On Applying Ambient Intelligence to Assist People with Profound Intellectual and Multiple Disabilities

Michal Kosiedowski[1], Arkadiusz Radziuk[1], Piotr Szymaniak[1], Wojciech Kapsa[1], Tomasz Rajtar[1], Maciej Stroinski[1], Carmen Campomanes-Alvarez[2], B. Rosario Campomanes-Alvarez[2], Mitja Lustrek[3], Matej Cigale[3], Erik Dovgan[3], and Gasper Slapnicar[3]

[1] Poznan Supercomputing and Networking Center, ul. Jana Pawla II 10, 61-139 Poznan, Poland
[2] CTIC Technological Centre, C/ Ada Byron, 39 Edificio Centros Tecnológicos Cabueñes s/n, Gijon, Spain
[3] Jozef Stefan Institute, Jamova cesta 39,Ljubljana, Slovenia
{kat,hamerhed,hollow,kapsa,stroins,ritter}@man.poznan.pl,
{carmen.campomanes,charo.campomanes}@fundacionctic.org,
{mitja.lustrek,matej.cigale,erik.dovgan,gasper.slapnicar}@ijs.si

Abstract. Advances in ambient intelligence technologies achieved over recent years allow building ICT systems capable of supporting people in executing previously difficult tasks. This includes providing new opportunities to people with special needs, such as people with disabilities. In this paper we discuss our approach at designing and developing a smart platform that can assist people with profound intellectual and multiple disabilities (PIMD) in achieving a portion of independence. We apply this approach in the INSENSION project executed within the Horizon 2020 research and innovation programme of the European Commission. In this paper, we describe the characteristics of the disability in question, focusing on the main challenge, which is the inability of individuals with PIMD to use symbols in their interaction. Due to the fact that, to our best knowledge, the topic of constructing a system that could assist this interaction has not been undertaken so far, we were required to thoroughly analyze how individuals with PIMD interact using non-symbolic behaviors. We then defined requirements for the platform capable of supporting their interaction with other people and possibly living environment, designed the architecture and built the first version with the use of AI methods. The evaluation of this version confirmed the soundness of our approach, which enables us to continue our work towards successful implementation of the planned ambient intelligence system and its validation in real-life scenarios.

Keywords: smart assistive technologies, non-symbolic interaction, accessible technologies

1 Introduction

The advancement of Information and Communication Technologies that can be observed in the recent years has allowed for the development of innovative

technological solutions capable of fulfilling the ever increasing percentage of the needs of vulnerable people. This relates to the needs arising as people age and to the needs of people living with disabilities. The former of these topics has been the center of attention of a significantly big community of researchers and technical experts. This has resulted in a number of innovative solutions supporting older adults, often developed with the financial support of dedicated research programs such as the Active and Assisted Living Program. These solutions, often referred to as ambient assisted living (AAL) solutions, have been attempting to use various forms of ambient intelligence. For example the Fearless project developed a system that uses vision based sensors to detect risks in daily life of the elderly [1]. The PersonAAL project proposed a platform detecting behavior changes in older adults in order to persuade them to undertake healthier behaviors [2]. In our own previous work we developed a system that unobtrusively monitors physical activity, mental stress and health of the workplace environment of older adults to recommend relevant actions aimed at decreasing the risk of illness [3].

Ambient intelligence technologies are also applied as support for people with disabilities. Some of examples include smart sensing solutions for visually impaired [4] or smart sign-language interpreters for the deaf [5]. Thanks to these innovative applications of ambient intelligence technologies an ever increasing number of people with disabilities receive tools supporting their self-reliance. Nevertheless, there still exist populations with disabilities who have not had a chance to use the potential of today's technologies. One such group are people with profound intellectual and multiple disabilities (PIMD – also referred to as PMLD – profound and multiple learning disabilities).

Individuals affected by profound and multiple disabilities are immobile or have severely restricted mobility and are subject to profound and multiple sensory impairment in combination with profound intellectual impairment [6]. Their capacity to perceive and act upon the interactive situation about them is significantly and severely diminished. Individuals affected by PIMD remain at a very early stage of development for a prolonged period of time, if not a lifetime. They often communicate on a pre-symbolic level and use unconventional behavioral signals (e.g. specific body movements or vocalizations) to express their needs. Consequently, they are usually unable to use existing technological devices, even those that use not using advanced ICT, as these tools require understanding of symbols. Moreover, the number of those interaction partners who are able of accurately perceiving and interpreting the specific and highly individual behavior signals is very limited in most cases. The exact understanding of the needs of people with PIMD is not often possible even for very familiar persons. This significantly restricts the participation of this group in all areas of life.

