Computational Intelligence, Volume ?, Number ?, 200?

IPRA — ENHANCING THE SENSING ABILITIES OF AMBIENT INTELLIGENCE

Holger Schultheis

SFB/TR 8 Spatial Cognition, Universität Bremen, Bremen, Germany

The ability of ambient intelligence to serve a person could be considerably enhanced, if intelligent devices would take into account the person's current cognitive and affective state. One primary source for state information are physiological measures. However, obtaining state information requires concertedly employing a multitude of different techniques. To facilitate the use of physiological signals as sources of state information and thereby allow improving ambient intelligence, we developed Ipra, an integrated pattern recognition approach which (a) employs all techniques necessary to appropriately analyze physiological data, (b) is broadly applicable, and (c) is easy to use even for non-experts. This approach together with results of its first application are presented in this contribution.

Key words: psychophysiological data, user modeling, pattern recognition

1. INTRODUCTION

An intelligent device in a person's environment will be all the more of use to the person if the services it makes available to her are appropriate with respect to her current situation. For example, presenting a person with new and potentially helpful information (e.g., route advice during driving) should be dependent on (a) whether the person currently has the cognitive capacity to process the information presented and (b) whether presentation of the information might distract the person's attention and thereby be more harmful than helpful (e.g., the person might cause an accident; cf. McFarlane & Latorella, 2002). Less critical but also desirable might be a preselection of services based on the current mood of a person like an intelligent HiFi device suggesting songs or films on the basis of her current emotional state. Consequently, truly useful ambient intelligence should take into account a person's cognitive and affective states.

In order to take a person's states into account, intelligent devices in the environment need to be provided with information about these states. But such information is, unfortunately, difficult to obtain, since cognitive and affective states of a person are not directly observable. Thus, cognitive and affective states need to be inferred from indirect measures. Four classes of measures can be distinguished: *analytic measures*, *subjective measures*, *performance measures*, and *physiological measures*.

Analytic measures derive estimates of cognitive and affective states from general (i.e., not interaction-specific) knowledge about situations and persons. In a usual analytic approach, the situations of interest (e.g., driving) are analyzed with respect to their cognitive demands and emotional affordances. These demands and affordances are then related to the cognitive and emotional attributes of some normative model of that group of users (e.g., expert vs. novice drivers) encountering the relevant situation. In doing so, analytic measures allow deriving predictions for person states in certain situations and are, therefore, most valuable in developing / designing AmI devices (Tsang & Wilson, 1997). Analytic measures are, however, less suitable to obtain state information for one particular person in one particular situation, since they neglect idiosyncrasies of particular persons and situations as well as interaction effects between persons and situations (Gopher & Donchin, 1986).

Subjective measures can be obtained by directly asking a person to judge her cognitive and affective state. Although this might seem, at first sight, the optimal way of obtaining state information, this approach has, several drawbacks: first, subjective estimates may be distorted due to memory and consciousness effects (O'Donnel & Eggemeier, 1986). Second, answering state questions concurrent to some activity might be annoying to the person and in some cases even dangerous (Picard, 2000). Answering questions after the respective activity, on the other hand, might provide state information too late for adjusting the intelligent device's behavior appropriately.

Performance measures rely on the person's overt behavior to identify her states. For example, the time it takes a person to react to some route advice or the extent of deviation from the middle of the driving lane could be taken as indicators for the amount of cognitive capacities still available for activities other than driving. Apart from the fact that not all state variations—especially affective state variations—might result in change of overt behavior (O'Donnel & Eggemeier, 1986), this technique allows to measure the states of the person only as often as overt behavior can be observed (Wickens & Hollands, 2000).

Physiological measures are based on the observation that various bodily processes and conditions (e.g., heart rate, skin conductance, brain activity; see Wilson & Eggemeier, 1991, for an overview) covary with changes in cognitive and affective states. Thus, monitoring of these body functions allows cognitive and affective states to be inferred. Compared to the other three types of measures physiological measures have the advantage to (i) be continuously available even without overt behavior of the person or the person answering questions, (ii) be independent of the person's memory or consciousness, (iii) usually change quickly (in the ms range) when cognitive or affective states change (again, independent of any overt behavior of the person), (iv) be quite sensitive, that is, even small changes in the person's states show up in changes of physiological signals, and (v) be immediately available even in situations not foreseen by a device's designer.

Thus, physiological measures exhibit a combination of properties which renders physiological signals highly valuable for sensing a person's cognitive and affective states. This is not to say that the other types of measures cannot be of use in inferring a person's states. Yet, given the advantages of physiological measures it seems desirable to not restrict ambient intelligence to the employment of analytic, subjective, and performance measures, but to complement these information sources by physiological data. Ideally, inference of a users state is based on many state indicators which are combined by information fusion approaches employing, for instance, dynamic Bayesian nets (Brandherm & Jameson, 2004).

Notwithstanding their advantages, physiological signals have one major drawback which often precludes their employment and impedes exploitation of their potential benefits for ambient intelligence: Due to their characteristics (see Sect. 2), it is quite hard and requires considerable expertise to reliably and correctly extract the information about a person's states from psychophysiological measurements. More precisely, to infer the states from the raw signals, a battery of machine learning techniques (see Sect. 3) needs to be applied. Because of this necessity for sophisticated data analyses, ambient intelligence researchers lacking the relevant machine learning expertise are usually barred from utilizing physiological signals.

Given the considerable potential physiological signals hold for ambient intelligence in providing information about a person's state, this seems to be a highly unsatisfactory situation. Desirable would be a machine learning approach and / or tool which both draws on the full battery of techniques required to properly analyze physiological signals and is easy to use, so that researchers wanting to endow their intelligent devices with the ability of sensing cognitive and affective states need not be(come) experts in analyzing the signals. Ideally, such an approach should also be general enough to be applicable to a wide range of physiological data. In this contribution we present such an approach, called IPRA.

The remainder of the article is structured as follows: first, the properties of physiological data necessitating the use of sophisticated techniques are shortly described (Sect. 2). Second, the general principles underlying the design of IPRA (Sect. 3.1), some necessary steps to preprocess given psychophysiological data (Sect. 3.2), and the specific methods employed in Ipra (Sect. 3.3) are expounded in Section 3. Having introduced Ipra we will present its evaluation in detailing analyses for two physiological data sets (Sect. 4) as well as discussing Ipra's suitability for ambient intelligence against the background of the analyses' results (Sect. 5). Finally, we will shortly highlight related work (Sect. 6) before some concluding remarks touch on open questions and future work (Sect. 7).

2. PROPERTIES OF PHYSIOLOGICAL DATA

Body conditions and processes are continuously available and measurement techniques allow to gauge the value of a physiological signal several hundred times a second. Furthermore, a multitude of different physiological signals (heart rate, positron emission tomography, skin conductance, electromyography of various muscles, etc.; see e.g. Stemmler et al., 2001; Kramer, 1991) have been proposed and are potentially available for inferring a person's states. As a consequence, the overall physiological data gathered within even short time spans can easily comprise more than 100 or actually several thousand values. If, for example, a person's states should be estimated every second, and a single physiological signal is sampled at 500 Hz, each measurement will already comprise 500 values.

To be able to estimate the current state of a person from such a 500 value vector one has to know how particular values of a measurement relate to certain states. Due to the large number of values, it is usually not obvious and for humans—even after long and thorough inspection—impossible to determine which combinations of values are characteristic for specific states. Besides the large number of values to consider, identifying the relation between values and states is further exacerbated by the fact that other factors than the current state of the person, called *noise*, influence the measured values. For one, bodily processes do not only change with changing cognitive or affective state. Heart rate, for instance, depends considerably stronger on bodily than on mental effort. Furthermore, physiological signals do vary seemingly random over time, that is, they vary even without recognizable external influences.

