Project Identification	
Project number	AAL 2012-5-199
Duration	1st May 2013– 30th April 2016
Coordinator	Martin Morandell
Coordinator Organisation	AIT Austrian Institute of Technology GmbH, Austria
Website	www.relaxedcare.eu

Report on behaviour pattern recognition for targeted areas

Document Identification	
Deliverable ID:	D-4.1 Report on behaviour pattern recognition for targeted areas
Release number/date	V0.3 1.10.2014
Checked and released by	Martin Morandell, AIT
Work Status	Finished
Review Status	Accepted

Key Information from "Description of Work"	
Deliverable Description	Description of relevant methods and algorithms for the BPR
Dissemination Level	PU=Public
Deliverable Type	R = Report
Original due date	Accepted

Authorship& Reviewer Information	
Editor	AIT, Jonathan Steinhart
Partners contributing	]
Reviewed by	Martin Biallas, IHL

The project RelaxedCare is co-funded by the European AAL JP and the following national authorities and R&D programmes from Austria, Switzerland, Slovenia and Spain

Release History

Release Number	Date	Author(s)	Release description /changes made Please make sure that the text you enter here is a brief summary of what was actually changed; do not just repeat information from the other columns.
V 01	1.12.2013	JS	First version with general overview
V02	30.06.2014	JS	Continuous updates due to agile approach
V03	1.10.2014	JS	First full version
V04	10.12.2014	MMo	Applied new template

Relaxed Care Consortium

Relaxed Care (AAL 2012-5-199.) is a project within the AAL Joint Programme Call 5

The consortium members are:

Partner 1	AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GmbH (AIT, Project Coordinator, AT)
Contact person:	Martin Morandell
Email:	Martin.morandell@ait.ac.at

Partner 2	*Hochschule Luzern Technik & Architektur – iHomeLab (IHL, CH)*
Contact person:	Martin Biallas
Email:	Martin.Biallas@hslu.ch

Partner 3	*50plus GmbH (50p, AT)*
Contact person:	Frauke Lettmann
Email:	frauke.lettmann@50plusgmbh.com

Partner 4	*New Design University (NDU, AT)*
Contact person:	Sandra Dittenberger
Email:	sandra.dittenberger@ndu.ac.at

Partner 5	*Mobili (MOB, SI)*
Contact person:	Jure Lampe
Email:	jure.lampe@mobili.si

Partner 6	*Szenographie (SZE, CH)*
Contact person:	Ralph Eichenberger
Email:	contact@szenografie.com

Partner 7	*Ibernex (IBE, ES)*
Contact person:	Antonio Remartinez
Email:	antonio.remartinez@ibernex.es

Partner 8	*Soultank AG (SOUL, CH)*
Contact person:	Bianca Redel
Email:	bianca.redel@soultank.ch

Partner 9	*Schweizerisches Rote Kreuz Luzern (SRK, CH)*
Contact person:	Nick Bolli
Email:	nick.bolli@srk-luzern.ch

Table of Contents

Table of Figures

List of Tables

Abbreviations

Abbrev.	*Description*
usw.:	und so weiter

Executive Summary

The RelaxedCare project aims to integrate off-the-shelf sensors into a solution which is easy for both informal caregivers and assisted persons. This will involve taking very low-level data (such as motion, opening and closing of door s and windows, etc.) and producing high-level abstractions (such as activity level) which are informative to the caregiver, but which do not violate the privacy of the assisted person. This task will involve the development of complex algorithms stemming from fields such as computer science, statistics, and machine learning. The purpose of the current document is to examine the task, given what information is already available concerning requirements and system specification, and develop a general approach to the problem which will help us to proceed as the task becomes clearer. We pay special attention to three areas which would be desirable to include in RelaxedCare: those of activity, emotion, and social interaction. We also attempt to identify the major challenges which will be encountered, as well as the keys to succeeding at our task as much as possible.

1 About this Document

1.1 Role of the deliverable

This deliverable aims to provide an overview of the methods that will be employed achieving the pattern recognition goals which arise from product requirements. The exact algorithms (as well as the processes used to implement them) are intimately and unavoidably connected to the exact details of the tasks — i.e., what data will be available to use, what exact form will they take, and what are the precise inferences we wish to make with them? Therefore, we must settle for a general overview of the task and how we'll approach it, and offer some background which will likely be relevant as we move forward.

1.2 Relationship to other Relaxed Care deliverables

The deliverable is related to the following Relaxed Care deliverables:

Deliv:	*Relation*
D2.4	Specificiation of technical requirements of system
D3.2	Description of the platform Description of the sensors

This deliverable depends upon both of the aforementioned ones for a more thorough analysis.

2 General aim of this workpackage

2.1 Introduction to this chapter

From the narrow point of view of this work package, the RelaxedCare system aims to provide a non-intrusive, user-friendly system which uses data from off-the-shelf sensors to ascertain complex states such as activity level, social interaction, and mood. Our goal here, then, is to fill in the black box that will turn the very low-level sensor data (comprising e.g. motion, door openings/closings, phone usage) into fairly high-level assessments such as those just listed. To do this, we will likely need to employ a variety of methods arising from the fields of mathematics and computer science, as well as some intuition about some underlying relationships between our inputs and outputs.

2.2 Requirements

2.2.1 Functionality

The overall goal of the RelaxedCare system is to "support the physical and mental care ability of the caregiver, including preventing psychological stress" — and to do so in an unobtrusive way. This workpackage involves the portion of the project lying between the collection of data from the environment of the cared-for user, and the conveyance of useful status information to the caregiver. Given whatever data are available, we must transform and analyze it in some way to make it meaningful to a caregiver. Although the exact "inputs" (raw sensor types) and "outputs" (caregiver-level status reports, based on inferred states) have not yet been concretely specified, it may be useful to group potentially inferrable states (outputs) into three categories to facilitate further discussion below:

- Low-level: In this category, we include states which can be determined with little to no abstraction or inference, given the types of sensors we will (probably) have. Examples of these would include whether or not the user has moved in the past n hours, given motion sensor data; whether the temperature is too high or too low given temperature sensor data; or whether the user has perhaps left doors or windows open during inclement weather given door closure and (externally available) current weather data.
- Mid-level: These would include states which involve a slightly higher level of abstraction from the raw-sensor level. This could include determining whether or not the user has had visitors, which might involve analyzing motion sensor data to determine if e.g. simultaneous motion had occurred at different places in the apartment, or if the amount of motion on a particular sensor was significantly greater than usual during some duration. Another example is looking at all of the sensor data as a whole over the course of days or weeks to get a picture of what normal daily routine looks like, and then notifying the caregiver in the event of a significant departure from that routine.
- High-level: In this category, we would place states which are significantly removed from sensor data. Examples of these might range from the relatively gross (like moods, such as lonely or depressed) to the fine (such as fluid intake, given only motion and door closure data).

Note that this grouping depends very heavily on the exact data which are available. Determining if someone is drinking water given only binary moving/not-moving data from a single motion sensor is impossibly high-level; but, at the other extreme, if the user only drinks water from a special pitcher which is outfitted with a outflow sensor, then the task becomes trivial.

2.2.2 Design

Various aspects of the aforementioned functionality will certainly be common to many current and future AAL projects, both within the RelaxedCare consortium and beyond. For this reason, there would be great value in a design which is as generic and modular as possible. As in any non-trivial system, however, there will be a delicate balance between this genericness on one side, and the ability for the system to perform the specific behaviors we need on the other — all while keeping things as simple and elegant as possible.

Our design should allow future implementors to experiment with new algorithms in a relatively unconstrained way. Inputs types should be standardized and the input data itself should be guaranteed free from the idiosyncrasies of individual sensors and sensor types. Useful metadata (such as sensor location or characteristics) can also be passed in in some standardized format — but care must be taken to keep this to a minimum, lest we circumvent the very modularity we're trying to build in. Similarly, outputs (e.g. state labels) should also be standardized, and new ones easily created for use with current and future algorithms. Inspiration for certain aspects of this design might come from RapidMiner or Weka, which likely use some abstractions and type hierarchies that would be quite similar to those we need for modeling inputs, outputs, and algorithms.

A clean and elegant solution to this design problem may be difficult, which is why we will suggest an iterative design process below.

2.3 Summary

The challenge of this work package is to take quite low-level data, and to synthesize them into variously complex assessments of abstract states including activity, mood, and social interaction. Furthermore, from a design aspect, care should be taken to implement this functionality in a modular and reusable way.

3 Introduction to behavior pattern analysis

3.1 Introduction to this chapter

We can motivate a discussion of some basic concepts in pattern recognition with a simple example. Imagine that there is a home with one person inside, and which is wired with 10 motion sensors such that any given spot is covered by at least one sensor. Every second, each sensor logs a 1 (if there is currently movement) or a 0 (if there is currently no movement). Now, given only data for the past 20 minutes (an example is shown here in tabular and graphical form in figures 1 and 2), what kinds of things can you imagine being able to determine about what's happening inside?

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Bedroom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Living Room	0	0	1	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	1	0	1
Bathroom	0	1	0	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	0	1	0
Hall	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0