Supporting people with PIMD is extremely challenging today as it requires constant support of a member of the very small – constituted by only a few people, group of professional and informal caregivers who are able to understand the communication signals expressed by a given individual with this type of disability. Nevertheless, in our opinion, the advancement of a number of technologies allows to construct a smart ICT solution capable of recognizing those behavioral communication signals, collecting information about context of these behaviors and

analyzing this information in order to find meaningful patterns and to interpret them as specific intents of the given individual with PIMD. This can significantly change the quality of life of people with PIMD allowing to improve assistance that can be provided to them and in consequence to empower them to a certain level of their individual potential. We believe this is possible because computer vision techniques, sound analysis techniques, Internet of Things technologies and machine learning techniques can be appropriately applied to construct a comprehensive system capable of assisting people with PIMD in interacting with other people and changing their living environment according to their current needs. We perform such application of the mentioned technologies within the INSENSION project in which we design and develop a personalized intelligent platform enabling interaction with digital services to individuals with PIMD. In this paper we present the results of the research conducted during the first year of the project. This work allowed us to analyze the requirements connected with building such a platform, to confirm usefulness of specific technologies for the delivery of specialized system components and finally to design the architecture of this platform. We discuss these in Sections 2, 3 and 4 respectively.

2 Requirements for ICT-Supported Communication of People with PIMD

We have already mentioned in Section 1 that the target users of the INSENSION platform communicate on a pre-symbolic level and that their interaction schemes are highly individual and known to only a few persons. This interaction is based on behaviors such as gestures, facial expressions or vocalizations of an individual with PIMD, or combinations of these. They can express demand, protest or comment towards a certain situation [7]. This situation is therefore the context of the specific meaningful behavior of an individual with PIMD. When attempting to interpret the need of the given individual with PIMD direct support persons link important elements of the situation happening around that individual with the particular behavior. This allows them to make a decision on what kind of support is needed at a given time. This may relate to prolonging an activity which is demanded by the individual with PIMD such as for example entertaining relaxation on a swing or playing with a favorite toy, or reacting adequately in the case of a protest against the crowdy room or cold temperature.

That means that the core functions of the INSENSION platform should be recognizing these behaviors as one of the three generalized statements mentioned above and monitoring key elements of the contextual situation happening around the supported individual with PIMD. We discuss these two elements in the current section, together with the results of our wider analysis of requirements towards the system that is capable of utilizing these core functions for the benefit of the given individual with PIMD (also referred to as ‘primary user’ throughout the remainder of this paper, as opposed to ‘secondary user’ which refers to persons providing direct support to a given individual with PIMD).

2.1 Non-Symbolic Behaviors of People with PIMD

The first step in designing the INSENSION platform was to understand more exactly what kind of gestures, facial expressions or vocalizations are used by people with PIMD.

Behaviors of a given individual with PIMD can be assessed using several specialized tools in order to gather a comprehensive view of the communication model of that individual. Such tools have been used by the INSENSION project to assess a group of 6 individuals with PIMD with characteristics as follows:

- male, 8 years old, with Struge-Wender syndrome;
- male, 9 years old, with cerebral palsy and epilepsy;
- male, 11 years old, with cerebral palsy, hypotonic form, quadriplegia and erectile pattern in the lower limbs.
- male, 14 years old, with post-inflammatory hydrocephalus and implanted placental system after purulent meningitis with etiology of e-coli in the non-infant period;
- female, 41 years old, with significant intellectual disability and able to walk;
- female, 30 years old, with cerebral palsy, hydrocephalus and epilepsy.

The assessment was performed by a team of special pedagogy experts supervised by Prof. Peter Zentel of the Heidelberg University of Education. Zentel and his colleagues collected information on the above-listed individuals with PIMD using an assessment system that included tools allowing to assess: the general competencies – these are oriented towards tools described in [8] and [9], communication skills using the Communication Matrix [10]; expressions of mood with the use of The Mood, Interest and Pleasure Questionnaire [11]; behaviors related to pain with the use of Non-communicating Children's Pain Checklist [12] and Non-communicating Adult Pain Scale [13]; and behaviors indicating pleasure and displeasure or distress with the use of Disability Distress Assessment Tool [14]. Next, the results of the performed assessment were combined with the video and audio recordings of the actual behaviors of the assessed individuals with PIMD. This was done with the use of the ELAN video annotation tool [15] and allowed to build non-symbolic behavior models for each of the assessed individuals. ELAN is applied in humanities and social sciences research. It provides a three step procedure in which the researcher who annotates the video material first defines tiers and their types, then selects time intervals and finally annotates the video. The software allows to use both: time and event sampling.

