Using cardiovascular measures for adaptive automation

Stuiver, Arjan

CHAPTER 2 - **METHODS**

This chapter has been published as Stuiver, A., and Mulder, L.J.M. *Artefact-free real-time computation of cardiovascular measures*. Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. IEEE, 2009.

Abstract

Cardiovascular measures can be used as indices of physiological and affective state. The theoretical and practical issues encountered in the sequence of data-acquisition, artefact handling and data (pre-) processing are described in this paper. The results of these processes are used to formulate suggestions how to develop an operator status model, a generic system for operator assessment.

Introduction

Researchers have used psychophysiological measures as indices of physiological state, affective state and mental effort for a long time. In recent years effort has been put into developing systems that can calculate such indices in real-time. The motivation for this work has often been to use the knowledge about a person's affective or psychophysiological state to adapt the environment or the current task demands to better fit his or her needs. This is indicated by the term adaptive automation (Haas and Hettinger, 2001).

One group of psychophysiological measures that can form a basis for adaptive automation has been cardiovascular measures. This group consists of the measures based on the interval time between successive heartbeats, such as heart rate and heart rate variability, as well as blood pressure, blood pressure variability and respiration measures. For example, when mental workload increases, heart rate variability usually decreases, in general caused by more invested mental effort. Heart rate, blood pressure and respiration frequency normally increase with a rise in mental workload. Also differences in mood have their effects on cardiovascular measures, although these are not fully described in literature (Thayer, 1989). Using other physiological measures to indicate affective state seems possible but has not reached maturity yet (Wang et al., 2003).

Different methods can be applied to construct indices from physiological input data. Our current approach is based on the COMPANION concept (Mulder et al., 2008). The goal of COMPANION was to develop better co-operation between a man-machine interface and the human operator, based on the human's (physiological) state. Our model developed from the COMPANION concept is called the Operator Status Model (OSM). Within the OSM, data acquisition, pre-processing, analysis and status decision-making are implemented in separate modules. These modules can form a basis for a system for adaptive automation.

During each phase in the computational process of deriving indices of operator state from cardiovascular measures some theoretical and practical problems must be solved. The different phases are represented in the different modules of the OSM. Our solutions to the problems at each level, including the arguments to use them, and some of the technical details to do so, are described in this paper.

Applied Methods

Data acquisition

Not only physiological sources can be used to construct indices for mental or affective state. For example, in a driving situation vehicle parameters, such as swerving behaviour, can be used besides cardiovascular measures. This means that several devices or computers dealing with different sources may send data to the OSM, which is realized by using Ethernet. It also means that specialized software must be used or written that can send data from acquisition devices or computers over the network in an appropriate format.

By using a general protocol and a general data format it is possible to send not only cardiovascular data, such as electrocardiogram (ECG), respiratory signal (RESP) and blood pressure (BP), but also other physiological and non-physiological data in the same manner.

Pre-processing cardiovascular data

For detecting the R-wave in the ECG signal some acquisition devices have a hardware detection function, while others use a software R-peak trigger. For full flexibility and ease of data communication between the acquisition software and the OSM, a software R-peak trigger is incorporated in our software.

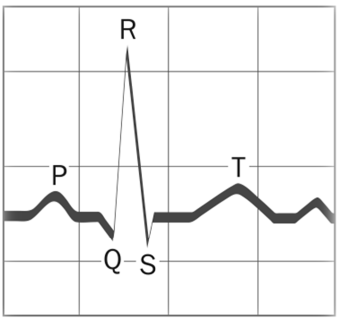

Figure 2: The PQRS-complex in the ECG.

The heartbeat detection algorithm creates a non-equidistant signal that contains only data points for every heartbeat occurrence time in the electrocardiogram. These data points are located at the R-peak in the PQRS-complex of the electrocardiogram, depicted in Fig. 1. The detection algorithm is therefore often called the R-peak trigger.

The basis for the R-peak trigger used in the OSM is level-triggering followed by maximum detection, but is extended to make the detection more accurate and robust. When the amplitude of the electrocardiogram rises above 10 percent of the preceding 200 ms of data, a PQRS-complex is identified. The maximum of the PQRS-complex is classified as an R-peak by looking where the gradient of the electrocardiogram changes from positive to negative. The new R-peak is classified as a heartbeat and added to the data if its amplitude is at least 70 percent of the mean amplitude of the three preceding (correct) R-peaks. To prevent triggering on the T-wave that immediately follows the PQRS-complex and that can have considerable amplitude, an inhibition period of 250 ms is used after triggering a (correct) R-peak.