Figure 1: Twenty minutes of hypothetical raw data from motion sensors

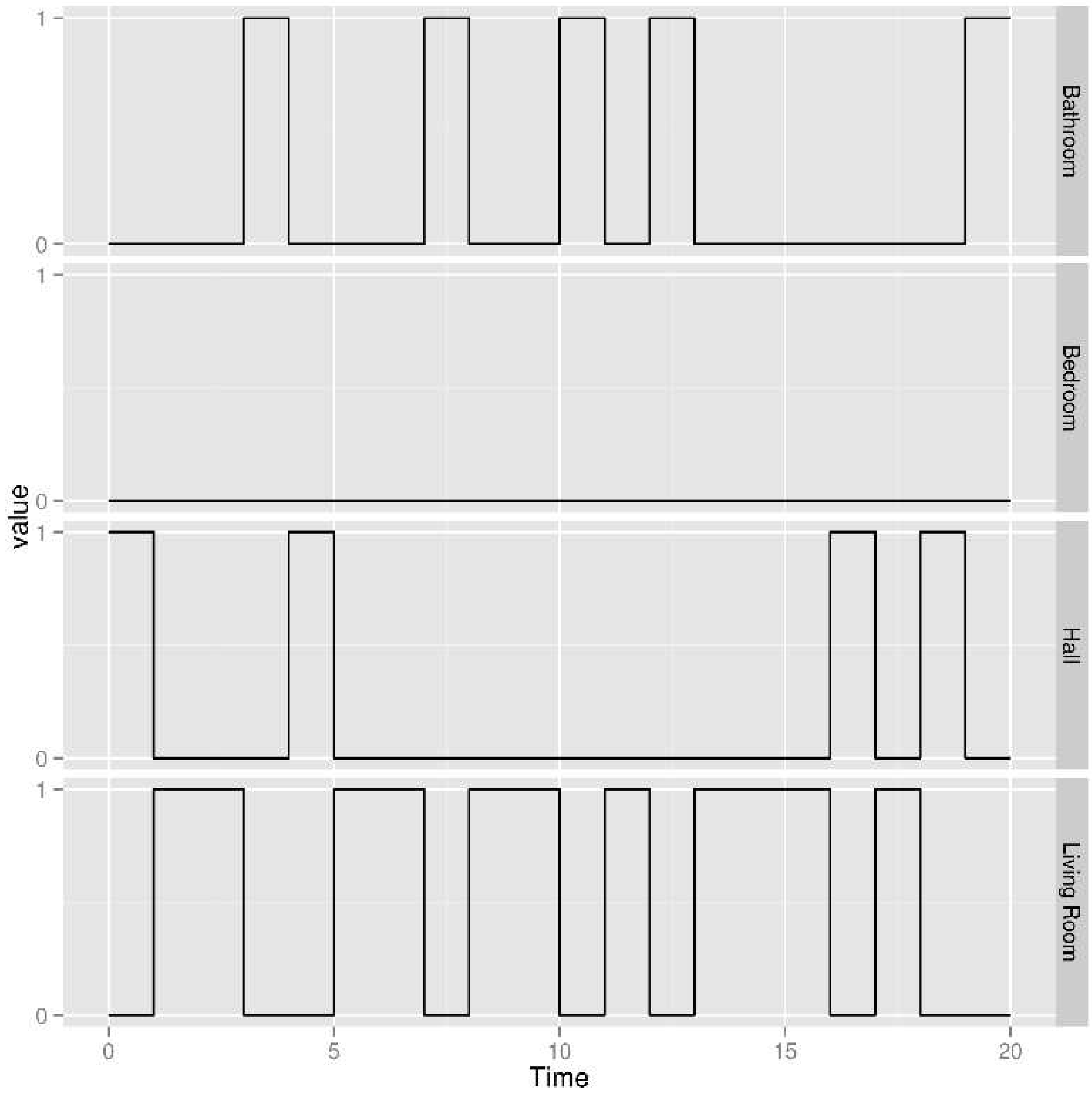

Figure 2: Graphical representation of data in figure 1

- One simple thing would be to determine approximately where the person is located — look for the sensor which has all 1's most recently (and if there's more than one, then they're in between them).
- A slightly more complicated example would be to determine some estimate of the person's "activity level" — maybe this could be the proportion of time periods during which they're moving (regardless of where), or perhaps the frequency with which they change rooms (which might actually be more like an "anxiety level").
- A more difficult task still would be to determine if the person socialized during this time. We could imagine that, if a visitor were present, there *could* be motion in more than one room at a time, but not necessarily.
- An impossibly difficult task would be to use these data to determine if the person drank water during this time.

Despite being a simple toy dataset, this gives us an example to use throughout the rest of this section, as a way of illustrating the basic tools available for doing pattern recognition on sensor data. It also demonstrates a rough guideline for what is or isn't possible with a given dataset — if you can't imagine a certain thing being "in the data" (e.g. water-drinking from the example data shown), then it's probably not[1].

Also, of course, real data are much messier: real sensors sometimes fail to send updates or even send false values (collectively referred to as "noise"), and a real dataset could have many more than 4 sensors (which might each emit more than one stream), each with hundreds of thousands of observations, spanning a year or more. Fortunately, dealing with noisy, complex data is just the sort of thing we will be developing algorithms to handle; and these algorithms need lots of data in order to learn the signal from the noise (and computers don't get fatigued).

3.2 A general formulation of the problem

Although the kinds of problems we can tackle with pattern recognition tools are quite diverse, considering a basic formulation which ties many of these together will give us a starting point for comparing and contrasting methods. In general, the task is:

- Given inputs $\boldsymbol{X}$, produce "best" output Y despite noise
    - For example, X might be the last 20 minutes of motion and door open/close sensor data; Y might be estimated location, the activity level, or whether or not someone is at home.
    - BUT there is a fan moving in one room, and a door blowing in the wind in another
- Possible values of X,Y determine potential algorithms
    - For example, should if Y is location, should it be (2D, continuous) coordinates of estimated position, or maybe a (categorical) variable representing which room?
    - If Y is instead the activity level, should it be some numerical scale, or perhaps a rough Low/Medium/High (ordinal) variable?
- Definition of "best" guides application of algorithm
    - If estimating the persons coordinates, the best guess might be the one with the least Euclidean distance from the true one. But this could place someone just on the other side of a wall, in a room which is not connected directly connected to the true room — so the room connectivity might need to be taken into account.
    - If determining home/not home, we might want to guess the state most likely to be true… *but* a person could be actually home and just not moving because of an emergency, which means we might want to only conclude "not home" when we're very sure.
- "Richness" of input limits performance
    - This is illustrated by the example in **Fehler! Verweisquelle konnte nicht gefunden werden.**: trying to detect water-drinking given only motion sensor data will never give results much better than chance. Imagine, on the other hand, trying to perform the same task using video cameras instead of motion sensors — the drastically increased richness of the data stream could make this "drinking detection" quite accurate, even

1 Although actually we might imagine using the general level of activity in the bathroom as a rough proxy for fluid intake! Although this particular example would probably be too error-prone, it shows the important role of creativity in attempting to find difficult patterns in data.

the point of discerning water from other fluids in many/most cases[2].

3.3 A rough taxonomy

With a basic understanding of the framework within which we will operate, it is now possible to discuss the basic differences between potential tools we might employ (including a few examples), with a focus on capabilities and data requirements.

It is important to note that no one approach will meet all (or even most) of our needs. Given the number and complexity of our likely goals with RelaxedCare, we will probably employ a broad mix of solutions, representing most of the categories discussed below.

3.3.1 Deterministic vs. Probabilistic

The first and most fundamental dichotomy between possible methods is between those which are designed to cope with uncertainty (probabilistic) and those which are not (deterministic).

A deterministic model assumes that the system evolves in the same way every time — e.g., if A occurs, then B will surely follow. Such a model may be expressed in terms of a set rules, a flowchart, or a finite state machine. For example, we could say that "if the front door opens and closes and then there is no movement for 10 minutes, then conclude that the person is not at home". This gives a natural way to express knowledge and beliefs about the relationship between what we observe and what we can conclude therefrom. It also has the benefit that in most cases, we wouldn't need data before developing the model (although of course we would need real data to test our models on in the end).

The downside of the deterministic approach is that often, in reality, things like sensor noise and unpredictable varations in individual behavior can sometimes make it difficult or impossible to come up with rules which apply in all cases. In the previous example, a faulty sensor which falsely registers movement a few times an hour could make us wrongly think that someone is home. Of course, we could add a rule to ignore reported "motion" below a certain threshold, but ad hoc patching can only take us so far (and can also result in a very large, unmanageable rule set).

When uncertainty plays a significant role, probabilistic methods provide a very powerful solution by allowing reasoning and decision-making within that uncertainty, providing a way to guess the most likely state, to determine how certain we are of our guesses, and even to plan a course of action which will have the greatest expected benefit (or, alternatively, the least risk). Because the ability to reason under uncertainty is such a fundamental need across all of science, economics, and other fields, these models represent a vast body of research in probability, statistics, machine learning, data mining, etc., and there are many options to choose from.

The downside of probabilistic models is that, except for the simplest cases (e.g. modeling a series of coin flips), we require example data in order to train the model to understand the relationships we're interested in making. The more complex the task, the more data is usually required. As discussed in **Fehler! Verweisquelle konnte nicht gefunden werden.**, and also below in **Fehler! Verweisquelle konnte nicht gefunden werden.**, certain kinds of tasks will require additional data (labels) in order to train our models.

3.3.2 Supervised vs. Unsupervised

Within the class of probabilistic methods, there are two main classes of algorithms, differing by the basic type of problem they're able to solve. In *supervised learning*, the system can learn by example to classify things or to estimate a function based on some observable features. For example as above we may want, given sensor data, to classify the current state as "at home" or

---

2 As above, we are assuming a human performing this task for demonstration purposes — while computers have come a long way at image/video analysis, they still have much further to go, and this level of automated analysis would still be quite challenging if not impossible.

"not at home", or we may want to estimate the number of hours during which the person was not at home. There are a very large number of algorithms which are specialized at handling any particular kind of data, ranging from classical methods like generalized linear models (including linear and logistic regression), to very new and high-powered methods like trees, support vector machines, and countless others. Given enough of the right kind of data, these methods are capable of learning astonishingly complex relationships with high accuracy.

This power, of course, comes at a cost. The word "supervised" implies that we are teaching the algorithm, for example, the difference between "at home" and "not at home" by showing it examples of each, and penalizing it when it makes a wrong guess. But this requires showing it what the sensor data looks like in both states, and — most importantly — that the examples of each are labeled as such in the dataset. This commonly (and in the case of RelaxedCare) means additional work: either a researcher has to observe the person (perhaps either live or on video) to take note their activities, or the person himself must track these (perhaps via a mobile app or a diary). Each state that we want to learn must also be tracked — so if we want presence, activity, and mood, we would need some record of "at time X, the person was at home, watching TV, and depressed". Clearly, this introduces significant overhead and intrusiveness[3]. Another complication is that for supervised learning to work, we generally need a sufficient number of both positive and negative examples — this is OK in the case of at-home/not-at-home, but in the case of, for instance, detecting an emergency state, we would hopefully not observe many positive examples.

Now suppose instead that we don't have labeled examples — an *unsupervised learning* task. In this case, we are not sure if we're looking at, for example, examples of data arising from a depressed vs. a happy state.

One thing we can do is look for patterns and attempt to group similar things together. This could be used in RelaxedCare, for instance, to detecting "normal behavior". Suppose the system is installed and, for 2 weeks, it only collects sensor data, and develops a model of the person's usual behavior which is robust to usual variations of wakeup time, movements, etc. Then if on one day there is a drastic departure from this routine, the system can detect this, perhaps even coming up with some estimate of how certain it is that today is different based on all observations to date. There exist quite powerful unsupervised algorithms for clustering all kinds of data. However, once we have the clusters, they may or not be separated along lines that are meaningful to us. For instance, we may cluster motion data and find that one cluster contains high-frequency motions, while the other contains low-frequency ones. But if what we are interested is mood, there's no way to tell if high-frequency means happy or agitated (or if low means depressed or relaxed).

On the other hand, tools like hidden Markov models, naive Bayes, and conditional random fields (all of which are special cases of the more general class of probabilistic graphical models) have already enjoyed great success in the context of ambient intelligence, and allow the modeler to incorporate her knowledge (or that of an expert in, say, geriatric medicine) into the structure of the model, resulting in even better predictions and more detailed guesses about the true state of things. However, it is an unavoidable fact that, without any labels on the data, the model is significantly less able to make detailed and accurate guesses about what's really happening.

3.4 Summary

There is an abundance of methods available for pattern recognition tasks which differ quite drastically based on data requirements, complexity, and capability. Once the exact requirements of RelaxedCare are more specifically defined, we will be able to decide on the appropriate tools, and to combine these into an effective solution.

---

3 Although, to be clear, we don't need this extra information from a normal RelaxedCare installation — this would be a dataset we only collect once, early in development, in order to train and test our system on.

4 Available Technology and tools within the consortium

4.1 Existing Technology at AIT

- HOMER: an open and flexible OSGi-based software platform which aims at the integration of various home automation sensors systems and consequential event and situation recognition for smart home and Ambient Assisted Living (AAL) applications

4.2 Existing Technology at IHL

- Matlab 2012: numerical computing environment and fourth-generation programming language
- RapidMiner: environment for machine learning, data mining, text mining, predictive analytics; also uses R modeling schemes. OpenSource (AGPL), Cross Platform

Weka (Waikato Environment for Knowledge Analysis). RapidMiner can also use Weka algorithms.

