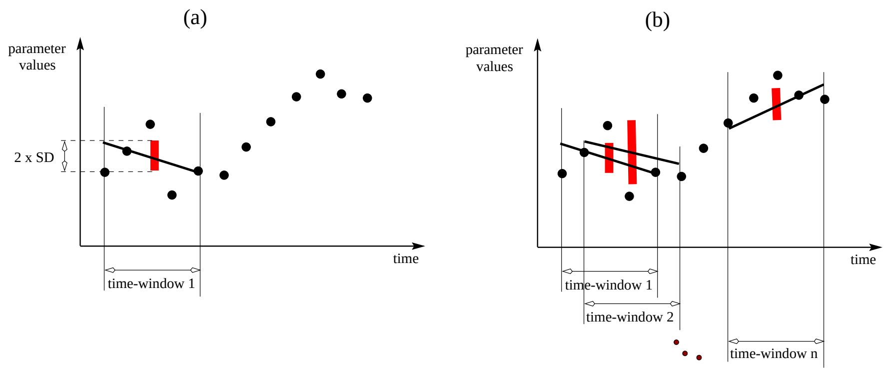


**Fig. 2.** To give an optical impression of the distribution of the data points around the regression line we vertically plot a measure for their distance like the standard deviation ($SD$) on the center of the line. (a) shows the construction of one vertical line while (b) shows a sequence of them.

Plotting the adapted standard error on the center of the distribution shows its error bar, which is a well-known means of visualizing statistical data. In the perfect case, in which all data points within the interval are valid, the width of the spread equals the (double of the) standard deviation while it will grow with an increasing number of invalid data points. Figure 2 illustrates the calculation error bars.

Connecting the upper and lower ends respectively of the error bars found for all time windows of a curve yields a band of variable width following the raw data in rather gentle bends which we call a *spread*. The narrower it is, the more concentrated the values around their mean. Figure 3 shows the final calculation of the spread.

### 3.3 Qualifying the Curve

Often the numerical value of a parameter is not itself interesting to the physician, but its qualitative abstraction like "very high" or "slightly low" or – most important – "normal". As indicated by the quotes, the exact definition of "normal" depends on the context in which the judgment is done [8].

A second characteristic of qualitative values in addition to being easy comprehensible is that they usually last for a longer time period. The resulting intervals are perceived for example as "$S_aO_2$ is high for 5 minutes". This implies that any short oscillation of the qualitative description must be avoided. While raw data typically oscillate and thus are not usable as a basis for finding reasonably long intervals wherein a qualitative value stays stable, the spread calculated above is a good ground to start at.

Figure 4 shows an example of a spread crossing borders only according to the overall tendency of the curve, skipping short-term peaks. Notice that nothing happens as long as only one margin of the spread crosses a border. Only when

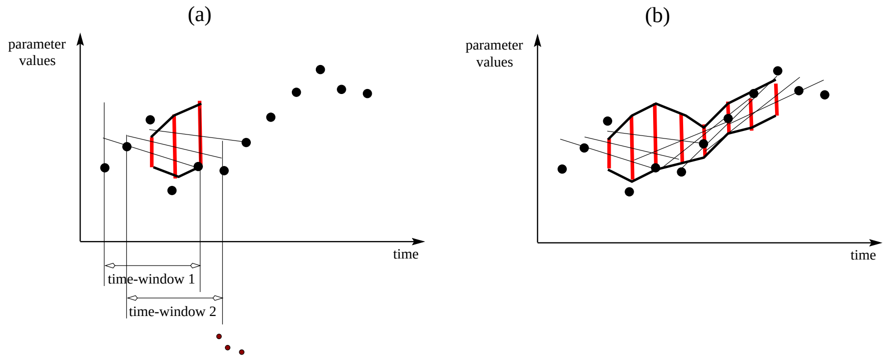


**Fig. 3.** Connecting all upper and lower ends respectively of the vertical lines (a) gives the upper and lower limits of a region containing most of the data points. The polygon constituted by these lines (b) gives an intuitive impression of the underlying data. The wider the band, the more uncertainty is involved in the calculation of the vertical position of the band representing the mean.

the other margin follows, the qualitative value changes. The spread is the wider, the more uncertainty is involved in calculating the mean of the time window. Thus changes to another region are less likely there. In contrast, a narrow spread strongly enforces the qualitative values to closely follow the quantitative value as it represents a sample of dense, noise free data points.