An accuracy of at least 2-4 ms is required for the resolution of the non-equidistant signal containing the R-peaks (Berntson et al., 1997). This corresponds to a sample frequency of at least 250 Hz in the original electrocardiogram. This accuracy is needed for the variability measures that are computed from the R-peaks. If the accuracy is too low the variability measures will not be sensitive enough. Knowing that the estimation errors have a flat distribution, the variance that is introduced by a resolution of 4 ms can be calculated by: ½ * (4 ms) 2 = 8 ms2. In most workload studies, the level of variance found in cardiovascular variability measures ranges from 200 to 10.000 ms2 (Mulder, 1992).

The non-equidistant series of R-peak data is used as a time-basis for blood pressure and respiration in a beat-to-beat approach. The values of mean, diastolic and systolic blood pressure and mean respiration in the intervals between two R-peaks are calculated and added to the R-peak data at the time of the second R-peak of the interval. The obtained data set can then be considered a non-equidistant sampled representation of the cardiovascular data in a beat-to-beat fashion (Mulder, 1992). In this way the different signals all have a similar format and can be used in the same routines.

Pre-processing respiratory data

Respiration is usually measured by using a flexible belt around the chest. Measuring principles vary from changes in belt-resistance to changes in induction, applying the so-called respitrace principle (Respitrace Inc.). Being less sensitive to movement artefacts, this latter principle is preferred above resistance measurements. The amplitude of the respiration signal is dependent on the placement of the equipment and body size. The respiration signal must therefore be calibrated for each person to make comparison between people possible. The standard procedure that can be found in the Respitrace manual is to have the subjects breathe in a plastic (spiro) bag with known volume (Respitrace Inc.).

Backs, Ryan and Wilson (1994) showed that respiration rate and depth can have different diagnostic power. The amplitude of the respiration signal can be used as a measure for respiration depth.

To identify breathing pattern and respiration rate from the respiration signal, a software respiration detection algorithm has been developed. The respiration trigger identifies the start of each breath and creates a non-equidistant signal containing breath inspiration times. Firstly, extremely high amplitudes of the respiration signal are cut off at about two times the mean amplitude of the data of the preceding 60 seconds. Secondly, a moving average low pass filter will smooth the data. The low pass filtered signal is a smooth version of the respiration signal that should only show maxima at exhalation-times and minima at inhalation-times. The interval between two inhalation-times is identified as a respiration cycle if the amplitude of the current respiration-cycle reaches at least 30 percent of the previous one.

The respiration period can be computed by taking the intervals between inspiration times. With zero-order interpolation, values for the respiration period are calculated for every R-peak time. This results in a signal that uses the last known value for respiration period for every R-peak, until a new respiration period value is available at the end of the next respiration cycle.

Artefact handling

The reliability of all physiological signals, and the derived measures, depends strongly on the quality of the recordings. The recording procedure, but also physiological sources, can introduce artefacts in the data. Erroneous data will result in unreliable measures. As such, the entire performance of an operator status assessment system depends on the quality of the input data. As a rule of thumb, one missing beat in 100 seconds of heart rate data accounts for the same amount of variability than can be found in 100 seconds of error-free data (Mulder, 1992). Therefore investing effort into an error-free measurement procedure and well-functioning equipment is obviously important.

Artefact correction is also very important for those artefacts that will still arise, if we want an operator state assessment system that functions properly. A distinction can be made between different types of artefacts, those caused by measuring and recording and those originating from the physiological system. Both types can induce problems for further analysis and use of the data. Although the non-physiological artefacts must always be corrected, this is not so straightforward for artefacts having a physiological background.

The most common artefacts in heart rate data are ectopic beats, a form of cardiac arrhythmia. These artefacts can be divided in several types: 1. a heartbeat with a very short preceding interval followed by a heartbeat with a very long interval; 2. a heartbeat with a very short interval followed by another very short interval, often caused by an extra systole; 3. a heartbeat with a very short preceding interval followed by heartbeats with normal intervals, causing sort of a phase shift in the data (Mulder, 1992).