4.3 Existing Technology at MOB

5 How to measure social interaction

5.1 Introduction to this chapter

In the last years many different research groups keep occupied with the topic of social interaction with elderly people, but they also deal with other questions with respect to this matter. There exists different strategies to come to a solution, [1]uses passive, unobtrusive sensors with a hidden Markov model and Bayesian updating. [2] deals in their work with a Sociometer. [3]examine different statistical models and compositions of sensors. [4]keep occupied with statistical analysis of F-formations. [5]deals with the subject of first-person perspective. [6]make some research on how to measure and compare social interactions with individual mental health. Also [7]try to measure social interactions and [8]attached to this work of them two. Measuring the amount of social interaction is done with mobile phone sensing mechanisms by [9],[10], [11].

5.2 Definition

As social interactions we intend the acts, actions, or practices of two or more people mutually oriented towards each other's selves, that is, any behavior that tries to affect or take account of each other's subjective experiences or intentions[4].The absence or presence of social interactions is a cue if an event is viewed to be memorable. So we intend that social interactions are coupled with whether a moment is worth to memory or not[5].This includes not only getting visits from relatives and friend, but also telephone calls and shopping.

5.3 Detection Techniques

To validate hypothesis, [1]employs tools to visualize activity levels in smart environments and identify times and places of resident interactions, further they design an learning algorithm to identify those interactions and last but not least they make use of a supervised learning technique to recognize the current activities. They have two different set-ups using only passive, unobtrusive sensors (e.g. motion sensor) an smart apartment testbed and a smart workplace environment. They collect sensor data in the testbeds and recognized that event density maps provide initial insights into resident behavior patterns. They employ Bayesian updating to track residents in a space and automatically detect interactions. Lastly, they set up a controlled experiment to determine their ability to recognize interactions and caregiver assistance using hidden Markov models. They inspected the data to see the automatically detected interaction and compared it with records of the students living in the apartment. A big problem was the recognition of interaction if the residents moved in close proximity to each other, for example the interaction while moving a couch. A large number of sensor events are suspected to be noisy events, triggered by floor heaters that generate warm air near a motion sensor in the night. They find out a different dynamic in the workplace and the apartment. They use the Viterbi algorithm to label the sensor event observations to compute the most likely sequence of hidden states. The hidden Markov model recognized the activities an accuracy of . To compare they also implemented a naive Bayesian classifier, but only get a accuracy of , it's obvious that the hidden Markov model is a more effective approach for this classification problem. They want to detect interaction between two residents and point out that interaction between groups would be a interesting topic, but is much more difficult.

[2]uses a sociometer, which is an adaptation of a wearable data acquisition board that stores information like identity of the wearer and speech information (only extract speech features and not the content). They detect speech and non-speech region with a hidden Markov model. This works very well with less than error even in an environment which is noisy. A huge problem that can be saved by mutual information of the two voice streams is, that they can't say with certainty that two people interact when they stay in front from each other. For the hand-labeled data-set, the

accuracy for detection conversation is overall and for conversations longer than one minute, this is due to the problem of multiple speaker detection. They also spend time to find out about the turn-taking dynamics within a conversation. The participant with the highest speaking time is considered to hold the 'turn' this is detected by speaker segmentation. They use a two-dimensional influence parameter to model the transition probability of interaction. They build a framework for automatic modeling of face-to-face interactions and integrated methods from speech processing and machine learning to demonstrate the possibility to extract information about patterns of communication.

[3]try to find out a range of plausible sensors using audio an visual data to detect social interactions. They build different statistical models for both individual sensors and multiple sensors using different machine learning methods. For elderly people the frequency, duration and type of social interaction of the patient with one another and their caregivers are an indicator for their health condition. Changes is such patterns can reflect mental and physical status of a person. For example, it's possible to use a radio frequency sensor to track the location of each or a speech detector from audio signals to automatically detect social interactions. They collect data from four cameras and four audio collectors and classify the data with a preprocessing algorithm into shots. They implement a background subtraction algorithm to detect human activities from video channel and a single power-based method for audio event detection. Further they fuse these events together and achieved through that recall and precision. Physical and algorithmic sensors are considered in their study, overall they obtained simulated sensors. To eliminate the non powerful sensors they first used a information gain technique and further a support vector machine method and gets the best fitting model. To find a proper classification model, they evaluated various machine learning algorithms: decision tree, naive Bayesian, Bayes network, logistic regression, support vector machine, adaboost, and logitboost. All of this models obtain good results, the three most important sensors where those for 'walking', 'talking' and 'leaving'. They can achieve high accuracy in detecting current interaction with temporal information. A drawback of the decision tree is that it is sensitive to noise in sensor outputs, the other models are more robust. With the three mentioned sensors the decision tree model achieves a accuracy of . The conclusion for them is that developing perfect sensors for the three terms above is more important instead of constructing complex ones.

[4]select another access, they focus on detection of F-formations. They present an approach for detecting social interactions in a crowded scene by employing only visual cues. The system gets as input the people in the scene and their orientation, with a voting strategy based on Hough transform, it recognized F-formations. This paper attempts to discover social interactions using statistical analysis of spatial-orientational arrangements. They analyze quasi-stationary people in an scenario identifying those subjects engaged in a face-to-face interaction. Their aim is to detect the o-space, therefore they designed a F-formation identifier. This algorithm is based on a Hough-voting strategy, which lies between an implicit shape model, where weighted local features vote for a location in the image plane, and a mere generalized Hough procedure where the local features have not to be in a fixed number as in the implicit shape model.

[5] presents a method to detect and recognize social interaction in a day-long first-person-video of a social event (e.g. picnic). They categorize social interactions in a egocentric video with a method differing between three subtypes: dialogue, discussion and monologue. Two sources are used faces and first-person motion with this information they track and cluster the faces and further compute the orientation and location of every face. They draw up a 3D space with orientation and location of every individual with a MRF model. To solve the temporal problem they used a Hidden Conditional Random Field which includes three features location, pattern of attention and role and pattern of first-person head movement. Attention and location based features perform better at detecting dialogue, discussion and monologue, while first-person motion features perform better on walk dialogue and walk discussion shown in Figure . This method constructs a description of the scene by transferring faces to 3D space and used patterns of attention to assign roles. The roles and locations of the individuals are analyzed over time to recognize the social interactions.

There are many different ways to track social interactions and extract behavioural cues for the analysis with audio/video data, wearable devices and others. [10], [11] mobile phone sensing mechanisms to automatic monitor social interaction. A device that is not built to infer face-to-face social interactions as the mobile phone, does not provide interpersonal distances and body orientations, in contrast to a special designed camera system. The mobile phone requires a complex interpretation of noisy data obtained from available embedded sensors. They were capturing feature vectors composed of interpersonal distances, relative body orientation, and standard deviation of relative body orientation and get a accuracy of . The distance estimation is done by RSSI analysis (Received Signal Strength Indicator). The body orientation is detected by the use of an embedded compass sensor and the standard deviation of this represents an index of stable relative position in a social get-together and recognizes not only social interaction, but also the type formal or informal social contexts. If speech activity is detected with a accelerometer and is embedded in the system they get even a accuracy of 90%.

5.4 Measurement Techniques

[6]measured the social interaction with a Business Microscope. This is a wearable sensing system, it contains a infrared sensor and detects face-to-face contact within a two meter radius. The sensor records face-to-face states as one minute, that means the time resolution of the system is equal to one minute. Face-to-face interaction logs were gathered through the period. Then, the amount of face-to-face contact time between every pair of participants was obtained. To analyze the data they use a social network analysis method. Because of the different sensor wearing time they need to normalize their data. They find out a big correlation in social interaction and individual mental health.

[7] measures social interaction with different methods. They give a good overview of how other scientist have measured social interactions. For example with only using aggregate information. Since social interaction create high levels of variance across space and time, by using the variance of aggregates, one can measure the extent of these interactions. Topa also uses aggregate-level variables to study social spillovers in employment status. The idea of the Topa-model is that the covariance between individuals (the degree of social interaction) is determined by spatial distance. Glaeser and Scheinkman focus on the size and nature of social interaction. The idea of one of their methodologies is to use the relationship between the variance of community level aggregates and the variance of individual data to estimate the size of the social interactions. Their goal is to estimate the parameter which captures the degree of social interaction. This basic formula for this parameter is that where and are the aggregate and individual level variances. They build a discrete model and later a dynamic one. The discrete version is implemented with two different methodologies. On the one hand Social Influence comes only from Unpredictable Elements of Decisions and on the other hand Allowing Control Variables to Influence Interactions. Therefore they estimate the different variances and find out the degree of social interactions. The dynamic model can fit the simulations quite well so this functional form is both reasonable and provides us a convenient measure of the degree of social interaction. [8]discusses how to measure social interactions via group selection. Therefore he tries to modify the model of Glaeser and Scheinkman and set up a model which satisfies the Henrich Equation. Simply a statistical relationship is expressed to measure a social interaction by decomposing both between-groups effects and within-group effects.

[9] uses smart phones microphones to measure social interaction. They recognizing subjects' voice and aimed to quantify the amount of social interactions during a working day. The audio data was split into one-minute samples and speaker was identified using short time spectrum through MFCC (Mel-Frequency Cepstral Coefficient) and GMM (Gaussian Mixture Model). Each minute that contained voice of monitored subjects was considered social interaction. Total amount of social interaction during working day was the sum of minutes in which the subject's voice was recognized.

In the RelaxedCare project a Social Interaction Module will monitor different activities, like voice/video communication and when the person leaves the house or gets visitors with a door sensor.

5.5 Summary

The aim of much prior research seems to be the detection, and not the measurement, of social interaction. Therefore RelaxedCare may have the opportunity to make novel contributions in this domain.

It seems to be state of the art to use Hidden Markov Models for detecting social interactions and most of the groups applied this. Therefore one difficulty is the parameter contribution which is dealt with in different ways. Most of the techniques achieve similar appropriate good results.

6 How to measure emotion

6.1 Introduction to this chapter

The RelaxedCare system can offer the more options the more concise it is able to determine the mood of the participating persons. The aim is to acquire more than the binary states “no emergency” and “emergency.” Being able to measure the mood of the assisted person can provide new options: Interventions from related persons can address sooner and more effective even in so simple cases as e.g. “I am sad because of bad news from my neighbors.” If the RC system could determine this mood, it would send this information the relatives, who in turn could visit or call the assisted person.

The more precise the status of the assisted person can be determined, the more concise the caregiving can be organized and planned.

Measuring emotion is a branch in the scientific community, which began in the 1960s. The interest in this field of research in unbroken, for there are many applications for emotion recognition. Some of which are consumer research, marketing and ergonomics. Consequently, there is a vast body of scientific literature. This chapter shall provide an overview of the facetted emotion recognition landscape and an insight in the state of the art.

6.2 Overview on emotion recognition assessment techniques

There are several methods of how to measure emotional responses. The following overview is structured by the categorization based on the review “Measuring Emotions in Interactive Contexts” from (Kong & Yang, 2009). Each method is described shortly and for further reading, references to scientific publications are provided. It is recommended to revert also to the review “Emotion Recognition – Theory or Practicality” from (Hurst, Jackson, & Glencross, 2012), which is the source of many examples listed below.

Altough it might be difficult for an individual to mask its own emotions, it can not be excluded that they are able to manage it to the extent that they can succeed to circumvent others or the methods listed below. In everyday life individuals could do this e.g. because they care about someone else, or to gain an advantage by masking their emotion (Hurst, Jackson, & Glencross, 2012).

6.2.1 Expressive reactions

Methods based on expressive reactions derive the emotional state of a test subject by analysing facial, vocal and postural expressions.

6.2.1.1 Facial emotion recognition

Deriving emotion from facial expressions seems to be common way in emotion recognition.