The large number of values to be considered together with the noise in the signal normally makes it impossible to identify the relationship between measurement values and a person's state with the naked eye. Therefore, pattern recognition techniques need to be applied to extract the relationship between the values of a measurement and the state the person was in when the measurement was taken. As will be explained in the following sections, however, applying pattern recognition is not as easy as it may sound. It requires careful consideration and preparation utilizing other techniques like feature selection and estimating intrinsic dimensionality. In other words, pattern recognition needs to be integrated with these additional techniques to yield an appropriate approach to the analysis of physiological data for ambient intelligence. The general principles and specific methods of Ipra which constitute such an integrated approach will be detailed in the following section.

3. IPRA — AN INTEGRATED APPROACH

The aims in developing Ipra were to provide:

1. an integrated approach: utilizing a comprehensive set of methods to ensure a satisfactory analysis of physiological data.
2. a broadly applicable approach: utilizing methods and procedures such that a wide range of different kinds of physiological data can be successfully analyzed with Ipra.

3. an easy to use approach: even non-experts should be able to use IPRA to extract the relevant state information from given physiological data.

The general principles and specific methods employed to achieve these aims are detailed in the next three sections.

3.1. General Principles

Aim 1: Integration. As already said, the particular nature of physiological data (Sect. 2) necessitates the use of pattern recognition techniques to extract relations between measured values and a person's states. More precisely, statistical pattern recognition (cf. Jain et al., 2000) seems most appropriate due to both (a) the lack of prior information about structure and (b) the noise in the data. The basic idea of statistical methods is to view the measurements as points in an n-dimensional space, where n is the number of values in each measurement, and to assume that measurements taken during one state of the person are distributed differently in this n-dimensional space than measurements taken during another state of the person. The aim of statistical pattern recognition is then to either directly or indirectly estimate those ($n-1$-dimensional) decision boundaries which minimize the probability to assign a measurement to a wrong state, that is, the probability to infer an incorrect state from a given measurement.

Trunk (1979) has shown that when estimating the decision boundaries the estimate will so much more be inaccurate as the number of dimensions will increase—a phenomenon termed the *curse of dimensionality.* More precisely, the accuracy of the estimation depends on the proportion of dimensions and number of examples from which to estimate the boundaries. Jain et al. (2000), for example, argue that one needs at least ten times as much measurements for each state to be distinguished than each measurement has dimensions.

If this ratio of measurements to dimensions is not given, one is faced with the dilemma (a) to preferably not neglect any important information, that is, to use as many dimensions as possible, but (b) to restrict the number of dimensions used to avoid severe false estimations of the decision boundaries. A possible escape from the dilemma, in principle, would be to increase the number of measurements. Yet, gathering arbitrarily many measurements can be infeasible and, more importantly, might be impractical. Especially in the scope of employing physiological signals (which may have several thousand dimensions) for adaptation in ambient intelligence, obtaining enough measurements to avoid the curse of dimensionality will be difficult if not impossible. As mentioned above, the main advantage of information about a person's state is the possibility to tailor the behavior of the ambient intelligence devices to this state. To do this on the basis of physiological data, existing classifiers (i.e., estimated decision boundaries) are necessary. As long as the classifiers are not available the user state cannot be determined and adaptation cannot take place. First gathering several thousand measurements to be able to estimate the decision boundaries using all dimensions would usually require to gather measurements for several days. That is, a person would have to wait for several days before the ambient intelligence devices could adapt to her state and, more importantly, would be required to keep a detailed diary of her varying emotional and cognitive states over the days. Consequently, with respect to user acceptance of ambient intelligence it will be necessary to shorten the phase of initial acquisition of physiological data. Thus, the number of measurements available will typically not be enough to avoid the curse of dimensionality.

An alternative to increasing the number of measurements is to decrease the number of dimensions. There are basically two types of methods available to do this: *feature extraction* and *feature selection.* Whereas feature extraction yields a reduced number of new dimensions by transforming and combining the original dimensions, feature selection reduces the number of dimensions by selecting a subset of the original dimensions. Given the aims in developing Ipra, the use of feature selection seemed to be more appropriate than feature extraction: first, feature extraction methods are often specific to particular physiological signals. For example, to determine various characteristics of heart activity determining the so called R *peak* in the raw signal is necessary (Kramer, 1991), whereas the brain source of certain parts of the raw signal is often extracted from electroencephalogram (EEG) data (Delorme & Makeig, 2004). Feature selection on the other hand is not dependent on the type of data analyzed. Thus, feature selection can be applied without modifications to all types of physiological signals which is more in accord with the second aim of Ipra of being broadly applicable. Second, feature extraction in and of itself is only instrumental if the extracted features replace the original ones. This, however, bears the danger of disposing original dimensions which might have been very informative—a particular problem in cases where the information of the relevant original dimension is not well represented in the extracted features. Against this background already Stearns (1976) proposed that feature extraction is important and helpful only as a first step. He advocated that after new dimensions have been extracted a suitable subset of dimensions should be selected from the combined set of all new and original dimensions. Consequently, feature extraction is more like a potentially advantageous extension than like an appropriate replacement for feature selection. Third, feature selection is preferable to feature extraction for practical reasons. As already said, extracted dimensions result from transforming and combining original dimensions. This entails that to compute the extracted dimensions in an ambient intelligence setting, all or most of the original dimensions have to be measured to be able to infer the state of a person. The result of feature selection on the contrary is a subset of the original dimensions. In particular, dimensions related to some of the originally employed physiological signals might not be part of that subset and thus need not be measured when inferring the person's states. Put differently, feature selection has the potential to reduce the number physiological signals that have to be measured and, therefore, can help to increase user acceptance (because less sensors need to be employed) and technical as well as computational complexity.

Against the background of these considerations we chose to employ feature selection in Ipra. Consequently, Ipra tries to achieve satisfactory decision boundary estimates by using as few, say k, dimensions as possible such that the dimensions used contain the maximal information for a subset of dimensions of cardinality k.

One way of finding this subset is to test every subset of dimensions for its suitability. However, considering each subset is only seldom possible, since the number of subsets is exponentially related to the number of dimensions. For physiological measurements in particular following this approach would include examining 2^{500} subsets or more which is clearly infeasible. Rather, the search space of dimension subsets needs to be searched using heuristics trying to find the best subset while visiting as few subsets as possible. Various such methods have been proposed over the years (see Liu & Motoda, 1998 for an overview). Although the different approaches differ quite a lot in how they attempt to search the subset space, most of them require the user to specify the maximal (and sometimes the minimal) cardinality of the subsets to examine. To constrain the search as strictly as possible it would be desirable to provide as the maximal cardinality a number which corresponds to the minimum number of dimensions needed to distinguish measurements belonging to different states.

An indicator for the minimum number of dimensions necessary is the *intrinsic dimensionality* of the given example data. Given a data set DS lying in a space S with d dimensions, the intrinsic dimensionality of DS is defined to be $d', d' \leq d$, such that there exists an S' with $S' \subseteq S$, S' has dimensionality d', and all elements of DS lie completely in S'. Usually, and especially for high dimensional, noisy data, d' cannot be determined exactly, but has to be estimated. But even if estimated d' yields important information for the pattern recognition process. If $d' < d$ holds, this indicates that some of the dimensions used to characterize the elements in DS are superfluous. This entails that some of the dimensions used in DS will be of no use in distinguishing measurements stemming from different states. Therefore, d' can be seen as a measure for the amount of dimensions necessary to retain all the information in a data set which are relevant for distinguishing between different states.

Summarizing the above, there are at least three steps necessary to appropriately analyze physiological signals: first, the intrinsic dimensionality of the given data set has to be assessed. Second, with the estimated intrinsic dimensionality as a guiding value for the cardinality, a subset of all available dimensions will be selected to avoid the curse of dimensionality. Finally, pattern recognition techniques can be applied to the data constrained to the dimensions found by the feature selection stage to estimate the optimal decision boundaries. By implementing this three step procedure Ipra realizes the first of the above stated aims of providing an integrated approach to the analysis of physiological data.