Artefact detection is more or less similar for all types of signals in our software. For a certain amount of data, for example a time window of 30 seconds, standard deviation (SD) of the time series is calculated. The current value is then compared with the mean plus or minus, for instance, four times the SD. How many times SD must be used depends on the type of data. If the value differs more than this criterion, the value is marked as an erroneous value. Another check is done by comparing successive differences between values, using the same approach of computing the SD but then from successive differences. If the current difference deviates more than a number of times the SD from the mean of the previous series of successive differences, the current value is marked as erroneous.

After detecting and marking the erroneous values, these values will be corrected, if correction is possible and only if it improves the data. Correction is done by removing the errors and filling the resulting gap by first-order linear interpolation between the two preceding and the two succeeding correct values (Mulder, 1992).

When many errors are found in a data sequence, there is probably something more complex happening than just an artefact. It can be a drift of the signal but there may also have been a large change in the task that causes a disturbance in the data, for instance starting to run after sitting for some time. Correction is often not appropriate in these cases. In the OSM, if more than 10 seconds of data is marked as erroneous, correction is stopped. The resulting problem should then be solved at the next higher analysis level, for example by keeping track of the reliability of the data as a function of time.

The difficulty of artefact detection ranges from fairly easy to very hard. Also the necessity of correction varies enormously over subjects. Some registrations show only a couple of artefacts in an hour of registration, while the data of other subjects can be useless because of hundreds of very specific artefacts, mostly related to ectopic beats. Many artefacts are easy recognizable for the detection algorithm and the interpolation gives excellent results. Other artefacts are more difficult to classify and are sometimes only interpretable for a human eye, which is why in offline analysis the automatic correction is always followed by a visual inspection. In real-time computation visual inspection is unfortunately not possible. A solution could be to keep track of the reliability of the data by constructing a confidence level of artefact detection and correction of the data. In the next processing phase the reliability or the unreliability of the data can be taken into account, for example by using the level of reliability to decide how much a measure contributes in a decision formula.

For correction of heart rate (R-peak) data a procedure is included that adds a small amount of noise to the interpolated values, related to preceding variability levels. Linear interpolated data has less variability than normal error-free data. By adding a small amount of noise, related to preceding variability level, introduction of a decrease in variability by the linear interpolation procedure is minimized.

Time domain measures

For analysis it will be beneficial to change the cardiovascular time series, which are recorded on a beat-to-beat basis, into stable and reliable profiles that no longer have the difficult property of being non-equidistant. This can be accomplished by calculating mean values of the obtained time series in a moving time window shifted at small intervals over the data, called profiles analysis. In the OSM implementation a 30 second window is used that is shifted by one second for every calculation-cycle. This means that the first 30 seconds give a mean value for a new data point at exactly 30 seconds. The window is then shifted one second and a new mean value is calculated for the next data point at 31 seconds. This results in a new signal (profile) with a sample frequency of exactly 1 Hz, the step size of the moving window. Using such a high 'sample rate' (compared to the length of the analysis window) gives the flexibility to select data segments from more precise periods of time for the next analysis step. It must be understood, however, that dependency between samples in the new profile data is fairly large. Every new sample will contain only 1/30 of new data. So, except for very large and steep changes the profile data will change only gradually, in relation to the new 1 second sample interval.

Physiological measures such as heart rate and blood pressure beat-to-beat values (systolic, diastolic, mean, pulse pressure) can be used in this procedure, resulting in mean heart rate and mean blood pressure values respectively. The same procedure, however, can also be used for most non-physiological measures in a similar way, resulting in data that can easily be combined with the physiological data. When using a driving simulator for example, performance measures can be calculated from swerving behaviour and integrated into the procedure with ease.

The procedure is of course not restricted to computation of mean values in the time window. Other possibilities are standard deviations or variation coefficients which might be more useful in some cases.

Spectral measures

Spectral analysis provides a means for separating sources of variance in heart rate and blood pressure, if the sources have different frequencies. For example, the influence of parasympathetic and sympathetic control can be partly separated (Akselrod et al., 1985).