For the material that was collected in the INSENSION project, our colleagues defined three main areas describing a specific situation observed in the recording. These were *Behaviors* indicating the actual behavior of the recorded individual, *Communication and Inner States* indicating the meaning of the recorded behavior and *Context* describing factors influencing the recorded individual to express through the current behavior. These areas were then divided into categories (such as for example *Facial Expressions* in the area *Behaviors*), categories into subcategories (such as for example *Apperance of Eyes*), and finally into nearly 100 tiers (such as for example

Eyebrow Movement or *Widened Eyes*) allowing to precisely annotate the recordings. Material annotated in such a way was the basis of experiments leading to confirmation of the possibility of recognizing the meaningful behaviors using relevant smart technologies as described in Section 3.

2.2 Context of Behaviors of People with PIMD

Behaviors of people with PIMD are usually a response to the situation that happens around them. This response relates to demanding that a certain situation continues, protesting against that situation or commenting it. Therefore, in order to interpret the actual meaning of these demands, protests and comments, we need to understand the situation that is concerned by them, i.e. the context of the behaviors of people with PIMD. To this end we used the results of the assessment of the six individuals with PIMD as described in Section 2.1 to derive the requirements for gathering information on the context of non-symbolic behaviors of people with PIMD. First, we created a list of all external circumstances influencing behaviors of the assessed individuals found in the assessment surveys. Next, we attempted to analyze them in relation to the technologies that could be used to monitor these circumstances. This allowed us to define requirements as to what the INSENSION platform should be monitoring in order to be able to correctly interpret specific behaviors of its target end users.

The performed analysis showed that the following technologies could be used in order to collect information on the context:

- video analysis, including identification of other people, monitoring position of the individual with PIMD, identification of specific objects such as toys, recognition of specific activities such as eating, and monitoring the temperature or other characteristics of objects with which the individual with PIMD interacts;
- sound analysis, related to recognizing various types of sounds, including music, singing, other people's voices and sounds of specific objects;
- monitoring ambient parameters, related to measuring such parameters as temperature or illuminance, and recognizing sudden changes in the environment, e.g. sudden loud noises.

2.3 Functional and Non-Functional Requirements for the Smart System Assisting Non-Symbolic Interaction

In sections 2.1 and 2.2 we described the approach we used to define requirements concerning the use of specific technologies to recognize behaviors of people with PIMD and their context. The components built on top of these technologies constitute the core functionality of the developed platform. However, it was also important for us to define how this core functionality should be integrated into a working system capable of everyday support of people with PIMD, and indirectly their caregivers. To this end we performed the assessment of the potential functional and non-functional requirements for the INSENSION platform using the following methodology.

First, we gathered an internal project group of experts on special pedagogy, business related to distribution of assistive technologies for people with disabilities and ICT. The group was constructed in such a way that the number of technological

(ICT) and domain (special pedagogy and business) were equal. This group of experts participated in a brainstorming session aimed at listing all potential characteristics of the developed system. These were defined from the point of view of all potential users, including primary users (people with PIMD), secondary users (informal and professional caregivers) and tertiary users (other people). As a result of this brainstorming session we created a list of nearly 80 potential features of the developed system.

Next, the members of the expert group assessed the defined system features according to their expert knowledge, rating each feature with one of the three grades: "OK", "discard", "decide later". The experts could also suggest that a given feature is out of the scope of the current ICT system. The features which have been assessed by any group member as 'out of the scope' or marked as 'discard', were abandoned.

The final step was to categorize the features according to the MoSCoW methodology [16]. This methodology allows to categorize the system features in order of their importance/priority. As a result we defined nine 'must' features that are critical for the success of the solution and its usefulness for the users, eight 'should' features that are equally as important, but could be implemented at a later stage, five 'could' features that are desirable but not necessary for the user satisfaction, and four 'would' features that are least critical, yet provide some added value for the end users. The 'must' features included, among others, making the platform to inform other people about the needs of the primary user – the individual with PIMD, interpreting the reactions of the primary user to the action undertaken by the platform as their feedback on the correctness of the behavior recognition and making sure that the direct support persons are not overwhelmed with the information provided by the platform to the them. The final list of prioritized system features was the basis for drawing the architecture of the INSENSION platform and defining its core mechanisms.