Aim 2: Broadly Applicable. To realize also the second aim, the methods employed in each of the three steps had to be chosen carefully. Over the years numerous methods for estimating intrinsic dimensional, feature selection, and pattern recognition have been proposed. Each particular method in each stage has its particular strengths and weaknesses and can be assumed to provide good or optimal results only for certain kinds of to be analyzed data. Regarding pattern recognition techniques, for example, Schaffer (1994) has shown that no single technique is superior in performance to any other technique when considered across all possible types of problems. This does not exclude one pattern recognition technique to be superior to other techniques for certain problem types. Rather, it shows that techniques superior to others for some problem type must be inferior to the other techniques when applied to a different problem type. Consequently, which technique is optimal is dependent on the type of problem considered. Since one usually cannot judge from the given data set to which type of problem it corresponds, it is impossible to know prior to analyzing the data which pattern recognition technique will give the best results. This leads to two important conclusions with respect to the second aim. On the one hand, to be broadly applicable a pattern recognition approach should not rely on one but on several pattern recognition techniques. On the other hand, techniques selected should be sufficiently different such that they can be assumed to be especially suitable for different kinds of problems.

Following a similar argument, applying several intrinsic dimensionality and feature selection methods which should be as different as possible in scope seems necessary. Adhering to these considerations, Ipra is equipped with three different methods for each of the three stages (see Sect. 3.3). One criterion—apart from methodological diversity—for selecting the employed techniques were the number of parameters a technique needs to have specified to be applicable: techniques with few parameters were prefered. Methods with many parameters seemed unsatisfactory, because without prior knowledge about the to be analyzed data—which usually is not available—either the user would have to provide the parameter values by hand or optimal parameter values have to be estimated from the data. The latter seems objectionable, as more values to be estimated exacerbate the curse of dimensionality. The former, in contrast, is in disagreement with the third aim of ease of use, since setting the parameter values by hand requires considerable experience and expertise. Therefore, although it was not possible to avoid parameters completely, methods were selected such as to minimize the number of parameters to specify.

Aim 3: Ease of Use. In order to realize the third aim all selected techniques for each of the three steps were combined into a single program. This program builds on the PRTools

This is an electronic version of an Article published in Computational Intelligence (Copyright 2007, Blackwell Publishing):
Schultheis, H. (2007). IPRA -- Enhancing the sensing abilities of ambient intelligence. Computational Intelligence, 23(4), 464 - 483.

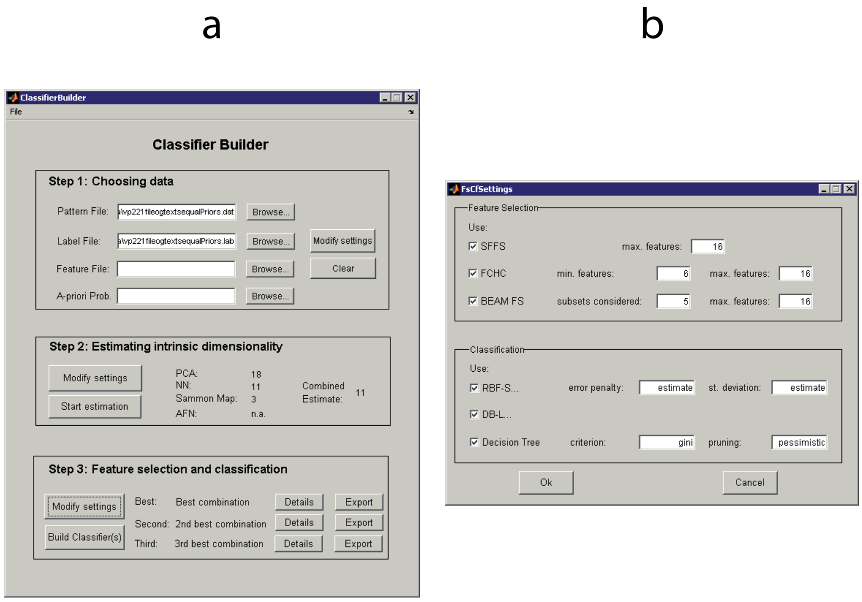

FIGURE 1. The IPRA GUI (a) and the dialog for parameterizing feature selection and pattern recognition (b).

software package version 3.0 of Duin (2000) and provides a GUI from which to access and control the employment of the techniques (see Figure 1). For each step of the general procedure, that is, for intrinsic dimensionality estimation, feature selection, and pattern recognition, three techniques can be freely chosen to be applied or not. Furthermore, for every technique chosen to be applied—where necessary—the parameters for the technique's application can either be set to some default value (extracted from the literature), be set to an arbitrary value the user might want to provide, or be set to be estimated from the given data set. In doing so, it is ensured that (a) novice users by relying on default values or estimates can apply IPRA successfully in most situations and (b) more experienced users have the full freedom to parameterize the techniques according to their needs.

As a result, the general approach just outlined has the potential to achieve all of the three aims underlying the development of IPRA. The success with which this potential is fulfilled, however, depends heavily on the specific techniques employed. In the following (Sect. 3.3) we will motivate and describe the specific methods which have been chosen for IPRA. Yet, before doing so, it seems sensible to briefly mention the preprocessing steps which are brought to bear on the to be analyzed data to set the stage for the application of the specific methods.

3.2. Data Preprocessing

Given that IPRA utilizes intrinsic dimensionality estimation and feature selection techniques before applying pattern recognition, one might wonder what additional preprocessing would be necessary. However, intrinsic dimensionality estimation, feature selection, and pattern recognition all three work on the same set of originally given measurements. For at least two reasons such an approach can be expected to yield suboptimal results.

First, after the given data set has been analyzed with the methods specified by the user, it is necessary to check how accurately the person's state can be inferred from the physiological measurements and, importantly, which of the employed pattern recognizers is most accurate for inferring a person's state. One way to evaluate the different classifiers, called *holdout method* is to use only a subset of the available data for building the classifiers and to employ the remaining part of the data for validation. Compared to other procedures such as bootstrapping or cross-validation (see, e.g., Kohavi, 1995) it is computationally cheap, but tends to underestimate the true accuracy (see Sect. 4.2), since not all of the available data is used for building the classifier(s). As IPRA in its current implementation already exhibits a considerably long runtime we chose to employ the holdout method. One dialog of IPRA allows to specify which fraction and / or how many instances of the available measurements is / are supposed to be used for the test set. The according number of test set measurements is then automatically and randomly chosen from the available data.

Second, the different dimensions of the measurements may differ in the extent of the range of values which they span. This potentially poses a problem for all those methods which rely on determining distances including several techniques for intrinsic dimensionality estimation, feature selection, and pattern recognition. To avoid such unwanted influences of irrelevant information a common approach is to scale all dimensions to the same range prior to analysis (see Hsu et al., 2004; Sarle, 1997). Accordingly, this option is also available in IPRA. The user can specify whether and to which range the available dimensions should be scaled. By default scaling is enabled and uses the commonly employed range of $[-1, 1]$.

After splitting and scaling (if requested) of the physiological data, the measurements are prepared for analysis by the specific intrinsic dimensionality estimation, feature selection, and pattern recognition techniques which will be described next.

3.3. Intrinsic Dimensionality, Feature Selection, & Pattern Recognition: Specific Methods

In this section the three specific methods utilized in IPRA for intrinsic dimensionality estimation, feature selection, and pattern recognition, respectively, will be introduced. Although for each stage of the analysis numerous methods have been proposed the techniques can usually be grouped regarding one or two fundamental aspects. In accord with the aim of using preferably diverse techniques, the specific techniques were selected such that they have different characteristics with respect to these fundamental aspects. For deciding between techniques having the same characteristics, those were chosen which need fewer parameters and are more strongly supported by evaluation studies in the literature.