Spectral analysis can be done in a way that is very comparable to the time domain analysis. Now instead of calculating the mean or standard deviation in a given time window, as a first step, a power density spectrum is calculated over that window. By integration of the spectrum between pre-defined frequency points it is possible to compute the variability in certain frequency bands. In this way, variability in the mid-frequency band (0.07-0.14Hz), associated with short-term blood pressure control, and the high-frequency band (0.15-0.40Hz), associated with respiration sinus arrhythmia, can separately be studied (Berntson et al., 1997).

The different frequency bands can give different information about operator status. In some studies the information in the high and mid frequency bands is similar, while other studies show interesting differences, related to the type of activity or the subjects' (mood) state. The mid-frequency range has been used in many studies as an index of mental effort (Mulder, 1992). The high frequency band has been used as an index of vagal inhibition (Grossman and Taylor, 2007).

A spectral profile can be created by shifting the analysis window at one second steps to compute the band values, according to exactly the same procedure used to create the time domain measures. This is done with spectral bands instead of mean values, resulting in a signal with a similar format, called spectral profiles.

The respiration period, described earlier, can be used as a measure by itself, but can also be used to define a variable respiration frequency band. As stated above the high frequency band is assumed to contain effects of respiration. The actual respiration frequency, however, is not always contained within the boundaries of the respiration frequency band. A new frequency band can be defined around the actual respiration frequency, resulting in variable respiration frequency band. The respiration rate, the inverse of the mean respiration period in the actual analysis window, is calculated as a first step. A frequency band can then be defined from 0.03 Hz below to 0.03 Hz above this respiration rate. Grossman and Taylor (2007) show the additional value of using specific respiration-area related cardiovascular activity instead of using information of the much broader high frequency band

Transfer functions can be computed using the same short-term approach with small analysis segments. This provides means to study the relation between blood pressure and heart rate, for example. The spectral transfer function (gain, phase shift and coherence) between blood pressure and heart rate shows the dependency of heart rate variability on blood pressure changes. The gain function of this relation in the mid frequency range is used as an estimate of baroreflex sensitivity (Robbe et al., 1987).

The spectral routines used in the OSM are based on the spectrum of counts as described by Rompelman (1985), a form of a Discrete Fourier Transform (DFT). In the spectrum of counts a delta pulse is placed at every R-peak occurrence time. The time information of each delta pulse, indicating the variations in R-R (interbeat) interval is inherently used in this algorithm. It can be shown that theoretically this spectrum is an ideal spectral representation of changing heart rate (Rompelman, 1985). Van Steenis (2002) has shown that with this method all frequency information is contained in the spectral result.

By using a DFT, spectral functions of each length can be handled without being restricted to powers of 2 or having to use zero padding to increase computational speed in FFT algorithms. The efficiency of the DFT algorithm is in practical situations sufficient, since sinus components only have to be calculated where the signal is not zero, which is only at R-peak occurrence times (Mulder, 1992).

The window length determines the frequency resolution of the spectra. For example, a 30 second window will result in a frequency resolution of 0.033 Hz. By interpolating between frequency points (using an integration algorithm) the spectral resolution can be increased to 0.01. The earlier mentioned frequency bands can be selected more accurately with a higher resolution. When the original frequency resolution of 0.033 Hz is used, irrelevant spectral power will be included or relevant spectral power will be excluded. For instance, in case of a resolution of 0.033 Hz, the best approximation of the mid frequency band would range from 0.066 Hz to 0.132 Hz, which is a bad approximation of the standard 0.07-0.14 Hz range.

The profiles analysis used in our system makes it possible to use window lengths of any size. This gives, via interpolation, a fixed frequency resolution. The same approach can be used for power spectral analysis as well as for transfer functions (gain, phase or coherence functions. Van Steenis (2002) argues that non-stationarities do not require a wavelet or a Wigner-Ville procedure; comparable results can be obtained using other algorithms such as implemented in our approach.

The described spectrum of counts and its implementation with the DFT algorithm delivers spectral measures in terms of heart rate (HR) changes (Mulder, 1992). In other procedures in literature (interpolated) interbeat interval times (IBI values) are often used as a basis instead (see Rompelman, 1985 for an overview of different methods). Users who are not familiar with the differences between these HR and IBI based measures would expect comparable variability measures between these two procedures. This relation, however, is not trivial and at least non-linear. As Mulder (1992) points out, spectral variability measures are strongly dependent on mean interbeat interval: the ratio of spectral power measures derived from HR spectra and the same measures derived from IBI spectra follow a curve with mean IBI to the power of 4. This means that either using HR or IBI measures will result in statistically different outcomes. Because this is an unwanted situation for indices of operator status we have chosen independent measures. These can be created by expressing the spectral values as a proportion of the mean, by dividing the values by the squared mean of the original signal, resulting in normalised spectral band values (Berntson et al., 1997).