3 Application of Advanced Technologies for Supporting Non-Symbolic Interaction

Following the definition of the requirements concerning construction of the smart ICT system capable of assisting non-symbolic interaction of people with PIMD, we performed a number of experiments with the technologies foreseen to enable recognition of the non-symbolic behaviors of people with PIMD. The goal of these experiments was to confirm that a given function of the INSENSION platform can be implemented successfully, to provide estimation of further work on each of the envisaged recognition components. In this section we discuss each of the considered recognition technologies and the resulting components. It is important to understand that all the reported experiments were performed using the excerpts from the video and audio recordings collected and annotated as described in section 2.1.

3.1 Facial Expression Recognition

Facial expressions are the facial changes in response to person's internal emotional states, intentions or social communications. From a computer vision point of view,

facial expression analysis refers to computer systems that attempt to automatically analyze and recognize facial feature changes from images. This analysis includes both measurement of facial motion and recognition of expressions. The general approach to automatic facial expression analysis (AFEA) consists of three steps: face acquisition, facial data extraction and representation, and facial expression recognition.

Studies of automatic facial expression recognition have made a significant progress in the last two decades due to the advances in machine learning and computer vision techniques. The current research can be classified in two types: the recognition of the appearance of facial actions and the recognition of the emotions conveyed by the actions. Following our previous experience we proposed to use the former. This kind of system usually relies on the facial action coding system (FACS) [17]. FACS consists of 44 facial action units (AU), which are codes that describe certain facial configurations.

We used the OpenPose library [18] for creating the component for the recognition of facial expressions. Four expressions – or AUs – were selected for performing the experiment. For each expression, a sample vector was built by means of computing the L2 Euclidean distance measure [19] from a reference key point (the tip of the nose) to the rest of face key points. Therefore, a vector of distances determines a particular facial expression. A different number of vectors of distances described each expression, because there could exist different head poses and orientations for the same expression, so the distances between key points vary for the same facial expression. Fig. 1 illustrates the location of the facial key points for which distances determining a facial expression are calculated.

After building a model with a machine-learning algorithm, a validation of the model was performed with the previously selected test dataset. The obtained error test rate was 0.0824. Therefore, using this preliminary training and test dataset, an accuracy equal to 0.9176 was achieved for the facial recognition model for a set of four facial expressions.

Fig. 1. Facial key points used for calculating distances determining facial expressions.

3.2 Gesture Recognition

A gesture is the use of motions of the limbs or body as a means of expression, to communicate an intention or feeling. Because gestures vary highly from one person to another, which is specifically true for people with PIMD, it is essential to capture the essence of the gesture – its invariant properties – and use this to represent the gesture. Besides the choice of representation itself, a significant issue in building gesture recognition systems is how to create and update the database of known gestures [20]. In general, a system needs to be trained through some kind of learning, there is often a tradeoff between accuracy and generality.

Static gesture or pose recognition can be accomplished using template matching, geometric feature classification, neural networks (NNs), or other standard pattern recognition techniques to classify pose. Dynamic gesture recognition, however, requires consideration of temporal events. This is typically accomplished by using techniques such as time-compressing templates, dynamic time warping, Hidden Markov Models (HMMs) and Bayesian networks.

Following the approach of the facial expression recognition, we used OpenPose to create the component for the gesture one. OpenPose's pose key points have been used to classify different gestures like hand on head or raising leg. In the frontal view of the human body OpenPose detects 18 key points as presented in Fig. 2. The gesture recognition component builds a sample vector by computing the distances between the reference key point to all other pose key points. We used the neck key point as the reference point. The distance metric allows to uniquely classify different postures.

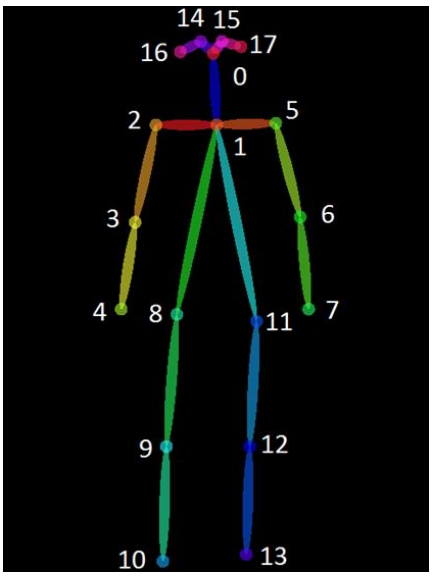

Fig. 2. OpenPose's key points of the human body used for calculating distances determining a given gesture.