In the following subsections the fundamental aspects used for selection are explained before briefly explicating which methods were chosen for which reasons.

Intrinsic Dimensionality. Methods to estimate the intrinsic dimensionality d'can be grouped according to two aspects. A first distinguishing characteristic is whether a technique is primarily sensitive to linear or non-linear relationships in the data. The second grouping factor discriminates between methods designed to directly estimate d' compared to methods estimating d' as a by-product.

The core component of virtually all linear estimation techniques is the *principal component analysis* (*PCA*). PCA has the advantage of being computationally efficient and needing only few parameters. The major disadvantage of the PCA is the fact that it only takes into account linear relationships between dimensions and might thus overestimate the true d'. As a consequence, several modifications of the basic PCA have been proposed (see, e.g., Brsuke & Sommer, 1998). The basic idea underlying most modifications is to first linearize all or some part of the data and then apply PCA. Because such modifications entail that (a) more parameters have to be specified, (b) the d' of the transformed data cannot always be guaranteed to be identical to the d' of the original data, and (c) the computational cost is increased, we decided to use the unmodified PCA as the linear estimation technique.

Still directly estimating the d', but also taking non-linear relationships into account is the nearest neighbor estimation (Verveer & Duin, 1995). This method is not only computationally efficient and needs no parameters to be specified, but also has been shown to be among the most accurate estimators (cf. Wyse et al., 1980). Consequently, it has been selected as the second technique for estimating d'.

The third chosen method belongs to the group of projection techniques which estimate d' indirectly. Of the projection methods available, like *auto-associative feed forward networks* (cf. Baldi & Hornik, 1989), *self-organizing maps* (Kohonen, 1997), or *Sammon's mapping* (Sammon, 1969), the last is most appropriate for physiological data, because its computational complexity increases with the number of measurements, but is independent of the number of dimensions.

The overall estimate of d' is the mean of the estimates of all applied methods.

Feature Selection. Since it is infeasible to consider all subsets during feature selection (Sect.3.1), selection techniques employ heuristics to guide the search. The criterion used for guidance in virtually all selection methods is the quality of the already considered subsets assuming that good subsets are more similar (with respect to the contained dimensions) to the best subset than bad subsets. However, problems vary regarding the degree to which this ordering assumption holds. As Cover & Van Campenhout (1977) have shown, there exist problems for which removing or adding one dimension to the best subset will result in the worst subset. Consequently, for some problems, relying on this assumption will lead the search astray, whereas for others it will lead to the best feature subset. In order to achieve good results for a wide variety of problems, selection methods were chosen such that they rely on the ordering assumption to a different degree.

The first method chosen was *first choice hill climbing* (*FCHC*). This method needs few parameters, is computationally efficient, and has been shown to be quite suitable especially for large search spaces (Russell & Norvig, 2003).

The *sequential floating forward selection* (*SFFS*) developed by Pudil et al. (1994) was chosen as the second method, because it needs only one parameter to be specified and has been shown in several evaluation studies to be as good as or better than other selection techniques (Reunanen, 2003; Molina et al., 2002).

As a third selection method a variant of the *beam search* (*BEAM*) proposed by Aha & Bankert (1996) has been chosen. Although this method is computationally more complex than the other two methods it will, in case the ordering assumption holds for the given problem, be better in finding the best subset.

In all of these methods—and indeed in virtually all feature selection techniques available—the goodness of feature subsets has to be determined repeatedly. But what makes a feature subset a good one? Loosely speaking a feature subset F_s can be assumed to be so much the better, the more similar measurements stemming from the same state and the more dissimilar measurements stemming from different states are, when considering only the measurement's features contained in F_s. Looking at the literature one can find various criteria like, for instance, the Mahalanobis distance (Liu & Motoda, 1998), the Bhattacharyya coefficient (Devijver & Kittler, 1982), or the directed divergence (Liu et al., 1998) which have been proposed as methods for gauging the goodness of feature subsets. These and related measures do only take the measurements themselves and their distribution in the reduced feature space into account. Such a restriction to only the (distribution of the) measurements, however, will normally yield only suboptimal results, as the *ugly duckling theorem* put forth and proven by Watanabe (1985) illustrates.

Basically, the ugly duckling theorem states that the notion of similarity is completely context dependent, that is, without any context information the similarity between two measurements m_1 and m_2 is identical to the similarity of any two measurements. Put differently, the notion of similarity does only make sense with respect to a certain purpose (i.e., context) which requires determining the similarity between two measurements. In the scope of Ipra this purpose is to build classifiers which allow to infer the current state of a person as good as possible. Although the above mentioned criteria for evaluating feature subsets reasonably account for this purpose by computing, for example, intra and inter class distances and variablities as well as conditional probabilities, they fall short of including one important context aspect, namely the pattern recognition technique utilized, in determining the subsets' goodness. This is illustrated by the fact that using the above mentioned criteria would result in selecting the same feature subset for all pattern recognition teqniques employed. Due to the considerable differences in the properties of the various techniques, however, it is unreasonable to assume that the best possible decision boundaries can be estimated with the same feature subset for all techniques. Accordingly, an appropriate criterion for judging subset goodness should take the idiosyncrasies of the particular pattern recognition techniques into account. One such criterion, which has been adopted in Ipra, is the classification accuracy of the respective technique for a given feature subset. By utilizing the pattern recognizer in evaluating the subsets it is guaranteed that the specific properties of the recognizer are taken into account and the best possible subset can be determined.

Pattern Recognition. The two aspects used to distinguish groups of pattern recognition techniques are whether the estimated boundaries are linear and whether the estimate relies on the distances between measurements.

Out of the set of available linear pattern recognition techniques *support vector machines* (*SVM*, cf. Burges, 1998) are used in Ipra, as they have been shown (a) to be comparatively insusceptible to the curse of dimensionality (Cristianini & Shawe-Taylor, 2000) and (b) to be effective in analyzing physiological data (Rani et al., 2006). Because SVMs, furthermore, need only two parameters to be specified they seemed to be a good choice for Ipra.

Another pattern recognition technique which has been applied with considerable success is the *learning vector quantization* (*LVQ*) developed by Kohonen (see Kohonen, 1997).

Although LVQ has been applied quite successfully, the original form has three major drawbacks. First, the number of codebook vectors used for pattern recognition has to be specified by hand and before learning starts. Choosing too few vectors might reduce the accuracy, whereas choosing too many unnecessarily raises the computational complexity. Second, the amount to which codebooks vectors are moved in each step is moderated by a parameter which also has to be specified by hand. If the parameter is set too high, learning might not converge, but setting it too low needlessly slows learning. In particular, both the optimal number of codebook vectors and the optimal learning rate can be assumed to be problem dependent. Since normally there is no information available on the optimal setting for a given recognition problem choosing values by hand will often result in suboptimal pattern recognition performance.

To eliminate these problems (Bermejo, 2000) and (Bermejo & Cabestany, 2000) have developed two LVQ variants. One, called *dynamic learning vector quantization* (*DLVQ*), solves the first of the mentioned problems, whereas the other called *batch learning vector quantization* (*BLVQ*) solves the second concern with the original LVQ.