The distribution of spectral power values and other variability measures is skewed, which underestimates smaller spectral values and overestimates larger ones during statistical analysis. Van Roon (1998) has shown that variability measures have a chi-square distribution and that using logarithmic values will result in a normal distribution.

Towards optimisation models

The obtained (derived) physiological measures for operator state can be the input for an optimisation algorithm that uses the different measures to make a decision on operator state. For example, the optimisation algorithm could be a mathematical combination of the (derived) physiological measures, and additional (task performance, task load) measures of other types. Multiple regression models [6, 10], fuzzy logic models and other methods have been used in this fashion. The output from such a model, a measure of operator state, can then be used by another system to make changes to the operator's environment or change the task specifications.

Discussion

A lot of work has already been done on using cardiovascular and other physiological measures as indices of affective or mental state. There have even been a few systems implemented that can assess operator status based on physiological measures, however some issues still have to be resolved in a systematic manner.

One of the differences between real-time and offline computation of indices for operator state is that in the offline case the center point of the computational window can be used with data originating from some time before the current time point and from some data after that point. In real-time this would result in a (larger) delay of the system; this is why only past data points are used. The real-time analysis is therefore slightly different from the offline analysis, which may increase complexity of comparing results. Missing sequences of data causes a difficult problem. A distinction can be made between missing values in the original data and in the computed profiles. Missing values in the original data must be corrected as much as possible, while keeping track of data-quality.

Of course, reliability will decrease when there are errors in the data, even if they are corrected. An estimate of how reliable the data still is can be constructed by keeping track of the artefacts found. If there is knowledge available from other domains about the reliability of the signal this can be used to adjust the reliability information. For example, a very irregular breathing pattern in the respiratory signal can indicate a non-stationary working situation which may cause artefacts in the heart rate data. As a next step, this information can be used in the computational process of updating profile data. If the profile data is known to be unreliable it might be wise to maintain the previous profile estimation up to the moment a new reliable data point can be derived.

Hoogeboom states that large delay times in system adaptation or changes in human machine interaction may result in unwanted surprise behavior (Hoogeboom and Mulder, 2004). For that reason it is important that a system responds with appropriate speed for the task/environment it is used in.

The adaptation speed of a system depends on how much data is needed to compute a reliable index, not whether there is enough time to finish all the required computations. If more data is used, from larger data segments, the outcomes of the (set of) measures will, in general, become more reliable. However, the response time of the system will decrease.

In our opinion, the status of the optimisation procedures used to estimate state changes on an individual basis is far from optimal in the field of adaptive automation. Researchers in the field of brain computer interfacing (BCI), for example, have developed examples of optimisation procedures that have proved themselves in laboratory settings. This has yet to be done for adaptive automation.

Conclusions

A generic approach, such as the COMPANION concept described here, offers good opportunities for integrating information from physiological and other sources. Physiological measures, such as cardiovascular measures, have shown to be indicative of affective and mental state. In many environments, task information can be used as indices of operator state as well. Brookhuis and De Waard (2001) have shown that driving parameters can be used for driver's assessment, for example.

A combination of these types of data requires a system that is less dependent on the technical tools and procedures, such as data-acquisition. Most methods described in this paper can be generalized, because the same principles can be applied for most data types. Changing all types of data into a similar format early in the pre-processing process will facilitate flexible use of the system. In this respect, the profiles analysis reported above is a step forward towards realizing a more generic approach.

An effective optimisation procedure is one of the challenges still remaining. Success will depend strongly on obtaining enough reliable individual data and on the procedures to combine that data in suitable algorithms.

Acknowledgments

Parts of this work were completed within the HILAS (Human Integration into the Lifecycle of Aviation Systems) project, a 6th Framework EU Programme (Contract No. 516181). Among many others, the contributions of Piet Hoogeboom and Arie van Roon to our research over the years in this area have been very valuable and are greatly acknowledged.