In 1971 it was demonstrated, that there is an association between particular facial muscular patterns and discrete emotions is universal across cultures (Ekman & Friesen, 1971). In their study, they reverted to six basic facial expressions/emotions: Happiness, anger, sadness, disgust and fear.

Ekman coded a selection of 44 different muscle movements (referred to as action units) such as a nose wrinkle (AU=9) and a lip stretcher (AU=20). Then a combination of these action units was mapped to each of the six emotions based on his classification system (Hurst, Jackson, & Glencross, 2012).

In their review “Automatic facial expression analysis: a survey” the authors recommend to distinguish consciously between “facial expression recognition” and “emotion recognition.” The difference between both is, that the latter is an interpretation attempt , which demands additional contextual information to be accurate (Fasel & Luettin, 2003). The mechanisms contributing to the facial expression is depicted in Figure 2.

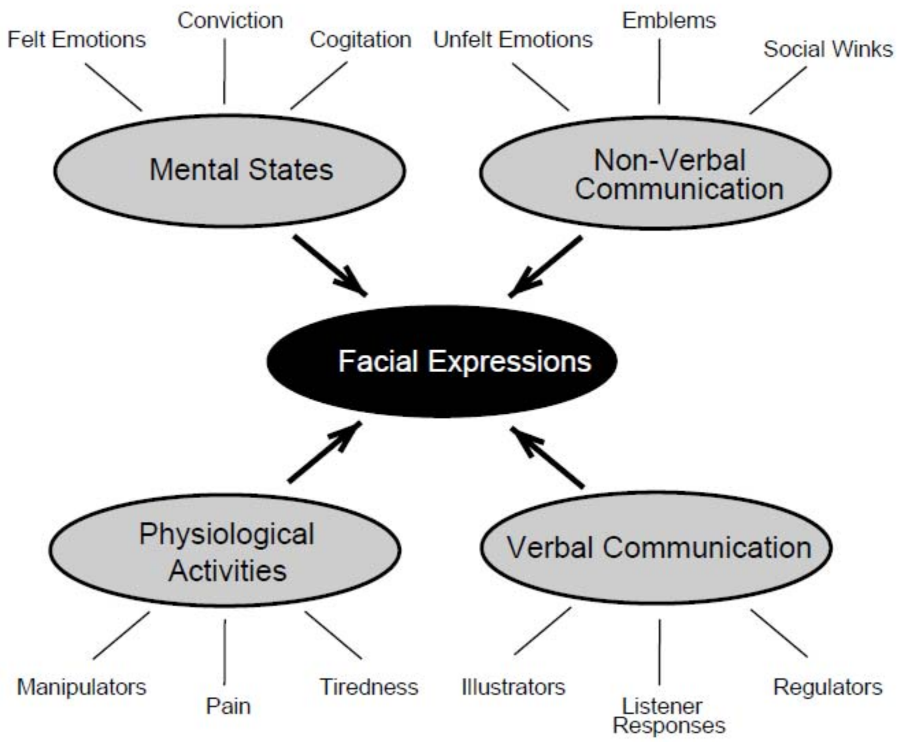

Figure 2 Sources of facial expressions (Fasel & Luettin, 2003)

An application on the field of AAL, utilizing facial expression to derive the six basic emotions, is the Magic Mirror Table. It is a half transparent mirror with built in video camera. It processes the facial expressions of the person in front, and outputs positive words when negative emotions are detected (Yu, You, & Tsai, 2012).

Multimodal approaches

(Chen, Huang, & Cook, 2005) compared the performance of acoustic and facial emotion recognition, and the combination of both modalities. The authors report, that their references video (including audio) tapes were not accurately categorized by human judges (50 % accuracy) under certain circumstances. Further their data supports the thesis, that the accuracy of the automated emotion detection system decreases when emotion categories are increased from 6 (5) to 7. Finally, the bimodal analysis outperformed both, the visual and acoustical analysis when they were applied isolated from each other. The achieved accuracies for are given in Figure 3.

	Neutral	Joy	Anger	Surprise	Disgust	Sadness	Fear	Avg
Visual Only for 6	0.78	0.65	0.87	0.77	0.83	0.89	n/a	0.79
Visual Only for 7	0.78	0.63	0.83	0.77	0.76	0.82	0.86	0.75
Acoustic Only for 5	0.83	0.63	0.82	n/a	n/a	0.67	0.70	0.73
Acoustic Only for 7	0.67	0.46	0.71	0.69	0.69	0.61	0.54	0.63
Direct Bimodal for 7	0.77	0.73	0.91	0.77	0.86	0.84	0.91	0.82
Balanced Bimodal for 7	0.78	0.69	0.97	0.77	0.86	0.91	0.89	0.84

Figure 3 Performance comparison for experiment setup (Chen, Huang, & Cook, 2005).

Many studies on emotion recognition based on facial expressions are interested in the best accuracy of their algorithms. In (Wagner, Andre, & Jung, 2009) the focus is shown on real-time emotion recognition. They introduce a framework, which uses the inputs of a microphone, a video camera and acceleration sensors.

6.2.1.2 Gestures

In comparison with other methods for emotion recognition, literature is sparse. Glowinski pursued the approach to track movements of the upper body extremities (hands and head) to derived their velocities, energy and perimeter, with the aim to distinguish anger, joy, relief and sadness. The authors found that the energy cue is the most significant in differentiating between the emotions.

The introduction of Microsoft's Kinect has elevated the research in gesture analysis. It is likely, that gesture based emotion recognition will profit from this updraft.

<u>Multimodal approaches</u>

Another approach to combine facial expression emotion recognition with a further modality is presented by Gunes. The authors fused facial expressions with gestures of the upper body and show, that this combination outperforms the isolated facial expression analysis. (Gunes & Piccardi, 2005)

6.2.1.3 Eye gaze

Eye gaze information as stand-alone method to conduct emotion recognition is very sparsely in literature. The review from Hurst, it is reported that the eye tracking technology has advanced drastically over the last decade. The effects used for emotion recognition are proximity and velocity. Proximity can be used to determine whether the individual is comfortable at looking at stimuli or not. On the other side, rapid moving eyes (i.e. measuring velocity) can be an indication for being in high alert (i.e. arousal emotion) situation (Glowinski, Camurri, Volpe, Dael, & Scherer, 2008). (Hurst, Jackson, & Glencross, 2012)

Multimodal approaches

Combining eye gaze in multimodal systems for emotion recognition seems to be rare. The few publications interconnect eye gaze with facial expression analysis (Zhao, Wang, & Petriu, 2011) or EEG and pupillary response (Soleymani, Pantic, & Pun, 2012).

Gaze-X is a software package, which relies eye gaze direction, speech, face and facial expression, and several standard HCI modalities, such as keystrokes, mouse movements and active software identification. (Maat & Pantic, 2007)

6.2.1.4 Textual emotion recognition

A separate field of research is the emotion recognition based on text analysis. Examples of implementations are from (Masum, Helmut, & Ishizuka, 2006) who developed a cognitively based approach to recognize emotion in text inputs. They also used their algorithm to interact (answer) to user input. An improvement of this approach was published by (Li, Pang, Guo, & Wang, Research on textual emotion recognition incorporating personality factor, 2007). Their result shows an increase in emotion recognition accuracy when the personality of the user is incorporated.

6.2.2 Physiological expressions

6.2.2.1 Brainwaves

Electroencephalography (EEG) is a wide spread, established method in research and with clinical applications. It provides the means to analyze the brain activity by detecting the changes in electrical potential caused by neuronal activity.

EEG based emotion recognition while listening to music seems possible. (Lin, et al., 2010) extracted from a study 30 subject independent features to classify emotions as joy, anger, sadness and pleasure.

A real time emotion recognition algorithm (fractal dimension based) is introduced by (Liu, Sourina, & Nguyen, 2010) to distinguish between 7 emotional states (sad, frustrated, fear, satisfied, pleasant, happy) und usage of only 3 electrodes.

The EEG headset “Neurosky Mindset” (see Figure 4) was assessed by (Maki, Sano, Kobashi, Nakamura, Kanoh, & Yamada, 2012). Compared to a standard EEG, it is more unobtrusive and handy to use. In their experiment, they asked test subjects to alternately rest and perform arithmetic calculations (stress situation). Furthermore they were asked to document their mood state during the experiment by selecting the corresponding face on a face scale (see Figure 5). The study led to the result, that the correlation between face scale and measurement was not sufficient, while the mental states could be assessed with an average accuracy of 83 %.

Figure 4 NeuroSky MindSet (Maki, Sano, Kobashi, Nakamura, Kanoh, & Yamada, 2012)

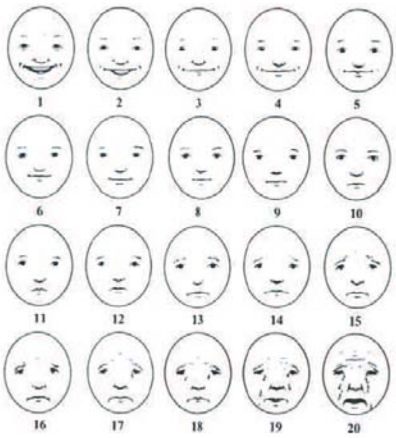

Figure 5 Face scale (Maki, Sano, Kobashi, Nakamura, Kanoh, & Yamada, 2012)

Multimodal approaches

An approach to recognize emotions (arousals) by combining event related potentials (19 channel EEG, stimulus International Affective Picture System (Lang, Bradley, & Cuthbert, 1999)) and skin conductance. The authors recommend to improve the system by including further modalities, such as ECG and EMG (Frantzidis, Lithari, Vivas, Papadelis, Pappas, & Bamidis, 2008).

6.2.2.2 Brain Oxygenation Levels

The success of neuronal functional imaging studies includes incorporates also emotion recognition. Results of the review “Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI” (55 PET and fMRI studies) are:

1. The medial prefrontal cortex had a general role in emotional processing;
2. fear specifically engaged the amygdala;
3. sadness was associated with activity in the subcallosal cingulate;
4. emotional induction by visual stimuli activated the occipital cortex and the amygdala;
5. induction by emotional recall/imagery recruited the anterior cingulate and insula;
6. emotional tasks with cognitive demand also involved the anterior cingulate and insula.

(Phan, Wager, Taylor, & Liberzon, 2002)

Due to the properties of light propagation in biological tissue, neurofunctional studies based on near infrared spectroscopy (fNIRS) are especially suitable for the cortical regions of the brain. In accordance with the findings cited above, fNIRS can also be a method for emotion recognition. One example is the study “Enhancement of Activity of the Primary Visual Cortex During Processing of Emotional Stimuli as Measured With Event-Related Functional Near-Infrared Spectroscopy and Event-Related Potentials” from (Herrmann, et al., 2008).

6.2.2.3 Heart Rate / Blood Flow

It was shown that heart rate increases in as the result of increased adrenaline levels. Monitoring heart rate can therefore be a method to measure favorably classes of emotions as distress, arousals and excitation. (Hurst, Jackson, & Glencross, 2012)

Multimodal approaches

A study in a driving simulator (Kato, Kawanaka, Bhuiyan, & Oguri, 2011) used the electrocardiogram (ECG) and the pulse wave to recognize negative (traffic jam) and positive (dissolving traffic jam) emotions. In further section (section 6.2.2.6, 6.2.2.4) there are also several examples of multimodal approach incorporating ECG.

(Tsai, Chen, & Lo, 2009) achieved an absolute accuracy of 92.5 % by combining blood volume pule, skin conductivity, respiration rate and electromyography. Their self-developed learning algorithm, which classified between pleasant, unpleasant, low arousal and high arousal.