Thus, BLVQ and DLVQ solve the problems associated with LVQ. Yet, both of them do so only partially: BLVQ still requires choosing the number of codebook vectors and DLVQ uses iterative learning with a learning parameter. In order to remedy all three drawbacks of LVQ we developed a new LVQ variant, called *dynamic batch learning vector quantization* (*DBLVQ*) by combining DLVQ and BLVQ. This new variant works as follows: DBLVQ is started with as many codebook vectors as there are states to distinguish. These codebook vectors are set to the mean of all measurements available for the corresponding state. Now the overall classification error given these vectors is computed and—in case the error is above 0—BLVQ is applied until the classification error does not decrease any further. In a next step the set of codebook vectors is extended. In doing so, the new codebook vectors are placed such that they can be expected to decrease the classification error. More precisely, first, for each codebook vector cv_i its classification error is determined and those codebook vectors with the highest error are selected for further refinement. Refinement involves determining that state to which most of the measurements misclassified by cv_i belong. Assume that M_i is the corresponding set of measurements for cv_i. Then a new codebook vector cv_{i+} is added and this new vector is computed as the mean of all measurements in M_i. Having extended the set of codebook vectors the next step consists of training all codebook vectors anew until the classification error does not decrease anymore. The cycle of extending and training the set of codebook vectors continues until adding and training new vectors does not decrease the classification error any further. Since iteratively adding new codebook vectors might result in redundancies or vectors which do not classify any measurements, the last step of DBLVQ consists of pruning all those vectors which (a) can be removed without increasing classification error and (b) do not classify any measurements.

Compared to the other variants, DBLVQ has the distinct advantage to be completely parameter free. The user of IPRA does not need to specify the learning parameter or the number of codebook vectors when employing DBLVQ. Furthermore, since DBLVQ is built on DLVQ and BLVQ which have been shown to be accurate for a range of pattern recognition tasks (Bermejo, 2000; Bermejo & Cabestany, 2000), DBLVQ can be assumed to give good results for various pattern recognition situations. Consequently, DBLVQ has been chosen as the second pattern recognition technique to be employed in IPRA.

Both SVM and DBLVQ rely on distances between measurements to estimate the decision boundaries. To employ a wide variety of conceptually different techniques we also chose a method which does not draw on distances. More precisely, we chose to use a *decision tree* (*DT*) as the third pattern recognition technique. Although decision trees have rarely been used to analyze psychophysiological data so far, existing results (e.g., Rani et al., 2006) indicate the suitability of this method. In addition, only two parameters have to be specified and the computational complexity of DTs is rather low.

As in the other two stages use of any or all of these three techniques can be controlled by the user.

4. EXAMPLE APPLICATION

As a first evaluation of the general principles and specific methods of IPRA as well as the newly developed pattern recognition technique DBLVQ, IPRA has been applied to physiological data from from two experiments—one related to cognitive and the other related to affective states. In this section we will first lay out the procedure used to anayze both data sets before the data proper and the analysis results for each of the two experiments are described in turn.

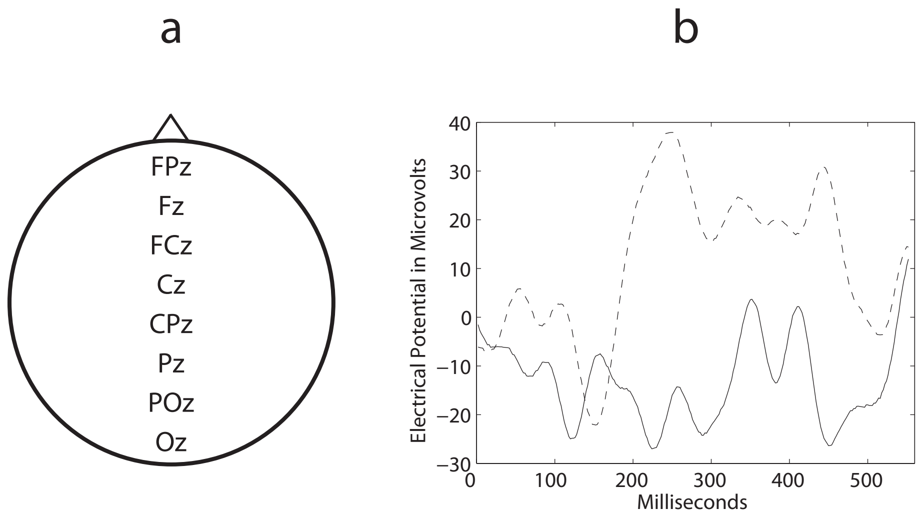

FIGURE 2. Position of the electrodes used to measure the P300 (a) and example EEG measurements for easy (solid line) and difficult (dashed line) reading (b).

4.1. Analyses Procedure

The procedure for analyzing the physiological data was the same for both data sets and was conducted employing the default parameter and method handling of IPRA. As a result, the analyses basically just required 6 button presses (see Fig. 1). First, the path to the file containing the measurements, the path to the file containing the labels for the measurements, and the path to the file containing the designations for the dimensions was chosen by using the three corresponding "Browse…" buttons. Subsequently, the to be analyzed data was loaded by pressing the "Load" button. After data loading was finished, intrinsic dimensionality was estimated by pressing the "Start estimation" button. The resulting estimate of the intrinsic dimensionality was automatically used to determine the range of dimensionalities to be considered in feature selection. As a consequence, pressing the "Building Classifier(s)" button sufficed to learn and evaluate the different classifiers, that is, to obtain the results presented in the following sections.

4.2. Cognitive States

The first set of physiological data we employed stems from an experiment conducted by Schultheis & Jameson (2004). In this experiment participants had to read texts of varying difficulty. It was hypothesized that reading easy texts would lead to a lower cognitive load than reading difficult texts. In accord with the hypothesis, several measures like subjective difficulty and interruptibility ratings, reading speed, text comprehension, and a component of the electroencephalogram (EEG) which were recorded during reading indicated that participants experienced higher load when reading difficult texts.

For the purpose of evaluating IPRA, the EEG component, namely the P300 is of main importance. The P300 is a positive variation of the ongoing EEG which usually occurs around 300 ms and which can be measured mainly in certain regions on the surface of the head. These regions are indicated in Figure 2a. Starting with the region closest to the forehead (labeled

FPz) the P300 mainly appears on a straight line to the back of the head (labeled Oz). Each of the labeled positions in Figure 2a also signifies a place where an EEG-electrode was attached to the scalp of the participants. The importance of the P300 for evaluating Ipra stems from the fact that this component of the EEG has repeatedly been shown to be sensitive to the cognitive load of a person: The higher the load, the smaller the P300 (Ullsperger et al., 2001; Sirevaag et al., 1989). Consequently, different cognitive states (high load vs. low load) lead to different EEG measurements and, thus, the current cognitive state of a user might be inferable from these measurements.

In the following the characteristics of the EEG data are described in more detail before the results of the analysis are presented.

The Data. EEG was measured at 500 Hz from the eight electrodes identified in Figure 2a. To identify the P300 in the measurements, EEG segments of 550 ms were considered for each electrode resulting in 276 measurements per electrode per trial. Furthermore, to allow considering the information available from all electrodes together, the 276 values of each electrode measurement were combined into a single measurement consisting of $8 * 276 = 2208$ values. For each participant, from 40 to 60 such measurements were recorded during the experiment. All measurements of one of the participants, participant 221, were used for evaluating Ipra. We chose this particular subject, because both a full 60 measurements were available for this participant and the P300 difference for the different cognitive states was clearly visible in the subject's average data.

However, this difference is far from obvious in the individual measurements as can be seen from Figure 2b. Although on average, that is, across all measurements, the P300 is clearly higher for easy texts, the relation might be reversed for individual measurements: due to the noise in the EEG signals, measurements from easy reading (solid line) might exhibit a smaller P300 than measurements from difficult reading (dashed line). Thus, the pattern recognition task can be assumed to be extremely difficult for the EEG data.

Analysis Results. The results of the analysis are displayed in Table 1. Each column presents the results for one of the pattern recognition techniques both without (first 3 columns) and with feature selection enabled (columns 4 – 6). The first row in each column indicates which feature selection method found the best dimension subset for the corresponding classifier. The second row signifies how many dimensions have been used for pattern recognition. The third row gives the accuracy of the pattern recognizers obtained by the analysis. The accuracy value was determined as the percent of measurements in the test set (see Sect. 3.2) which were assigned to the correct state. Although dimensions were selected with all feature selection techniques for each pattern recognition method, for the sake of clarity only the results of each pattern recognition method for the best selected dimension subset are shown.