Retrospective analysis of data allows to select the time point at which the qualitative value changes. One may chose the intersection of the margin first crossing the border with that border, the intersection of the second margin with the border or the middle in between these points. In the example in figure 4 we have chosen the middle.

Another reasonable point to set the event of change is the intersection of the middle of the spread (in the value axis) – representing the average of its surrounding – with the border. Unfortunately, the middle can have several intersections with the border during the interval in question. So we need a rule which intersection should be taken: the first, the last or the middle between the first and the last intersection.

## 4 Parameters in the Abstraction Process

In the following we discuss the influence of different parameters on the abstraction process.

**Length of Time Window.** The length of the time window is the most influential parameter. It drives the amount of change of the curve which goes into the abstracted spread. The longer the considered interval, the smaller the influence of insular peaks in the raw data. If you want to get the overall estimation of a minute, the length of the time windows will be 60 seconds. If

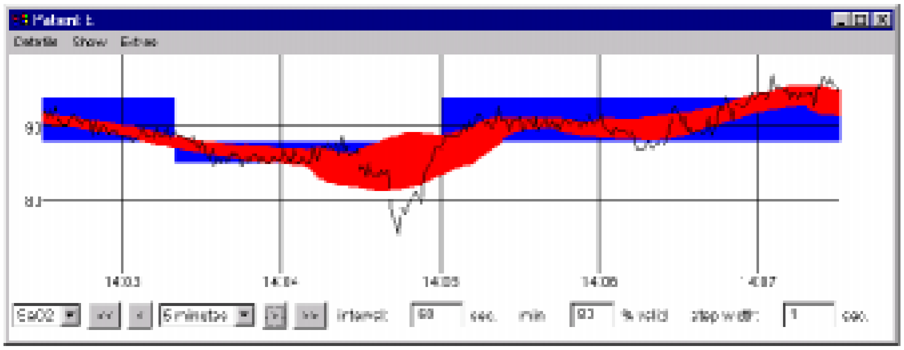


**Fig. 4.** The thin line shows the raw data. The red (light gray) area depict the *spread*, the blue (dark gray) rectangles represent the derived temporal intervals of steady qualitative values. Increased oscillation leads to increased width of the spread but not to a change of the qualitative value. The lower part of the screen shot shows the parameters used.

a decrease during 5 seconds is considered significant, the length should not be much longer than 5 seconds.

**Permitted Gaps.** In real-world situations there is always a certain amount of data points which are missing or get discarded by the error detection performed in step 1. If the amount of such points becomes too large, the linear regression calculated from the remaining points might not be too reliable and it should be visualized clearly that there is no usable input at that point. Such a situation happens frequently in daily practice.

To handle such situations we define both a maximum duration of a single gap in error-free input data, and a minimum percentage of valid data points within the time interval in which the linear regression is computed. If a gap in the (error-free) input data exceeds the maximum allowed duration, this gap is propagated through all levels of abstraction and cannot be closed by higher level abstractions. If the number of valid data points in one time window does not reach the required limit, the calculation of the regression line is skipped. Since the lines usually overlap, a gap only appears if the ends of neighboring lines do not touch. Still the reduced number of data points is visible because it leads to an increase in the adapted standard error and thus in the width of the spread.

**Point of Changing the Qualitative Value.** As described above the point of time at which the qualitative value of a parameter is changed can be set at will within the period in which the spread intersects the border between the old and the new qualitative value in retrospective analysis.

The point of change can either relate to the margins or to the middle (on the value axis) of the spread. Since the middle can intersect the border several times during the period between the first margin crossing the border and the second margin following, we can generally only speak of an interval between the first and the last crossing of the border by the middle of the spread.

For both intervals – defined by the intersections of the margins and by intersections of the middle – the beginning, end, and middle of the interval are possible choices. Among the more plausible ones are the last intersection of the middle of the spread and the middle between the first and the last intersection of the middle with the border.

**Step Width.** While as a default the algorithm calculates one linear regression within the defined time window for every data point measured, under many circumstances this can mean a lot of unnecessary computation. E.g. if the length of the considered time window is one minute, a step width of 10 seconds will still yield a smooth curve. This example shows that for best results the step width should always be some fraction of the length of the considered time window.