6.2.2.4 Muscle Contraction

As the brain controls many muscles in the human body, analyzing the brain is indirectly possible by measuring activity in muscles. Electromyography can even pick up very small muscle activations, which are not visible for the naked eye (Hurst, Jackson, & Glencross, 2012). While measuring specific muscle contractions is useful in many research applications, its echo is rather small in the field of emotion recognition.

Multimodal approaches

Facial electromyographic (EMG) activity of the zygomatic major (“smile”) and corrugator supercilii (“frown”) muscle regions was recorded in combination with the heart rate. The latter were acquired to be able to detect the arousal component of the emotional reaction. Their goal was to study the relationship between self-reported and physiological reactivity to pleasant, unpleasant and neutral emotion-eliciting stimuli and experiential avoidance. Findings indicated, that high experiential avoidance participants reported greater emotional experience to both, unpleasant and pleasant stimuli. (Sloan, 2004)

6.2.2.5 Speech (acoustic, voice) emotion recognition

As speech is a fast and natural method of communication between humans. Researchers and scientists in the field of HCI realized, that speech could be an effective means of communication with computers. However, bare voice recognition is not sufficient, since it lacks the emotional content inherent in speech. Therefore a new field of research emerged. An overview on that field is given in the “Survey on speech emotion recognition: Features, classification schemes, and databases” (Ayadi, Kamel, & Karray, 2011) and in “Behavioral Signal Processing: Deriving Human Behavioral; Informatics From Speech and Language” from (Narayanan & Georgiou, 2013).

In the study "Detection of Negative Emotional States in Real-World Scenario" from (Kostoulas, Ganchev, Mporas, & Fakotakis, 2007) speech emotion recognition is used to detect negative emotions of test individuals in a smart home environment. However,
with an accuracy of 50 % - 60 % there is room for improvement, the authors point out that it is even for humans difficult to judge only from speech.

A military study followed up the question, whether it is possible to recognize emotion in the natural speaking voice. They focused on detecting changes in the voice when mental stress situations occur. Their interest was to learn whether this technique could be utilized to screen the mental status in military situations. As final result, they concluded that further development is necessary. (Tokuno, et al., 2011)

Hassan et. al. claim in their publication (Hassan, Damper, & Niranjan, 2013) to be able to significantly improve the speech emotion recognition in many circumstances through modelling channel and speaker differences as covariate shift.

One real time recognition software package, available under GNU-Licence, is downloadable under http://www.informatik.uni-augsburg.de/lehrstuehle/hcm/projects/tools/emovoice/ (Vogt, Andr{\'e}, & Bee, 2008)

6.2.2.6 Skin Conductance

Emotions of an individual can trigger change in skin conductance. As perspiration is water based and a good conductor for electrical currents, skin conductance can be higher in nervous persons, than in happy persons. (Hurst, Jackson, & Glencross, 2012)

Today, skin conductance measurements are mostly employed as one modality in a multimodal system.

Multimodal approaches

(Lundqvist, Carlsson, Hilmersson, & Juslin, 2009) researched the emotional response to music. The authors found out, that "Happy music generated more zygomatic facial muscle activity, greater skin conductance, lower finger temperature, more happiness and less sadness than sad music."

In another study heart rate, skin conductance and skin temperature were continuously monitored to deduct and an algorithm (based on clustering techniques) for the recognition of emotions was developed. The authors hope to be able to discriminate between the emotional states happy, sad and angry in the future. (Quazi, Mukhopadhyay, Suryadevara, & Huang, 2012)

The same modalities (ECG, galvanic skin response and skin temperature) were used by (Wu, 2010) to determine due to their newly developed support vector machine, in real time, the emotional state of individuals. They achieved an average accuracy for all models of about 70 %.

By acquiring skin conductivity, ECG, photoplethysmography (PPG) and skin temperature, (Jang, Park, Kim, Eum, & Sohn, 2011) developed a support vector machine algorithm, which achieved an accuracy of 99.05 % in classifying the measurements into the groups happiness, sadness, anger, fear, disgust, surprise and stress.

Kim et. al. chosen an emotion-specific multilevel dichotomous classification approach to classify physiological signals acquired during music listening. They reverted to measuring electromyogram (electrodes located at the upper trapezius (near the neck)), electrocardiogram, skin conductivity and respiration changes. In their finding, they report an emotion recognition accuracy of 95 % for subject dependent and 70 % of subject independent classification into positive/negative high arousal, positive/negative low arousal. (Kim & Andre, 2008)

Further multimodal approaches that include skin conductivity can be found in section 6.2.2.3 .

6.2.2.7 Other approaches to recognize emotions

6.2.2.7.1 Music

An interrelation of music and mood has been proven. A vast body of literature exists on the topic of how to determine the mood of a song as stated in the review “MUSIC EMOTION RECOGNITION: A STATE OF THE ART REVIEW” (Kim, et al., 2010) . The effect of music on an individual depends on many variables, among which there is also body state (Dibben, 2004) and personality (Vuoskoski & Eerola, 2011). Reversing the question (i.e. can music be an indicator for the individuals mood) seems not to have given much attention in literature. In one, very recently published, study, the question was posed whether the individuals short-term music preference may reveal the emotional state of the individual (Yang & Liu, 2013). Basis for their examinations was an online repository where bloggers can attach music titles to their post (see Figure 6). A result of the study is presented in Figure 7, showing that people with different personalities prefer different music even when being in the same mood.

The authors mention also methodical imperfections in their set up (i.e. there was no way to verify, whether the blogger was listening to the music identified by the tag while writing the blog entry).

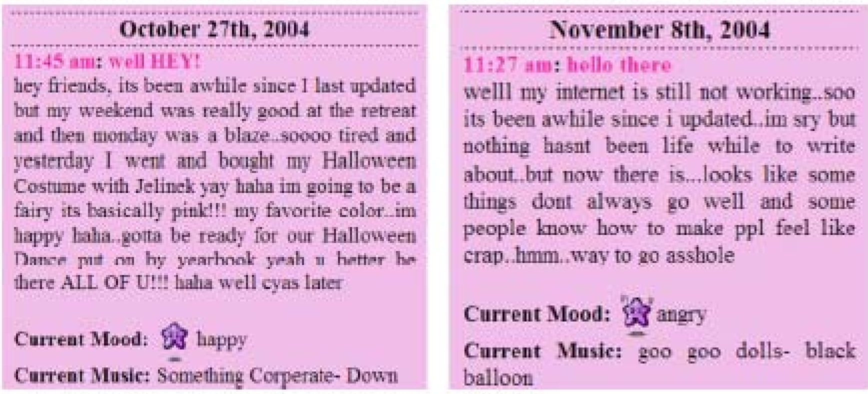

Figure 6 Two blog entries of a user with mood and music tags attached (Yang & Liu, 2013).

Personality	User mood	Music emotion
High in Extraversion	loved	more party music
High in Extraversion	sick	more party music
Low in Extraversion	anxious	less angry music
Low in Extraversion	awake	less angry music
Low in Extraversion	blank	less party music
Low in Extraversion	bored	less angry music
Low in Extraversion	bored	more peaceful music
Low in Extraversion	cheerful	less angry music
Low in Extraversion	drained	less angry music
Low in Extraversion	drained	more peaceful music
Low in Agreeableness	chipper	less sad music
Low in Agreeableness	chipper	more angry music
Low in Agreeableness	cold	less sad music
Low in Agreeableness	frustrated	less sad music
Low in Agreeableness	lonely	less sad music
Low in Agreeableness	lonely	more angry music
Low in Conscientiousness	depressed	less sad music
Low in Conscientiousness	depressed	more angry music

Figure 7 Preference of certain music as function of user personality and user mood (Yang & Liu, 2013).

6.2.2.7.1.1 Acceleration

A more unobtrusive approach to perform emotion recognition is suggested by Lee et. al. They limited their sensor inputs to the accelerometer of a smartphone. They claim to be able to classify 7 user emotions (happiness, surprise, anger, disgust, sadness, fear, neutral) with high accuracy. Figure 8 depicts the concept of emotion oriented social communication. (Lee, Choi, Lee, & Park, 2012)

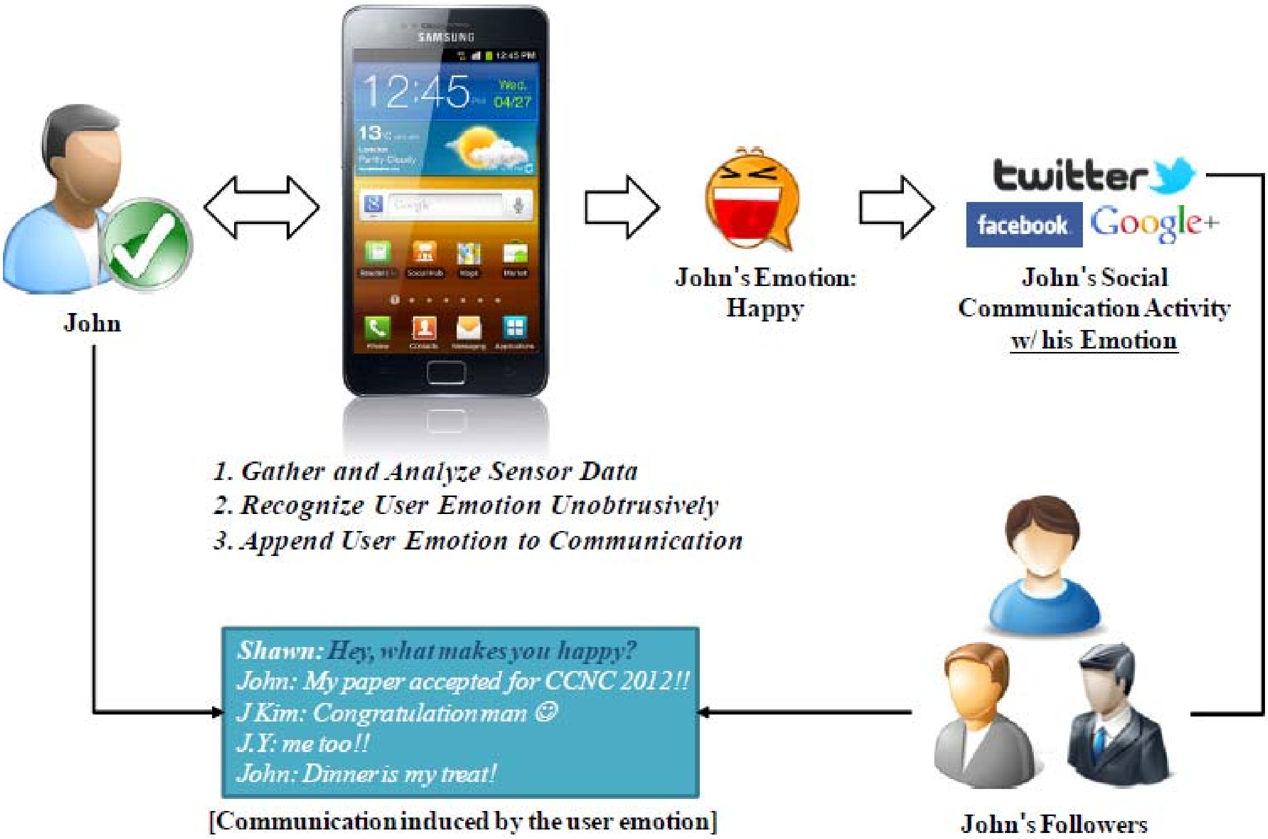

Figure 8 Concept of affective social communication. Note: Acceleration sensor of smartphone is only sensor in use (Lee, Choi, Lee, & Park, 2012).