Several points are noteworthy regarding the results: first, although the pattern recognizers perform above chance, the accuracies are low to moderate, the highest accuracy achieved being 70% correct. In part this can be assumed to be due to the difficulty of the problem (see above). Furthermore, however, the low accuracies might be an artifact of the way the recognizers were evaluated, since splitting the original data in training and test set has been shown to underestimate the true accuracy of pattern recognizers (cf. Jain et al., 2000).

Second, comparing columns 4 – 6 with columns 1 – 3 shows that feature selection is highly desirable: each of the three techniques exhibits improved accuracy with the reduced set of dimensions. Third, the estimated intrinsic dimensionality helped to considerably constrain the search space during feature selection. Although the estimated d' of 11 is still several points above the number of dimensions finally used, it is 200 times smaller than the original

TABLE 1. Classification results in percent correct for participant 221 of the experiment by Schultheis & Jameson (2004). Only the best results for each classifier are shown.

	Pattern Recognition Techniques					
	SVM	DBLVQ	DT	SVM	DBLVQ	DT
FS used	–	–	–	BEAM	SFFS	FCHC
#dimensions	all	all	all	1	1	2
accuracy	50%	60%	45%	65%	70%	60%

Estimated d' was 11

number of dimensions: instead of exploring the full power set of size 2^{2208} feature selection could be constrained to subsets containing 11 features at maximum.

Fourth, using several different feature selection and pattern recognition techniques pays off. Not only are there marked differences in the accuracies of the different pattern recognizers, but also the feature selection method giving the best dimension subset is different for the different classifiers. Especially, given the (little) available prior knowledge about the data, it would not have been possible to predict that the particular combination of SFFS and DBLVQ would give the best result for the EEG data. Finally, the newly developed DBLVQ did quite well in finding regularities in the data. Although DBLVQ's accuracy is still in the moderate range it achieved better recognition results than the two established pattern recognition techniques.

4.3. Emotional States

The second data set we considered in our application of IPRA comprised psychophysiological signals collected by Stemmler et al. (2001). In their experiment Stemmler and colleagues collected psychophysiological data related to heart activity, respiratory activity, muscle activity, skin conductance, and skin temperature during several experimental phases. Importantly, in three of the phases the experimental context was set up to induce anger in the participants. These phases were accompanied—among others—by phases of rest, where the subjects were instructed to relax but to keep their eyes open. After each induction and rest phase participants were asked to rate on a 11 point scale ranging from 0 to 10 how angry they were. From these anger self-reports and the measured physiological activity a data set was constructed and analyzed as described in the following.

The Data. From the measured signals, 29 physiological indicators such as heart rate, blood pressure, skin conductance, and forehead temperature were available for each participant in each phase (for a more detailed description of the indicators see Stemmler et al., 2001). As a result, 3 measurements of dimension 29 collected during phases of induced anger existed for each subject. To yield a balanced data set, the group of these 3 anger measurements was extended by 3 measurements from rest phases giving an overall of 6 measurements for each participant. Since 6 measurements seemed too few for applying pattern recognition even when utilizing feature selection, the data of 8 subjects was combined for analysis. These subjects were selected on the basis of their anger self-reports: self-reports for anger phases had to be above or equal to 5 and self-reports for rest phases were not allowed to exceed 1.

TABLE 2. Classification results in percent correct for participants 1, 15, 20, 44, 49, 65, 73, and 94 of the experiment by Stemmler et al. (2001). Only the best results for each classifier are shown.

	Pattern Recognition Techniques					
	SVM	DBLVQ	DT	SVM	DBLVQ	DT
FS used	–	–	–	BEAM	FCHC	SFFS
#dimensions	all	all	all	1	1	1
accuracy	67%	67%	67%	75%	83%	75%

Estimated d' was 5

Because physiological signals of different persons usually differ considerably from each other even when the persons report experiencing the same (intensity of) emotion, we included a unique identification for each participant as an additional dimension into the data set. Consequently, the data set finally analyzed consisted of 48 measurements with 30 dimensions each. Although compared to the EEG data these are few dimensions, the curse of dimensionality can still be expected to be a major problem also with this data set, as the number of measurements is low and, in particular, considerably below the amount of 600 measurements which would be required to avoid the curse (Jain et al., 2000).

Analysis Results. The results of the analysis are shown in Table 2. As can be seen from the last row the accuracies are generally higher than for the EEG data. Reaching a maximal accuracy of 83%, IPRA yields classification rates which commonly are deemed good for emotion data (cf. Rani et al., 2006)—in particular, considering that the data set combined the physiological measurements of 8 different persons.

Apart from this general accuracy difference the results are analogous to the results for the EEG data, and once more lend support to the rationale underlying the design of IPRA: (1) feature selection improves classification accuracy, (2) estimating intrinsic dimensionality considerably constrains the search space, (3) employing multiple feature selection and pattern recognition techniques is advantageous, and (4) DBLVQ outperforms SVM and DT.

4.4. Application Summary

Taken together, the two above reported analyses indicate that the three aims motivating the development of IPRA have been achieved. First, IPRA is an integrated approach in complementing pattern recognition with intrinsic dimensionality estimation and feature selection. Without d' estimation, feature selection, especially for the EEG data set, would not have been feasible and without feature selection, pattern recognition results would have been considerably worse. Second, the data sets analyzed cover various physiological signals ranging from EEG over skin conductance, heart, and respiration activity to muscle activity as well as cognitive and emotional states. The recognition results achieved across these different signals and types of states show that IPRA is broadly applicable which was the second central aim in developing IPRA. Finally, as explicated in Section 4.1 achieving the analyses results basically required pressing 6 buttons of the IPRA GUI. Thus, even non-experts can easily extract the relevant state information from given physiological data by employing IPRA.

Furthermore, the results of the analyses allow a first evaluation of the newly developed pattern recognition technique DBLVQ. Overall the accuracies indicate that this technique is suitable both for very difficult pattern recognition problems and for a variety of different types of physiological signals. In particular, for both data sets the method outperformed the established pattern recognition techniques.

5. IPRA FOR AMBIENT INTELLIGENCE

Ipra was designed to enhance the sensing abilities of ambient intelligence by making the information about a person's state inherent in physiological signals easily available. Judging from the results of Ipra's first applications (see Sect. 4) this aim seems to have been achieved. One might argue, however, that the analyzed data was collected in situations too different from real-life ambient intelligence settings. More precisely, reservations to accept Ipra's suitability for real-life settings might stem from the objections that (a) it is unrealistic to obtain such a big number of different physiological signals from a person in real-life, since the involved sensors and cables would be too annoying; (b) data obtained from persons behaving in the real world will be very noisy due to artifacts from movement and / or physical strain; and (c) even if recognition rates as achieved in the above data analyzes could be obtained in real-life situations, one would not want to base the adaptation of ambient intelligence devices on states predicted with 70% or 83% accuracy.

Although the final proof can only be given through applying Ipra to real life ambient intelligence situations, we think that there are good reasons to belief that Ipra's suitability is not compromised by these objections. Regarding objection (a), technological development has provided and will continue to provide small and unobtrusive sensors to measure physiological signals. One prime example is the LifeShirt developed by Wilhelm et al. (2006). This wearable device allows monitoring a multitude of physiological indicators as well as wireless transfer of the measured signals. Even EEG might be recordable without attaching electrodes to a person's head in the near future (Harland et al., 2002). Regarding objection (b), one can indeed assume that physical activity and movement will lead to noisy data in real-life ambient intelligence settings. However, this does not seem to pose a serious problem to Ipra. For one, as the analysis of the EEG data set has shown (see Sect. 4.2), Ipra is able to successfully cope with extremely noisy data. Furthermore, a person's movement might not always be a hindrance to but might even facilitate the recognition of states (Wilamowitz–Moellendorff et al., 2005). Regarding objection (c), as mentioned in the introduction, we do not propose to employ information from physiological measurements as the only source. Rather it should be integrated with state information from other sources using suitable information fusion approaches. As a complementary information source recognition rates in the range of 80% seem well suited to enhance the sensing abilities of ambient intelligence.