Next, we selected four gestures for the experiment. These were: "hand_on_hand", "foot_on_foot", "raising_right_arm" and "raising_left_arm". A sample vector codified each gesture by calculating the L2 Euclidean distance measure [19] from a reference key point to the rest of points that represent the body. As with facial expressions, a different number of vectors of distances described each gesture, because there could exist different head poses and orientations for the same expression, so the distances between key points vary for the same gesture.

The gesture recognition was evaluated on excerpts from video files described in section 2.1. One part of the data was selected for the training phase and the other part was used for the test stage in a proportion of 67:33, respectively. A Time Distributed Feed Forward (dense) NN was used to train the recognition system using the set of training gestures.

Finally, the validation of the constructed recognition component has been performed. The obtained accuracy was equal to 90% for the gesture recognition model for a set of four body gestures.

3.3 Vocalization Recognition

Vocalization recognition aims at detecting instances of non-linguistic sounds produced vocally by an individual under surveillance. Such sounds include laughing, wailing, heavy breathing etc. Similarly to other media of expression (i.e. facial, gestural), not all of these sounds have to correspond to interpretable messages communicated by a particular individual with PIMD.

The classical approach to the vocalization recognition utilizes input signal parameterization of input signal and statistical analysis of extracted parameters in order to find out what predefined category it represents. The general procedure in these studies is as follows: (1) a set of characteristic features is extracted from short packages of sampled signal using statistical measures (mean, standard deviation or more complex, like Mel Spectrum or MFCC); (2) extracted features are classified into one of predefined states [21,22]. The rapid development of artificial neural networks (ANN) in recent years allowed to utilize ANN for detection and classification of voice signals. Here ANN are used on first or second step in connection with the classical feature extractor or classifier [23,24,25]. For the first trials of vocalization recognition in the INSENSION project we used the classical approach.

The vocalization recognition is a two-stage process. First, signal parametrization for uniformly spread short-time audio frames is performed with mathematical transformations (e.g. *Mel-frequency cepstral coefficients* (MFCC)). In this process a portion of acquired audio signals is taken, and from each such portion a vector of features is extracted. Our frames were overlapping windows of 25 ms of audio signal starting every 10 ms. Subsequently, these extracted feature vectors are fed into a statistical process-modeling framework based on Hidden Markov Models. The analysis based on statistical learning models allows to classify/categorize (by assigning labels, "tags") particular groups of samples that were characterized by vectors of extracted features at the first stage of the detection process. The labels produced this way contain also information on the calculated accuracy of the recognition.

In our solution, vocalization types are represented by a list of distinct states for which a state transition matrix is defined. Each state corresponds to one stationary segment of audio observations. The stationary signal in a given state is thus represented by its Gaussian Mixture Model (GMM) that describes distributions of parameters extracted from the audio signal during the parameterization phase. It is also possible to extend the models by explicit state duration distributions, which formally makes them semi-Markov rather than HMM.

The training procedure consists of two phases. In the first phase unsupervised audio frame clustering is performed using a GMM-based method. The second phase deals with re-estimation of the model parameters using Expectation-Maximization (EM) method whose objective is to increase the degree to which the model matches the training data.

A separate model is constructed for each vocalization type. Input audio stream is processed using each of these trained models. Therefore, one observation can be classified as belonging to different vocalization types with various degrees of confidence.

For labeling the samples we implemented a Token-Passing algorithm. A token represents a hypothesis that the given sample should be tagged with a certain label, because it is partially matched to some range of input observations, and 'currently' occupies one of the states of the trained model. Each token remembers its supposed start time, as well as its accumulated cost from this point; acoustics (GMM emission probability), transition/duration distributions and constant insertion penalties all contribute to the total cost. With each observation, i.e. short-time audio frame, tokens are passed from state to state according to the transition matrix. Tokens may be cloned, which happens when a state has more than one possible successor. Whenever two tokens reach the same state of the model at the same time, the more expensive one (according to the cost) is discarded. To speed up computation, heuristic pruning is implemented, that puts out least promising candidates within a set of tokens belonging to the same model. Whenever a token reaches a final state of the model, it undergoes a final scoring; a weighted sum of acoustic and transition/duration average per frame costs is computed. The final scoring must be lower than a preconfigured threshold to treat the token as a candidate for output decision. Finally, only the most likely candidates are preserved from those overlapping ones.

Using the approach described above, we have conducted an experiment to assess accuracy of our system using recordings of behaviors of two participants with PIMD. For them we have identified 7 distinct vocalization categories. Due to the limited number of vocalization instances (the highest number of occurrences of the same vocalization category was 9), we have experimented with the test-on-train setup (ToT). Each model was build based on all its known examples, and with the pseudo-cross-validated (pXV) setup, in which for every category three separate models were built. For each model about ~1/3 of vocalization instances were excluded from the training process. All models were tested on all vocalization instances identified for the given individual.