6.2.2.7.1.2 Emotion prediction

While this chapter on emotion recognition gives an overview on the most relevant and/or creative methods to determine the emotion of an individual, in the context of RelaxedCare it can also be of interest to predict emotion.

(Zhang, Tang, Sun, Chen, & Rao, 2010) used information from a mobile social network to forecast the mood of users. With access to all-day communication (by SMS and call), calendar, alarm, WIFI signal, GPS location, activity and (self-reported) mood information, the authors developed a dynamic continuous factor graph and achieved an accuracy in prediction of emotion of about 62 %.

6.2.2.7.1.3 Weather

There are studies, which have correlated weather and psychological changes. A relevant publication is (Keller, et al., 2005), where “pleasant weather (higher temperature or barometric pressure) was related to higher mood, better memory, and a ‘broadened’ cognitive style during the spring as time outside increased.” Unfortunately, there were no significant findings for the other times of the year, except that hotter weather in summer resulted in lower mood. They point out, that there are parallels with findings on the seasonal affective disorder.

In an online diary (N>1200) study weather station data (temperature, wind power, precipitation, air pressure, photoperiod and sunlight) was linked to mood (positive affect, negative affect, and tiredness). The scientists, who conducted the study could not proof “the idea that pleasant weather increases people’s positive mood in general.” (Denissen, Butalid, Penke, & Van, 2008). An effect of sunlight on tiredness. In sunlight the production of vitamin D_3 leads to increased levels of serotonin, which could account for changes in mood, was significant. Also significant were main effects of temperature, wind power, and sunlight on negative affect were found.

6.2.2.7.1.4 Localization

In an experimental setting of a case study, several PIR sensors were distributed in an apartment-like space within a hospital. By analyzing the movement pattern of the patient within the apartment, it was found that a high rate of activity corresponds to periods of great complaining. (Noury, et al., 2008)

6.2.3 Self-report of subjective experience

Although all of the methods listed below have their own advantages and limitations, they share also common ones. While self reported emotions are likely to be more valid because the subjective interpretation of the emotions can be acquired. However, the common limitation is that not all test subjects are capable of reporting, forget the reporting their emotional states (Kong & Yang, 2009), or report what they think is expected from them.

6.2.3.1 Verbal

It is claimed that subjective feelings (subjective emotional experience) can only be measured through self report. Often the Osgood’s semantic differential technique has been used (e.g. 1=very fun, 2=fun, 3= neutral, 4=boring and 5=very boring).

The limitations of this technique is that emotions are difficult to verbalize in some cases, and finding semantic opposites can also be a challenge (Kong & Yang, 2009).

6.2.3.2 Pictograms

To overcome the limitations imposed by verbal self reporting, pictograms can be utilized

Figure 9 Pictograms for non-verbal measurement of emotions (Kong & Yang, 2009).

(see Figure 9). Pictograms bridge also the language gap and are therefore cross cultural deployable. However, this approach lacks the practical possibility to measure distinct emotions.

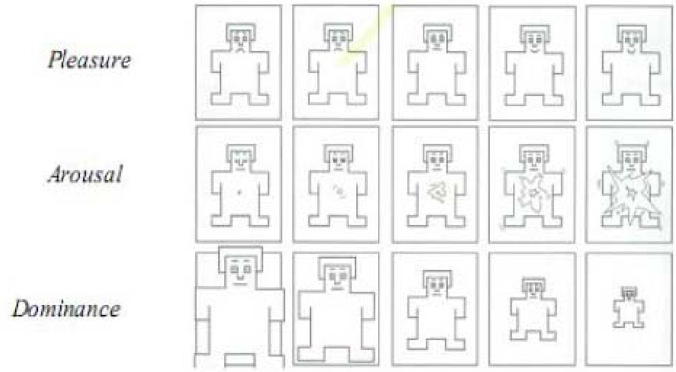

Figure-1 SAM

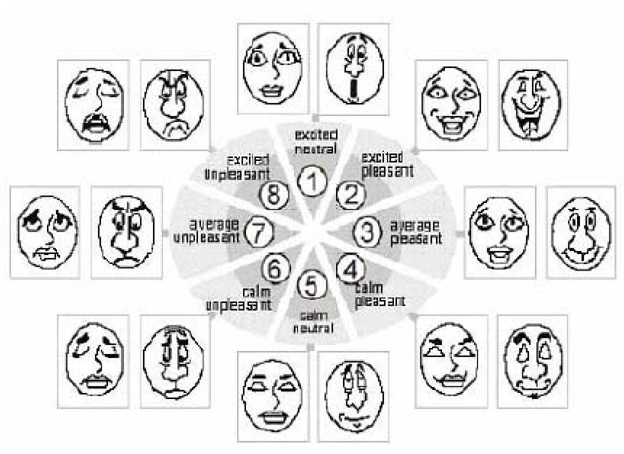

Figure-2 Emocard

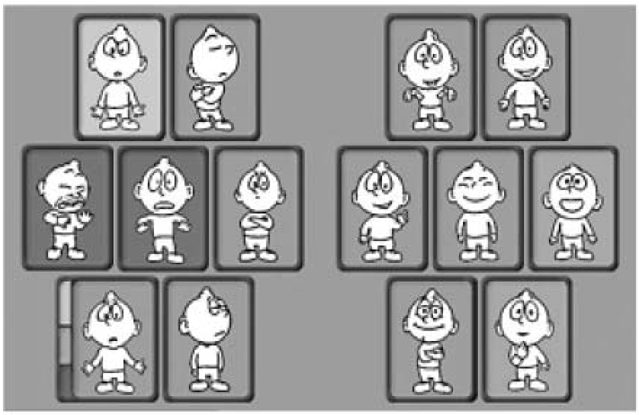

Figure-3 PrEmo

Therefore pictograms are commonly utilized for reporting generalized emotional states (Kong & Yang, 2009).

6.3 Measuring mood and well-being

Emotions, as they are detected by almost all the literature cited here, are events of shorter duration. In the scope of the RelaxedCare project, it seems more relevant to measure mood. Mood is considered the general emotion, an individual has for longer periods of time, e.g. during the day. Technical solutions to detect the mood are scarce, or revert to classical emotion recognition methods. E.g. (Peter, Waterworth, Waterworth, & Voskamp, 2007) in the publication “Sensing Mood to Counteract Dementia” integrate electroics in furniture (see) and attempt to interpret the following cognitive states

“The patient is:

- interested in doing “something”, e.g. when looking for the games box in the cupboard

- in need of help, for example when groveling over the floor with the pills drawer opened

- interested in communication, e.g. by sitting eagerly in front of the turned off multipurpose display or playing with the mobile phone

- needs attention, for example when sitting in the armchair with apathetic face, eyes opened, and low physiology - would like some attention, e.g. when sitting in the armchair,

looking at the turned off display, with physiology aroused and active face, limbs and fingers

- in need of quietness, e.g. when sitting in the armchair, looking at the turned off display, with physiology being low, active face and passive limbs and fingers;

- being entertained, e.g. when actively watching a film

- needs changes to her daily routine of the patient."

Instead of searching for literature in the area of emotion recognition, there were also searches performed on “well-being”. However there are plenty of publication available, “well-being” rather refers to something as the mood in life. Well-being is assessed most of the times with questionnaires. The questions therein have the character of “are you content with your life, so far?” and “do you have a purpose in life?”.

6.4 Discussion

The overview on modalities employed for emotion recognition is widely facetted. While facial expression recognition and emotional interpretation is a well-established method in this field of research. With RelaxedCare in mind, all detection modalities, which do not rely on body worn sensors should be preferred. Although facial emotion recognition does not require the user to wear body attached devices, the installation of cameras in the environment of users is delicate from the perspective of user acceptance.

Another contactless and wide spread method in emotion recognition is speech analysis. Although speech recognition is often addressed by publications, the experimental set-ups in laboratories do not represent real-life environments, where different voices from radio and TV and background noise is present often. To enhance the robustness of a such a recognition system, a voice recognition should precede the emotion recognition, ensuring that only the speech of the user is analysed.

Even if installations of microphones in apartments is less critical than with cameras, it will be most likely far from being accepted easily by the user. This means, if cameras and microphones are to be utilized, the underlying user-design must not be underestimated. Thinkable are solutions as a magic mirror, which is confined to a clearly defined area within the apartment, and only activated when the user stands before it. Thus it is not recording the user continuously. Obviously, such a solution will prevent a system, which is capable of monitoring the emotions of the user continuously.

Despite of the challenges mentioned, it is good practice to combine multiple modalities and contextual information. The latter comprises a range of possibilities, from keystrokes on the computer's keyboard, body movement, perhaps weather, to the personality of the user. As the review of (Kong & Yang, 2009) concludes: no single emotion detection method can be regarded as the "gold standard." Each isolated method has its advantages and weaknesses, which lead to an incomplete picture of the underlying emotion processes.

Body worn sensors have to be attached by the user to himself. While this could be regarded as a disadvantage, these sensors offer access to all of the physiological modalities introduced in the subchapters before. By attempting to be in accordance with the idea of unobtrusiveness of the RelaxedCare system, it is difficult to see, how even a partial EEG can be derived. The same holds true for the determination of brain oxygenation levels. Attaching devices as MRI scanners to users is next to impossible at the present time. PET scanners suffer from the same problem, and on top of that, the user must be administered a marker. This leaves ECG, heart rate, EMG and skin conductance as most reasonable candidates for emotion recognition based on physiological signals.

These physiological signals could be combined with the expressive reaction (facial expression, speech, gestures, eye gaze). If the RelaxedCare system has access to writings of the user, textual emotion recognition could be a further useful modality.

Promising are also methods based on movement patterns of the user inside the apartment as measure of well-being. Another creative approach was to use the accelerometer date of smartphones for emotion recognition.

Rather powerful and simple are the ways of self-reporting emotions. If all else fails, the user can be asked to report his emotions. There are limitations of this method, such as an user unable to articulate his emotions, or the limited resolution/pool of possible choices in case pictograms are used.

The unrivalled advantage of self-reporting is the possibility to get subjective emotions. In context of the RelaxedCare system, it cannot be excluded that users do not report what they feel, but what they think is expected from them. Motivations users could have are e.g.: "I feel very bad, but I know how little time my relatives have; I do not want to be a burden for my relatives, so I pretend to be OK" or "I would like to get more calls. If I pretend to be very depressed, more relatives will call me."

In general, it is unclear, how easily emotion recognition systems can be deceived. In fact, several authors of publications on expressive reactions were reverting to actors in order to develop their emotion recognition algorithms. If an actor can pretend emotions, without actually feeling them, and convince a emotion recognition algorithm, then it cannot be excluded that users will exploit this possibility.

Besides the question of how robust the systems are against attempts of deceptions, a more fundamental question is how reliable are the systems? Due to the lack of clear information on accuracy, it is difficult to report a percentage number. However numbers reported vary between smaller 50 % to more than 90 % accuracy, the set-ups of the experiments are hardly comparable and practically always under laboratory conditions. Multiple simultaneously voices, background noise, insufficient contrast/light within video pictures, faces from different angles, moving persons, etc. are only rarely accounted for – if at all. It is safe to conclude, that under less than good conditions, the accuracy will drop quickly. Possible measures to counteract this trend could be, to minimize the number of emotion categories and/or broaden them. In context of RelaxedCare, it could perhaps be more important to focus on well-being in general. It could also be argued, that detecting "not being well" is of higher priority than being well, for the consequences could be more severe and need to impair is more latent. E.g. it could be considered much more helpful to call a person when he is not well, instead of calling the user when in high spirits.