**Position of the Error Bar.** In the above text we silently presumed that it would be most suitable to visualize the entity of the data points in the considered by a vertical bar in the middle of the line produced by the linear regression. This means that one time window of the spread represents $x/2$ data points before and $x/2$ data points after the position of the time window where $x$ is the number of data points involved in the calculation of the time window.

While this symmetrically smoothes out disturbances in the curve in retrospective analysis, it does not properly reflect the situation of on-line monitoring where the values before the actual point of time are not available of course . In such a situation one would only consider the past and deduct only from it – the left-hand side of the curve – some abstraction of the data at the current time point. While the appearance of the spread shows some difference between these two modes of visualization, differences in the qualitative intervals abstracted from the two variants are rare.

## 5 Discussion and Further Development

Abstracting raw data to spreads and deducting intervals, in which qualitative descriptions hold, is an important step toward better visualization and comprehension of high-frequency data. The output of the algorithm presented can be used for three distinct though related purposes.

1. **Visualization of quantitative data.** While the raw data when plotted "as it is" are rather confusing, the spread gives an intuitive impression of the data by showing both the value - by its position - and the amount of uncertainty in that value - by its width. It is thus a useful tool for visualizing the quantitative input itself.
2. **Abstracting qualitative descriptions over time intervals.** Based on the need of the practitioners we display the data as a sequence of intervals, during which the values of a parameter take one qualitative value (e.g. high).
3. **Finding suitable therapeutic actions.** The qualitative descriptions are a solid basis for recommending changes of the ventilator settings and for

intelligent alarming. These are nontrivial tasks and need a knowledge base with sophisticated temporal inference capabilities.

Our future efforts will be dedicated to the integration of the visualization tool into the bigger context of a knowledge-based system using the Asgaard framework [10] for temporal planning and developing our tool from a retrospective analyzing tool towards an on-line monitoring and alarming system.

**Acknowledgments.** We thank Klaus Hammermüller, Robert Kosara and Georg Duftschmid for their useful comments. This project is supported by "Fonds zur Förderung der wissenschaftlichen Forschung - FWF" (Austrian Science Foundation), P12797-INF. We greatly appreciate the support given to the Austrian Research Institute of Artificial Intelligence (ÖFAI) by the Austrian Federal Ministry of Science and Transport, Vienna.

## References

1. R.K. Avent and J.D. Charlton. A critical review of trend-detection methologies for biomedical monitoring systems. *Critical Reviews in Biomedical Engineering*, 17(6):621–659, 1990.
2. I. J. Haimowitz and I. S. Kohane. Managing temporal worlds for medical tread diagnosis. *Artificial Intelligence in Medicine, Special Issue Temporal Reasoning in Medicine*, 8(3):299–321, 1996.
3. W. Horn, S. Miksch, G. Egghart, C. Popow, and F. Paky. Effective data validation of high-frequency data: Time-point-, time-interval-, and trend-based methods. *Computer in Biology and Medicine, Special Issue: Time-Oriented Systems in Medicine*, 27(5):389–409, 1997.
4. J. Hunter. Knowledge-based interpretation of time series data from the neonatal ICU. presentation, 1998.
5. Goldsmith J.P. and Karotkin E.H. *Assisted Ventilation of the Neonates*. Saunders, Philadelphia, 1996.
6. E. T. Keravnou. Temporal abstraction of medical data: Deriving periodicity. In N. Lavrac, et. al., editors, *Intelligent Data Analysis in Medicine and Pharmacology*, pages 61–79. Kluwer Academic Publisher, Boston, 1997.
7. C. Larizza, R. Bellazzi, and A. Riva. Temporal abstractions for diabetic patients management. In *Proceedings of the Artificial Intelligence in Medicine, 6th Conference on Artificial Intelligence in Medicine Europe (AIME-97)*, pages 319–30, Berlin, 1997. Springer.
8. S. Miksch, W. Horn, C. Popow, and F. Paky. Utilizing temporal data abstraction for data validation and therapy planning for artificially ventilated newborn infants. *Artificial Intelligence in Medicine*, 8(6):543–576, 1996.
9. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipies in C.* Cambridge University Press, Cambridge, 1992.
10. Y. Shahar, S. Miksch, and P. Johnson. The Asgaard Project: A task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine*, 14:29–51, 1998.
11. Y. Shahar and M. A. Musen. Knowledge-based temporal abstraction in clinical domains. *Artificial Intelligence in Medicine, Special Issue Temporal Reasoning in Medicine*, 8(3):267–98, 1996.