Consequently—together with the fact that Ipra gives good recognition results with few training samples and helps reducing computational complexity by selecting the most informative dimensions (see Sect. 3.1)—Ipra seems well suited for the application to real-life ambient intelligence settings.

6. RELATED WORK

Due to their advantages there has been a growing interest to use physiological signals as a source of information about a person's affective and cognitive states. This interest has led to advances of both available tools (e.g., De Clercq et al., 2006; Delorme & Makeig, 2004) and specific approaches for analyzing physiological data (e.g., Blankertz et al., 2007; Healey & Picard, 2005; Kulic & Croft, 2007; Leon et al., 2007; Lisetti & Nasoz, 2004; Picard et al., 2001; Rani et al., 2007).

Tools such as PSPHA (De Clercq et al., 2006) and EEGLAB (Delorme & Makeig, 2004) mainly provide functions for preprocessing (e.g., removing artifacts) of and feature extraction from physiological signals; they do not, however, provide functions to extract state information from the signals. Thus, Ipra can be viewed as complementing the existing tools by providing functionality they do not offer.

In contrast to the available tools, the main aim of the specific approaches for analyzing physiological data is to infer a person's states by employing machine learning techniques. In doing so, all of the current approaches do not utilize methods for intrinsic dimensionality estimation and—apart from few exceptions (Leon et al., 2007; Picard et al., 2001)—no methods for feature selection. One reason for this might be that the number of dimensions of the analyzed data sets is rarely above 15 (Kulic & Croft, 2007; Lisetti & Nasoz, 2004; Rani et al., 2007). Such low numbers of dimensions are usually achieved by extracting dimensions from the raw signals before applying pattern recognition and / or by restricting the number of considered physiological signals in the first place. As previously said, this has the disadvantages (a) of potentially neglecting important information, (b) that the approaches are not easy to transfer to other physiological measures, and (c) that considerable expertise is required to extract the most informative dimensions. More generally, most of the approaches (Blankertz et al., 2007; Healey & Picard, 2005; Kulic & Croft, 2007; Leon et al., 2007; Lisetti & Nasoz, 2004; Rani et al., 2007) employ highly specific, hand-tailored machine learning techniques. This not only makes it more difficult for the non-expert to adopt these techniques, but also raises doubt as to the suitability of these techniques for a different set of signals and / or extracted dimensions than the ones employed by the authors.

A notable exception to this common practice is the work of Picard et al. (2001) which uses well known feature selection and pattern recognition approaches in their analyzes. Yet, they do not use intrinsic dimensionality estimation and only apply one feature selection and only two pattern recognition techniques which again questions the suitability of this approach for physiological data in general. At any rate, in contrast to Ipra none of the specific approaches is currently available as a tool which renders the use of these approaches virtually impossible for non-experts.

7. CONCLUSIONS

Physiological data constitutes a valuable source of information about a human person and by exploiting this source ambient intelligence applications could be improved. However, since (a) extracting information from physiological signals requires the principled use of several machine learning techniques and (b) existing approaches lack in methodology and ease of use, ambient intelligence applications are usually barred from this important source of information. In response to this lack, we developed Ipra, an integrated pattern recognition approach which (a) employs all techniques necessary to appropriately analyze physiological data, (b) is broadly applicable, and (c) is easy to use even for non-experts. Furthermore, we developed a new pattern recognition technique, called dynamic batch learning vector quantization (DBLVQ) for use with physiological data. This new technique not only has the advantage of being parameter free, but also, due to its theoretical basis can be assumed to be superior to conventional learning vector approaches.

Both Ipra and DBLVQ have been evaluated with analyses of two physiological data sets and found to be suitable tools for the extraction of information about a person's cognitive and affective state from psychophysiological signals. In particular, the general principles underlying the construction of IPRA, namely its three step approach (first determine intrinsic dimensionality, then select appropriate features, then apply pattern recognition techniques) and the use of several distinct methods for each of these three steps, have been validated by the analyses.

There are three major issues regarding future work. For one, since IPRA is currently implemented in Matlab its computational efficiency is suboptimal. To further advance the approach, it would be advantageous to reimplement IPRA in a more machine oriented programming language. Assuming a suitable speed-up can be achieved through reimplementation, it seems, moreover, desirable to improve the current evaluation procedure. Instead of using training and test sets, for example, boot strapping could be employed. Finally, IPRA's evaluation needs to continue with respect to both additional physiological data and IPRA's usability to accrue more evidence of the capabilities of the system.

8. ACKNOWLEDGEMENTS

In this paper work done in the project READY of the SFB 378 Resource-adaptive Cognitive Processes and in the project R1-[ImageSpace] of the SFB/TR 8 Spatial Cognition is presented. Funding by the German Research Foundation (DFG) is gratefully acknowledged. We also thank the reviewers for their helpful comments and Gerhard Stemmler for providing the emotion data set.

REFERENCES

Aha, D. W., and R. L. Bankert. 1996. 'A comparative evaluation of sequential feature selection algorithms', in Artificial Intelligence and Statistics, eds., D. Fisher and Lenz J.-H., 199 – 206, Springer, New York.

Baldi, P., and K. Hornik. 1989. 'Neural networks and principal component analysis: Learning from examples without local minima', Neural Networks, **2(1)**, 53 – 58.

Bermejo, S. 2000. 'A dynamic LVQ algorithm for improving the generalisation of nearest neighbour classifiers', in Learning with Nearest Neighbour Classifiers, ed., S. Bermejo, 6–1 – 6–12, Universitat Politècnica de Catalunya, Barcelona.

Bermejo, S., and J. Cabestany. 2000. 'A batch learning vector quantization algorithm for nearest neighbour classification', Neural Processing Letters, **11**, 173 – 184.

Blankertz, B., G. Dornhege, M. Krauledat, G. Curio, and K.-R. Müller. 2007. The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. NeuroImage, **37(2)**, 539 – 550.

Brandherm, B., and A. Jameson. 2004. An Extension of the Differential Approach for Bayesian Network Inference to Dynamic Bayesian Networks, Special Issue "Uncertain Reasoning", International Journal of Intelligent Systems, **19(8)**, 727 - 748.

Bruske, J., and G. Sommer. 1998. 'Intrinsic dimensionality estimation with optimally topology prserving maps', IEEE Transactions on pattern analysis and machine intelligence, **20(5)**, 572 – 575.

Burges, C. J. C. 1998. 'A tutorial on support vector machines for pattern recognition'. Data Mining and Knowledge discovery, **2(2)**, 121 – 167.

Cover, T. M., and J. M. Van Campenhout. 1977. 'On the possible orderings in the measurement selection problem', IEEE Transactions on systems, man and cybernetics, **7(9)**, 657 – 661, (1977).

Cristianini N., and J. Shawe-Taylor. 2000. An Introduction to Support Vector Machines, Cambridge University Press, Cambridge.

De Clercq, A., B. Verschuere, G. Crombez, and P. De Vlieger. 2006. Psychophysiological Analysis [PSPHA]: A modular script based program for analyzing psychophysiological data, Behavior Research Methods, **38(3)**, 504 – 510.

Delorme, A., and S. Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, Journal of Neuroscience Methods, **134**, 9 – 21.