6.5 Summary

Emotion of humans is interconnected with physiological expressions (detectable by EEG, fNIRS, fMRI, PET, MEG, ECG, skin conductance, etc.) and expressive reactions (facial expression, gestures, speech, etc.). It can also be obtained by self-reports. However, creative approaches comprise the attempts to derive emotion from acceleration data (smartphones), movement in apartments, music listening behavior and weather, it becomes obvious that these modalities alone will not be sufficiently accurate. Thie impression is reflected also in literature, which indicates, that emotion recognition remains even today a challenge. To enhance the accuracy, there is a trend to multimodal measurements. Incorporating contextual and personality increases the chance of further increasing the accuracy.

In the scope of the RelaxedCare project, it seems recommendable to base the emotion recognition on many sensors (ambient and body worn) and much contextual information as possible and practical, and use the pattern recognition approaches to correlate emotion with sensor inputs, to filter out the relevant input modalities for future commercial products. To increase the potential of obtaining satisfactory results, it is further recommended to limit the number of different emotion categories and demands on continuous emotion recognition. Perhaps it could be sufficient to focus more on negative/bad emotions during half a day, instead of attempting to recognize the full range of possible human emotions with a high time resolution.

Based on the goals and possibilities of RelaxedCare, utilizing a wide range of sensors and contextual information, spanning from presence detectors, over ECG belts, number/time of visitors, the weather, number and time of incoming/outgoing phone calls, to the agenda stored on the computer of the user, the project is rather likely to contribute a deeper insight into the applied research of emotion recognition, than to be able to base the emotion recognition on commercially available technological solutions.

7 How to measure inactivity / activity

7.1 Introduction to this chapter

Parallelling the progression from low- to high-level described in **Fehler! Verweisquelle konnte nicht gefunden werden.**, activity measurement may be construed with varying levels of abstraction. We will consider the idea at several such levels, but ideally RelaxedCare will be "aware" on all of them.

7.2 Basic Activity Level

On the most basic level, we are interested in the presence or absence of any activity on the part of the elderly. To this end, we could easily employ any available sensors which detect state changes, such as infrared motion detectors, accelerometers (which may be worn, embedded in cell phones, or attached to household objects), door contact sensors, and the like.

On a slightly higher level, we may be interested in not just in a binary "active" vs. "not active" designation, but further in an estimation of *how* active the person is at present. One can imagine a variety of simple solutions to this, given the type of data considered in the previous paragraph. By looking at per-time period statistics (e.g. number of events) or event frequency, various measures which are proportional to actual physical activity could be realized.

Of special importance are accelerometers, which may be worn (e.g. in a wrist watch) and are also included in most smart phones. *[12]*gives an overview of the available technology and the considerations involved in using the resulting data for activity monitoring. *[13]*does an analysis of the reliability of several off-the-shelf devices. And *[14]*demonstrates a successful application of this, showing that triaxial accelerometer data, integrated over time slices, is highly correlated with energy expenditure (which is a standard measure of physical activity).

7.3 Activity Level Over Time

Given a simple measure of activity as above, the next obvious question is how this changes over time. Most importantly, perhaps, is whether or not today's activity level is different from usual, or different from yesterday's. Again, a variety of approaches could be applied to this. From a time series perspective, we could consider parametric modeling of past data to compare predicted vs. actual current state (*[15]*provides background for many of these methods) or clustering series (*[16]*gives a survey of algorithms and similarity measures).

Another approach is to consider the sequence and apply a graphical model (exhaustively examined in *[17]*), either with directed or undirected nodes (i.e. either Bayesian or Markov) with one or more latent switching states (perhaps representing e.g. 'high activity', 'low activity', or 'alarm') underlying the observables. *[18]*gives a semi-supervised approach to this using HMM's.

7.4 Activity Recognition

At the most abstract level, we have the actual identification of specific activities, such as walking, watching television, etc. As stated in the introduction to this section, this is further removed from the raw data, and is thus a more complex task. However, since it is of great interest within the AAL and ambient intelligence fields in general, much work has been done in this area, particularly with supervised learning.

Several notable examples include:

- *[19]*, who collected data from 20 subjects performing 20 different activities in a lab setting, which they were then able to predict with fair accuracy (although they noted that certain types of activities were harder to generalize across subjects, and hypothesized why)

- *[20]*, who collected data from a couple living for 10 weeks in MIT's Place Lab, a custom apartment outfitted with over 900 sensors of diverse types, and then experimented with activity classification, examining minimal and most-informative sensor subsets
- *[21]*and *[22]* who brought the data collection into a real home environment, with the latter having especially good classifier accuracy
- algorithmically, *[23]*shows a boosting method which gives improved performance to previous results on common benchmark datasets; and *[24]*develops a model for "zero-shot" learning — a form of semi-supervised learning where training occurs on labeled examples, but where new previously unseen activities may be encountered during testing.

Note that many of these datasets are now available to researchers — these will be an indispensable resource in developing algorithms for RelaxedCare.

7.5 Summary

Perhaps more easily definable than problems related to social interaction and emotion detection, detection of activity level and – especially – of specific activities is another complex task, the solutions to which must also employ varied approaches. Fortunately, as a very active field of research, there exists a lot of precedent in this direction, and we can make great use of prior knowledge in construction our solution.

8 Conclusion / Further Work

8.1 Challenges

Creating a system as described above presents numerous technical challenges. Some of these arise from an essential opposition implied by the requirements themselves. In particular, there are forces (cost, complexity of installation, use, and maintenance, privacy) pushing towards a paucity of data on the input side; but these are in conflict with the desire to have a rich hierarchy states which can be somewhat confidently detected on the output side.

The transparency and ease-of-use requirement seems to restrict us to "ambient" type sensors (motion, temperature, etc.), and probably excludes more intrusive (but still passive) sensors, such as wearable ones, as well as more active input devices like touchscreens, etc. Possible exceptions to this could include mobile phones and/or watches (since these are things that many/most users may reasonably be expected to already be carrying in 5+ years). Furthermore, the need for strict privacy means that even more richly detailed passive sensors such as video or sound are also excluded — we may only be able to collect data which are low-frequency and quite low resolution (often, perhaps, only binary). Since there are only so many such sensors which can be placed before returns steeply diminish, this could also restrict us to fairly low-dimensional input overall, as well.

Considering the effects of the no-effort and full-privacy requirements together, we also derive another limitation on the data: it will likely be unlabeled (unless we can arrange to collect a small set of labeled data for training and research purposes)[4]. This means that we might never, for any portion of sensor data, know what "true" states (e.g. eating, visiting with friends, sleeping, in some emergency state) correspond to that sensor data: there may be no way (e.g. input by touchscreen or cellphone app, keeping a log) for the user to tell us, and there is no way (e.g. reviewing video) for us to find out ourselves. This makes it considerably harder to apply some probabilistic methods to the state detection task, and severely limits the types of inferences we can make, as well as the certainty with which we can make them.

Thus our basic requirements of privacy and non-intrusiveness have significant impact upon our potential input data which, in turn, has an inescapable limiting effect on our outputs. Progressing through the classification outlined in **Fehler! Verweisquelle konnte nicht gefunden werden.** above, from low- to high-level states, we'll find increasing difficulty in reaching practically useful conclusions with a fair level of certainty (which is important, since we also want to be non-intrusive with the caregiver, and not bother them with spurious alerts). The low-level states don't pose much of a problem: they are still close to the data and can be inferred mostly deterministically, with only slight uncertainty arising from noise. On the other end, the high-level states could be nearly impossible to detect with the sort of data described – they are abstract and very far removed from the data.

Another challenge arises from the essential conflict between genericness and inferential complexity. Ideally, a new customer would install the sensors and plug them into our "black box", and we would begin usefully determining status and detecting problem states. However, what is normal daily behavior for one person might be quite abnormal for another. And even more concretely, different apartments, installations, sensor types, and even individual sensors will all have idiosyncratic behavior. Therefore, we must employ methods which are generic enough to cope with these differences. On the other hand, though, the more generic the algorithm, the less abstract and confident the outputs can be. For example, "not moving for n hours" looks roughly the same for any person with any motion detectors in any apartment. But "depressed" might look drastically different for a person who becomes agitated and paces between his living room and kitchen, versus another who stays in bed all day — and it could be quite hard for us to say what is

4 For a further discussion of labeled vs. unlabeled data, see **Fehler! Verweisquelle konnte nicht gefunden werden.**.

happening with any confidence, without having labeled examples of "depressed" sensor data from specific individuals.

An overall challenge arises from the general complexity of the task we've described so far. There are many ways to approach this problem theoretically, but many will probably not be very good, while a few could be surprisingly accurate. Since it may not be trivial to see in advance which will be most fruitful, we will likely have to sift through lots of solutions in order to find the good ones.

8.2 Keys to success

8.2.1 Define Success!

Although it may seem obvious, this is a crucial step in focusing effort towards the requirements. For instance, a requirement to "detect if there is a moderate drop in activity level" is very different from a requirement to "detect a moderate drop in activity level which is probably due to illness or depression". Both depend first upon defining activity level in terms of available sensor data, and then being able to robustly estimate these in real time despite lost data, infrequent updates, etc.; but the latter requirement adds a quite non-trivial extra element, that of determining the *reason* for the drop, and for establishing some notion of confidence in our assessment. This type of inference requires a whole additional set of machinery, and introduces another significant layer of uncertainty (and thus error) into the process.

This example also shows how our definition of success will depend very heavily on whether we have (at least some) data with labels pertaining to the differences we wish to detect. In the above example, if our data include examples of activity drops which we are clearly told are due to depression, illness, excessive heat, etc. then we can test our accuracy on predicting these from the sensor data. If, on the other hand, we know nothing about the conditions under which some observed drops occur, we can only employ some mixture of human review of the data, clustering seemingly-similar segments into groups and attempting to guess what the groups are, and employing (semi-)deterministic tools like rule sets or automata to try to pin down the underlying causes. In this case, our success measure may also have to contain more of a human element — do our conclusions seem optimal to us? Or at least good enough?

In any case, once "success" has been as clearly defined as possible, hopefully using some concrete error measures, it is essential to begin measuring the exact performance of attempted solutions (and to do so over a reasonable range of the parameter space, when appropriate). When evaluating many competing models, it is easy to get lost in the details of individual solutions; but the only thing that matters is how well they are able to take information from old data and make the required inferences on new data with the lowest error possible.

8.2.2 Iterative Design

Attaining an elegant, modular design to accommodate the sorts of behaviors we require is probably a difficult problem, and we may not have much prior work for reference. Furthermore, due to the nature of the challenges described above in **Fehler! Verweisquelle konnte nicht gefunden werden.**, it is likely that our specified requirements will change as a result of discoveries and pitfalls during design, implementation, and testing. For this reason, the so-called "waterfall" development model — which tries to do a complete specification, followed by a complete design, followed by a complete implementation of that design, etc. — may not be a fruitful path. Instead, an iterative, incremental design process could yield more elegant results with less total effort.

As applied in the "lean" methodologies, this method focuses on adding just one feature at a time to the existing system, adding only enough code to make it work. One should only refactor or increase the level of abstraction when doing so clarifies the relationships between entities, or simplifies the process of adding new features.