For these experiments, the maximal achieved F1 score was 0.814 for the pXV setup, for the vocalization category with 9 instances, while the combined F1 score for all tested vocalization categories in this setup was 0.593. The difference between the result for the vocalization category with the highest number of instances and result for

all categories indicates that building a personalized vocalization recognition component for an individual with PIMD requires set of data larger than collected for the experiment. Nevertheless, the obtained results are promising as we were able to implement a vocalization recognition component capable of recognizing most of the identified vocalization instances for the most numerous category. Providing training data sets containing more samples (instances) should increase the accuracy in the future.

3.4 Video-based Recognition of Physiological State

In addition to the fact that non-symbolic behaviors constitute the core of the interaction models of people with PIMD we have also assumed that some additional information concerning the message communicated by an individual with PIMD can be derived from their physiological response. This assumption is based on literature reports which suggest that understanding the meaning of the observed non-symbolic behaviors in people with PIMD may be strengthened by supportive use of physiological parameters monitoring [26]. Studies have shown that "heart rate and skin temperature can give information about the emotions of persons with severe and profound ID", similar to people without disability [27]. For example, it has been observed that participants to the studies showed "frequent consistent physiological reactions" to stimuli [28], and a "shallow, fast breathing pattern, used less thoracic breathing, had a higher skin conductance and had less RSA when experiencing positive emotions then when experiencing negative emotions" [29]. Due to this we aim to attempt using physiological parameters such as heart rate (HR), heart rate variability (HRV), breathing rate (BR), etc. to determine the physiological state to support the detection of intents of people with PIMD.

Determining the physiological state from physiological data requires that the following three steps are completed: (1) the reconstruction of physiological signals from sensor data, (2) the calculation of physiological parameters from the physiological signals if this is not performed by the sensor device such as in the case of cameras and microphones, and finally (3) the determination of the physiological state from the physiological parameters. Due to the fact that we try to make the INSENSION platform a contactless system for the primary users, we look into the video-based methods for monitoring of the physiological parameters, particularly remote photoplethysmography (PPG). PPG values reflect the volume of blood in tissues, which increases when the heart pushes blood towards the periphery of the body, and decreases when the blood returns to the heart. PPG is typically measured with wristbands and fingertip devices, so it is called remote PPG or rPPG when retrieved from video.

Two main approaches are used for PPG reconstruction using RGB cameras. The first – color-based – approach uses the same physiological phenomena as wristbands do, i.e., it analyzes the changes in color of the skin that corresponds to blood volume changes, in order to reconstruct the PPG signal. As a light source, this approach uses the ambient light, which is less predictable than the light source of the wristband, which is in contact with the skin. Consequently, this approach is very sensitive to different environmental conditions. Therefore, it is no surprise that an independent evaluation on a publicly available dataset showed that several methods reported in

literature are not precise enough to be used in real-world scenarios [30]. More precisely, this evaluation included three state-of-the-art methods for retrieving rPPG and the results show that there is low correlation between the reconstructed and true PPG. The second – motion-based – approach analyzes the small head movements that are induced by the pumping of blood into the head [31]. However, it should be noted that these small movements are very subtle and might not be recognized with a low quality camera.

All of the reported camera-based methods use the face of the subject as the "field of interest" in their research, which also needs to be detected for facial expression recognition described in Section 3.1.

Since retrieving the rPPG signal from video recording appears to be quite difficult, we tried five different methods for the color-based approach inspired by the related work. We also tried one motion-based method [31]. In this method, we track the vertical motion of facial pixels with the Lucas-Kanade flow-tracking algorithm [32]. Once we select the most PPG-like signal as a result of each method, we further refine it using a deep NN that takes a window of noisy rPPG signal as input and outputs a correct PPG signal trained on a reference signal from a fingertip device. Experimental evaluation showed that the NN-based method outperformed all six methods inspired by related work. We present example 10 seconds of the rPPG signal reconstructed using this method in Fig. 3. After obtaining the rPPG signal, physiological parameters are estimated by detecting the peaks in the signal, which correspond to heartbeats. This way, heart rate and heart-rate variability can be derived. In addition, features will be computed from the morphology of the segmented PPG cycles, with which we will attempt to estimate respiratory rate and blood pressure.