Devijver, P. A., and J. Kittler. 1982. Pattern Recognition: A Statistical Approach, Prentice/Hall, New Jersey, 1982.

Duin, R. P. W. 2000. PRTools Version 3.0: A Matlab toolbox for pattern recognition.

Gopher, D., and E. Donchin. 1986. Workload - An Examination of the Concept, in Handbook of Perception and Human Performance, Vol. 2, eds., K. Boff, L. Kauffmann, and J. Thomas, 41-1 – 41-49, Wiley, New York.

Harland, C. J., T. D. Clark, and R. J. Prance. 2002. Remote Detection of Human Electroencephalograms Using Ultrahigh Input Impendance Electric Potential Sensors, Applied Physics Letters, **81(17)**, 3284 – 3286.

Healey J. A., and R. W. Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors, IEEE Transactions on Intelligent Transportation Systems, **6(2)**, 156 – 166.

Hsu, C., C. Chang, and C. Lin. 2004 'A practical guide to support vector classification', in LIBSVM: A Library for Support Vector Machines. Software Available at Http://Www.Csie.Ntu.Edu.Tw/ cjlin/Libsvm.

Jain, A. K., R. P. W. Duin, and J. Mao. 2000. 'Statistical pattern recognition: A review', IEEE Transactions on Pattern Analysis and Machine Intelligence, **22(1)**, 4 – 37.

Kohavi, R. 1995. 'A study of cross-validation and bootstrap for accuracy estimation and model selection', in International Joint Conference on Artificial Intelligence, ed., C. S. Mellish, 1137 – 1143, Morgan Kaufmann.

Kohonen, T. 1997. Self-Organizing Maps, Springer, Berlin.

Kramer, A. F. 1991. Physiological Metrics of Mental Workload: A Review of Recent Progress, in Multiple-Task Performance, ed., D. L. Damos, 279 – 327, Taylor and Francis, London.

Kulic D., and E. Croft. 2007. Physiological and subjective responses to articulated robot motion, Robotica **25**, 13 – 27.

Leon E., G. Clarke, V. Callaghan, and F. Sepulveda. 2007. A user-independent real-time emotion recognition system for software agents in domestic environments, Engineering Applications of Artificial Intelligence, **20(3)**, 337 – 345.

Lisetti C. L., and F. Nasoz. 2004. Using noninvasive wearable computers to recognize human emotions from physiological signals, Journal on applied signal processing, **11**, 1672 – 1687.

Liu, H., and H. Motoda. 1998. Feature Selection for Knowledge Discovery and Data Mining, Kluwer, Boston.

McFarlane, D. C. and K. A. Latorella. 2002. 'The scope and importance of human interruption in HumanComputer interaction design.', Human-Computer Interaction, **17**, 1 – 61.

Molina, L. C., L. Belanche, and A. Nebot. 2002. 'Evaluating Feature Selection Algorithms', in Topics in Artificial Intelligence: 5th Catalonian Conference on AI, ed., M. T. Escrig, F. Toledo, and E. Golobardes, 216 – 227, Springer, Heidelberg.

O'Donnell, R. D., and F. T. Eggemeier. 1986. Workload Assessment Methodology, in Handbook of Perception and Human Performance, Vol. 2, eds., K. Boff, L. Kauffmann, and J. Thomas, 42-1 – 42-49, Wiley, New York.

Picard, R. W. 2000. Toward computers that recognize and respond to user emotion, IBM Systems Journal, **39(3&4)**, 705 – 719.

Picard, R. W., E. Vyzas, and J. Healey. 2001. 'Toward machine emotional intelligence: Analysis of affective physiological state', IEEE Transactions on pattern analysis and machine intelligence, **23(10)**, 1175 – 1191.

Pudil, P., J. Novovicova, and J. Kittler. 1994. 'Floating search methods in feature selection', Pattern recognition letters, **15**, 1119 – 1125.

Rani P., C. C. Liu, N. Sarkar, and E. Vanman. 2006. An empirical study of machine learning techniques for affect recognition in human-robot interaction, Pattern Analysis and Applications, **9(1)**, 58 – 69.

Rani P., N. Sarkar, J. Adams. 2007. Anxiety-based affective communication for implicit human-machine interaction, Advanced Engineering Informatics, **21(3)**, 323 – 334.

Reunanen, J. 2003. Overfitting in Making Comparisons Between Variable Selection Methods, Journal of Machine Learning Research, **3**, 1371 – 1382.

Russell, S., and P. Norvig. 2003. Artificial Intelligence: A Modern Approach, Prentice Hall, Upper Saddle River.

Sammon, J. W. 1969. 'A nonlinear mapping for data structure analysis', IEEE Transactions on computers, **18**, 401 – 409.

Sarle, W. S. 1997. Neural network FAQ. Periodic posting to the usenet newsgroup comp.ai.neural-nets.

Schaffer, C. 1994. 'A conservation law for generalization performance', in Proceedings of the Eleventh International Conference on Machine Learning, eds., William W. Cohen and Haym Hirsh, 259 – 265, Morgan Kaufmann, San Francisco.

Schultheis, H., and A. Jameson. 2004. 'Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods', in Adaptive hypermedia and adaptive web-based systems: Proceedings of AH 2004, eds., Wolfgang Neijdl and Paul De Bra, Springer.

Sirevaag, E.J., A.F. Kramer, M.G.H. Coles, and E. Donchin. 1989. 'Resource reciprocity: An event-related brain potential analysis.', Acta Psychologica, **70**, 77 – 97.

Stearns, S. D. 1976. On Selecting Features for Pattern Classifiers, in Proceedings of the Third International Conference on Pattern Recognition, 71 – 75, Coronado, CA.

Stemmler, G., M. Heldmann, C. A. Pauls, and T. Scherer. 2001. Constraints for emotion specificity in fear and anger: The context counts, Psychophysiology, **38**, 275 – 291.

Trunk, G. V. 1979. 'A problem of dimensionality: A simple example', IEEE Transactions on Pattern Analysis and Machine Intelligence, **1(3)**, 306 – 307.

Tsang, P., and G. F. Wilson. 1997. Mental Workload, in Handbook of Human Factors and Ergonomics, ed., G. Salvendy, 417 – 449, Wiley, New York.

Ullsperger, P., G. Freude, and U. Erdmann. 2001. 'Auditory probe sensitivity to mental workload changes - an event-related potential study', International Journal of Psychophysiology, **40**, 201 – 209.

Verveer, P. J., and R. P. W. Duin. 1995. 'An evaluation of intrinsic dimensionality estimators'. IEEE Transactions on pattern analysis and machine intelligence, **17(1)**, 81 – 86.

Watanabe, S. 1985. Pattern Recognition: Human and Mechanical, Wiley, New York.

Wickens, C., J. G. Hollands. 2000. Engineering Psychology and Human Performance, Prentice Hall, New Jersey.

Wilamowitz–Moellendorff, M., C. Müller, A. Jameson, B. Brandherm, and T. Schwartz. 2005. Recognition of Time Pressure via Physiological Sensors: Is the User's Motion a Help or a Hindrance?, in Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors in conjunction with the User Modeling (UM05) conference, Edinburgh, UK.

Wilhelm, F. H., M. C. Pfaltza, and P. Grossman. 2006. Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion, Interacting with Computers, **18**, 171 – 186.

Wilson, G. F., and F. T. Eggemeier. 1991. 'Psychophysiological assessment of workload in multi-task environments.', in Multiple-Task Performance, ed., D. L. Damos, 329 – 360, Taylor and Francis, London.

Wyse, N., R. Dubes, and A. K. Jain. 1980. 'A critical evaluation of intrinsic dimensionality algorithms', in Pattern Recognition in Practice, eds., Edzard S. Gelsema and Laveen N. Kanal, 415 – 425, North-Holland Publishing Company, Amsterdam.