The benefit of this is that, rather than spending lots of time and effort upfront trying to guess an optimal solution to a problem that you don't yet fully understand, given requirements that will probably change anyway, you instead offload this effort into smaller pieces which come up only at the times when you better understand and are prepared to tackle them. In this problem-centered approach, the software abstractions, such as entity boundaries and hierarchies, are allowed to emerge from the naturally from the problem context, instead of being imposed upon it. This can often lead to a simpler, more elegant design, which is only as complex as it needs to be at present.

8.2.3 Collaborate on "input" side

As should be clear by now, both the definition and success of this task rely heavily upon the type of input data we're given. Thus, it is crucial that we have two-way communication with those who are specifying the hardware and software platforms. From them, we need to understand the types of sensors that will be used and how many there will be. Another important question is the parameters of on installations in general: how will they be different or the same from each other?

Another question to pursue is whether it would be possible to collect some labeled data, possibly using a small pilot group of users who agree to either track certain activities, or to allow us to do so. This might require some additional hardware or software, for instance a mobile device, wearable sensors, NFC sensors, etc. — and these might be most efficiently handled by those whose expertise lies in this area.

In general, we need to communicate the heavy dependence of the final product upon the richness of the inputs, and to facilitate the move in that direction wherever possible.

8.2.4 Collaborate on "output" side

This is the converse of the previous item: we must also communicate the central importance of the outputs — these requirements must be as practically useful and rich as possible, but can only be so much as is allowed by the data.

As shown above, the types of situations we must detect and the level of certainty we must have before reporting them are crucial questions, and so it is crucial that we answer them collaboratively along with those who are specifying system requirements. As also discussed above, it is necessary to clarify each requirement to make it as specific as possible, in order to focus effort towards the proper targets; and this will also help us to identify targets which, when fully specified, are not as feasible as they may seem in the abstract.

8.2.5 Be creative and economical

In order to get the most value possible from the data that we will have, it will be crucial that we consider innovative and perhaps less-than-obvious interpretations of these sensor data. One important example of this would be using proxies for information of interest: bathroom motion level throughout the day as a proxy for fluid intake, or television usage to (inversely) estimate physical activity level. Brainstorming for such ideas should be ongoing, and as inclusive of all team members as possible, since novel ideas could come from anyone, not just the more technically oriented.

8.3 Recommendations for further work

Clearly, this document represents only an early attempt at framing the problem of this work package, and at filling it in with a few broad strokes. As the exact details of the design and specification become clearer, a more detailed review of available literature should be conducted, focusing more concretely on the exact tasks at hand. Also, obtaining data arising from early prototypes (or comparable systems elsewhere) will help both to set realistic goals, which will also help to define the scope of this work package and to focus a more detailed examination.

Works Cited

Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from Text: Machine Learning for Text-based Emotion Prediction. (pp. 579-586). Stroudsburg, PA, USA: Association for Computational Linguistics.

Ayadi, M. E., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases . *Pattern Recognition , 44*(3), 572-587.

Chen, C.-Y., Huang, Y.-K., & Cook, P. (2005). Visual/Acoustic Emotion Recognition., (pp. 1468-1471).

Denissen, J. J., Butalid, L., Penke, L., & Van, M. A. (2008). The effects of weather on daily mood: a multilevel approach. *Emotion, 8*(5), 662.

Dibben, N. (2004). The Role of Peripheral Feedback in Emotional Experience With Music. *Music Perception: An Interdisciplinary Journal, 22*(1), pp. 79-115.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology, 17*(2), 124.

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: a survey . *Pattern Recognition , 36*(1), 259-275.

Frantzidis, C., Lithari, C., Vivas, A., Papadelis, C., Pappas, C., & Bamidis, P. (2008). Towards emotion aware computing: A study of arousal modulation with multichannel event-related potentials, delta oscillatory activity and skin conductivity responses., (pp. 1-6).

Glowinski, D., Camurri, A., Volpe, G., Dael, N., & Scherer, K. (2008). Technique for automatic emotion recognition by body gesture analysis., (pp. 1-6).

Gunes, H., & Piccardi, M. (2005). Fusing face and body gesture for machine recognition of emotions., (pp. 306-311).

Hassan, A., Damper, R., & Niranjan, M. (2013). On Acoustic Emotion Recognition: Compensating for Covariate Shift. *Audio, Speech, and Language Processing, IEEE Transactions on, 21*(7), 1458-1468.

Herrmann, M. J., Huter, T., Plichta, M. M., Ehlis, A.-C., Alpers, G. W., Mühlberger, A., et al. (2008). Enhancement of activity of the primary visual cortex during processing of emotional stimuli as measured with event-related functional near-infrared spectroscopy and event-related potentials. *Human Brain Mapping, 29*(1), 28-35.

Hurst, M., Jackson, T., & Glencross, M. (2012). Emotion recognition - Theory or practicality., (pp. 1-6).

Jang, E. H., Park, B. J., Kim, S. H., Eum, Y., & Sohn, J. (2011). Identification of the optimal emotion recognition algorithm using physiological signals., (pp. 1-6).

Kato, T., Kawanaka, H., Bhuiyan, M., & Oguri, K. (2011). Classification of positive and negative emotion evoked by traffic jam based on electrocardiogram (ECG) and Pulse wave., (pp. 1217-1222).

Keller, M. C., Fredrickson, B. L., Ybarra, O., C{\^o}t{\'e}, S., Johnson, K., Mikels, J., et al. (2005). A warm heart and a clear head the contingent effects of weather on mood and cognition. *Psychological Science, 16*(9), 724-731.

Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30*(12), 2067-2083.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., et al. (2010). Music emotion recognition: A state of the art review. *Citeseer*, (pp. 255-266).

Kong, X., & Yang, Y. (2009). Measuring emotions in interactive contexts., (pp. 526-530).

Kostoulas, T., Ganchev, T., Mporas, I., & Fakotakis, N. (2007). Detection of Negative Emotional States in Real-World Scenario., *2*, pp. 502-509.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings. *International affective picture system (IAPS): Technical manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Lee, H., Choi, Y. S., Lee, S., & Park, I. (2012). Towards unobtrusive emotion recognition for affective social communication., (pp. 260-264).

Li, H., Pang, N., Guo, S., & Wang, H. (2007). Research on textual emotion recognition incorporating personality factor., (pp. 2222-2227).

Li, H., Pang, N., Guo, S., & Wang, H. (2007). Research on textual emotion recognition incorporating personality factor., (pp. 2222-2227).

Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., et al. (2010). EEG-Based Emotion Recognition in Music Listening. *Biomedical Engineering, IEEE Transactions on, 57*(7), 1798-1806.

Liu, Y., Sourina, O., & Nguyen, M. K. (2010). Real-Time EEG-Based Human Emotion Recognition and Visualization., (pp. 262-269).

Lu, L., Liu, D., & Zhang, H.-J. (2006). Automatic mood detection and tracking of music audio signals. *Audio, Speech, and Language Processing, IEEE Transactions on, 14*(1), 5-18.

Lundqvist, L.-O., Carlsson, F., Hilmersson, P., & Juslin, P. N. (2009). Emotional responses to music: experience, expression, and physiology. *Psychology of Music, 37*(1), 61-90.

Maat, L., & Pantic, M. (2007). Gaze-x: Adaptive, affective, multimodal interface for single-user office scenarios. In *Artifical Intelligence for Human Computing* (pp. 251-271). Springer.

Maki, Y., Sano, G., Kobashi, Y., Nakamura, T., Kanoh, M., & Yamada, K. (2012). Estimating subjective assessments using a simple biosignal sensor., (pp. 1-6).

Masum, M. A., Helmut, P., & Ishizuka, M. (2006). A Cognitively Based Approach to Affect Sensing from Text. (pp. 303-305). New York, NY, USA: ACM.

Mowafey, S., & Gardner, S. (2012). A novel adaptive approach for home care ambient intelligent environments with an emotion-aware system., (pp. 771-777).

Narayanan, S., & Georgiou, P. (2013). Behavioral Signal Processing: Deriving Human Behavioral Informatics From Speech and Language. *Proceedings of the IEEE, 101*(5), 1203-1233.

Noury, N., Hadidi, T., Laila, M., Fleury, A., Villemazet, C., Rialle, V., et al. (2008). Level of activity, night and day alternation, and well being measured in a smart hospital suite., (pp. 3328-3331).

Patrick, D. L., Bush, J. W., & Chen, M. M. (1973). Methods for measuring levels of well-being for a health status index. *Health services research, 8*(3), 228.

Peter, C., Waterworth, J., Waterworth, E., & Voskamp, J. (2007). Sensing Mood to Counteract Dementia.

Phan, K., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in \{PET\} and fMRI . *NeuroImage , 16*(2), 331-348.

Quazi, M. T., Mukhopadhyay, S., Suryadevara, N., & Huang, Y. M. (2012). Towards the smart sensors based human emotion recognition., (pp. 2365-2370).

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music, 35*(1), 88-109.

Sloan, D. M. (2004). Emotion regulation in action: emotional reactivity in experiential avoidance . *Behaviour Research and Therapy , 42*(11), 1257-1270.

Soleymani, M., Pantic, M., & Pun, T. (2012). Multimodal Emotion Recognition in Response to Videos. *Affective Computing, IEEE Transactions on, 3*(2), 211-223.

Tawari, A., & Trivedi, M. (2010). Speech based emotion classification framework for driver assistance system., (pp. 174-178).

Tokuno, S., Tsumatori, G., Shono, S., Takei, E., Suzuki, G., Yamamoto, T., et al. (2011). Usage of emotion recognition in military health care., (pp. 1-5).

Trabelsi, A., & Frasson, C. (2010). The Emotional Machine: A Machine Learning Approach to Online Prediction of User's Emotion and Intensity., (pp. 613-617).

Tsai, T.-C., Chen, J.-J., & Lo, W.-C. (2009). Design and Implementation of Mobile Personal Emotion Monitoring System., (pp. 430-435).

Vella, K., & Johnson, D. (2012). Flourishing and video games. (pp. 19:1--19:3). New York, NY, USA: ACM.

Vogt, T., Andr{\'e}, E., & Bee, N. (2008). EmoVoice—A framework for online recognition of emotions from voice. In *Perception in Multimodal Dialogue Systems* (pp. 188-199). Springer.

Vuoskoski, J. K., & Eerola, T. (2011). The role of mood and personality in the perception of emotions represented by music . *Cortex , 47*(9), 1099-1106.

Wagner, J., Andre, E., & Jung, F. (2009). Smart sensor integration: A framework for multimodal emotion recognition in real-time., (pp. 1-8).

Wu, Q. (2010). Real-Time Emotion Assessment Method Based on Physiological Signals., *2*, pp. 225-228.

Yang, Y.-H., & Liu, J.-Y. (2013). Quantitative Study of Music Listening Behavior in a Social and Affective Context. *Multimedia, IEEE Transactions on, 15*(6), 1304-1315.

Yu, Y.-C., You, S., & Tsai, D.-R. (2012). Magic mirror table with social-emotion awareness for the smart home., (pp. 185-186).

Zhang, Y., Tang, J., Sun, J., Chen, Y., & Rao, J. (2010). MoodCast: Emotion Prediction via Dynamic Continuous Factor Graph Model., (pp. 1193-1198).

Zhang, Z., Singh, V., Slowe, T., Tulyakov, S., & Govindaraju, V. (2007). Real-time Automatic Deceit Detection from Involuntary Facial Expressions., (pp. 1-6).

Zhao, Y., Wang, X., & Petriu, E. (2011). Facial expression anlysis using eye gaze information., (pp. 1-4).

Zheng, Q., & Qin, J. (2011). Evaluating the emotion based on ontology., (pp. 32-36).