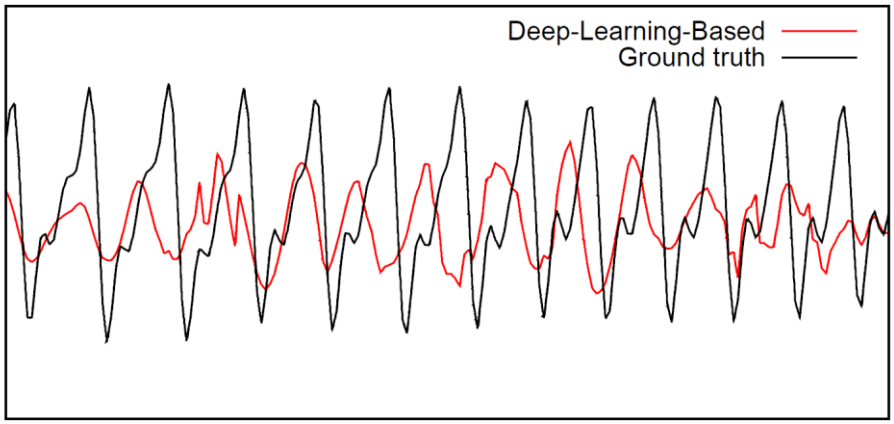

Fig. 3. Example 10 seconds of the rPPG signal reconstructed using our experimental deep neural network

Due to the fact that studies show that it might be extremely difficult to create a component for video-based monitoring of physiological parameters working in real-

world settings, we assumed that the INSENSION platform should enable a fallback solution based on a state-of-the-art wristband device. A simple experiment of testing whether people with PIMD find wristbands obtrusive that we performed with the participation of our test group has shown that such devices are in principle accepted by this population group.

3.5 Behavior Pattern Recognition

The interpretation of behaviors of people with PIMD requires that we are able to map the recognized behavior features (specific gestures, facial expressions and vocalizations) onto their meaning. In other words, we need to define behavior patterns and then recognize them as specific messages. The development of such a component requires that we extract relevant information from the components for recognizing gestures, facial expressions and vocalizations, the data collected with the use of the assessment system as described in Section 2.1, and real time feedback from secondary users and other ICT components such as assistive applications discussed in Section 4.4.

In order to learn the basic state of an individual with PIMD we first try to determine the behavior state (‘pleasure’, ‘displeasure’ or ‘neutral’) and communication attempt (‘comment’, ‘demand’, ‘protest’) from gestures, facial expressions and vocalizations. Since the capabilities of the users are not uniform, the models for extraction are user specific. To learn the gestures, facial expressions and vocalizations that are associated with the behavior state, we look at all the gestures, facial expressions and vocalizations that are associated with one state and are not part of the other behavior states. For example, if we want to extrapolate the behaviors associated with pleasure, we look at all the gestures, facial expressions and vocalizations that are associated with ‘pleasure’, but not ‘displeasure’ or ‘neutral’.

When we are presented with an unclassified behavioral state we look at the gestures, facial expressions or vocalizations that are associated with it. For example, if we find behaviors that are associated with ‘pleasure’, we decide that the unclassified behavioral state is ‘pleasure’. While this is a rather naive approach, it provides the best results on the current dataset. The communication attempt is classified in the same manner with three classification classes (‘comment’, ‘demand’, ‘protest’).

To validate our approach, we created a Prolog-based software component. It learned the communication attempt or behavioral state on a subset of all the annotations and validated the results by trying to classify the remaining annotations. We used annotations of the recordings collected from the test group as done by the special pedagogy experts (see Section 2.1). Due to the sample size we removed one example of annotated data for each class (i.e., we removed one example of ‘pleasure’, ‘displeasure’, ‘neutral’) and tried to classify the behavioral state for all possible combinations of removed behavioral state for each user.

The results of the validation of the constructed method show that the ‘neutral’ state is the least accurate and that the accuracy ranges from 45 % to 80 %. ‘Displeasure’ is the most robustly classified, being accurately predicted from 83.3 % to 96 %, while ‘pleasure’ is correctly classified from 60 % to 86 %. Classification accuracy for communication attempts ranges from 70 % to 100 %, with the best results for ‘demand’, where the lowest achieved accuracy is 89 %. When interpreting these

results one must note that they were achieved based on the annotations done by the special pedagogy experts rather than the recognition components described above. That means that they were noise-free as opposed to the foreseen results of the real-time automatic recognition.

4 Insension Platform for Personalized Assistance of Non-Symbolic Interaction of People with PIMD

The components discussed in section 3 provide the key functionality of the intelligent system capable of assisting the non-symbolic interaction of people with PIMD. However, successful utilization of this functionality requires integrating it with other important functionalities such as data acquisition or data storage into a comprehensive environment, and finally enabling the results of the non-symbolic behavior interpretation for practical consumption. The latter is related to integrating those results to instruct a range of digital applications to undertake specific actions on behalf of the given individual with PIMD, typically in assistive scenarios. The primary foreseen scenario is communication of an individual with PIMD with other people, most importantly caregivers. Others may include controlling smart room facilities such as heating or playing favorite music based on automatic recognition of needs expressed by the primary user of the system.

In the current section we discuss the issues related to the design of the platform and its practical utilization in real-life scenarios.

4.1 Platform Architecture

The analysis of requirements reveals that the main control flow of the platform consists of processing input video, audio, ambient and physiological data streams, detecting significant events and generating messages based on the result of this processing. Generated messages become the input data for further processing components that determine user intents. Finally, the user intents may be used by applications and services, typically of assistive nature as we state above.

This kind of processing naturally indicates that the control flow is driven by events. Taking this into account, we proposed to use an Event Driven Architecture (EDA) [33] for the platform. EDA is an architectural pattern which is based on creation, consumption and reaction to events. Occurrence of an event causes generation of the notification which can be used by other system components or third-party services to change their internal state or trigger an action assigned to the event. EDA allows to create loosely coupled and highly distributed systems.

The architecture of the INSENSION platform is presented in Fig. 4. The Sensors Layer represents the hardware components concerned with the collection of raw data from the user and their living environment. These are controlled by the Sensor Data Acquisition Layer which is responsible for the preparation and delivery of the data streams as expected by the recognizers that are the components that we discussed in more detail in Section 3. We describe this layer in more detail in Section 4.2. Recognizers are grouped on the Recognizers Layer. Due to the fact that multiple

sensors of various types may be used to acquire data and at the same time it is crucial that these data are synchronized in time, we plan to use a Time Synchronization Server. Next, the platform features four specialized services. The Interaction Decision Support Service is a service that performs the contextualized behavior pattern recognition (as described in Section 3.5) on the one hand and is foreseen to produce the recommendations of actions to be performed on behalf of the given primary user, here referred to also as user intents, on the other. The Application Access Management and Control service is responsible for the communication with assistive applications that are external to the. We shortly discuss main assumptions concerning this service in Section 4.3. The Platform Management Service is an application with a graphical user interface for secondary and tertiary users to control, configure and interact with the platform and its services. Finally, the Repository is the data storage, capable of storing all kinds of data acquired and produced by the platform. The platform services are integrated through a messaging mechanism, i.e. the Message Broker. We based this component on the RabbitMQ framework.

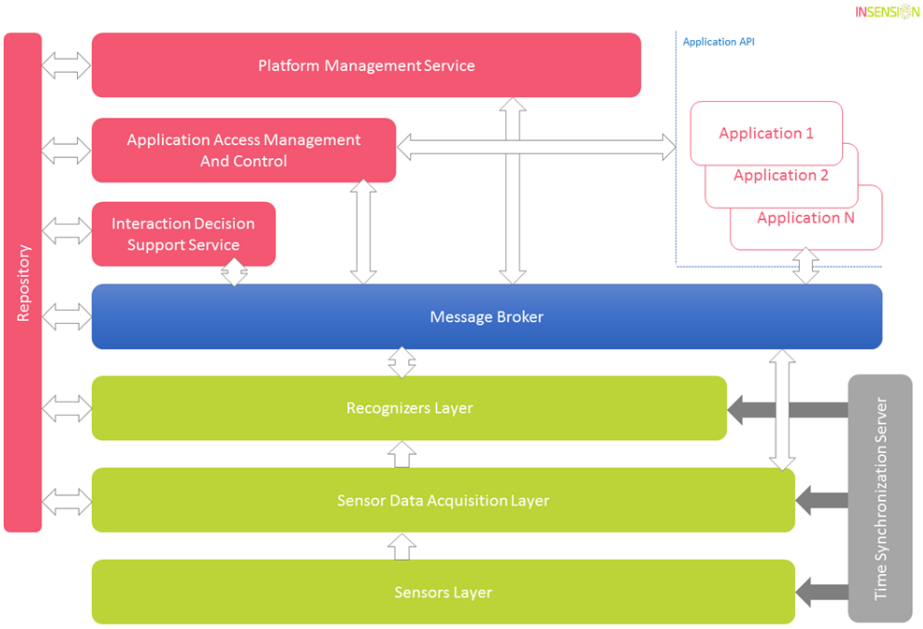

Fig. 4. The architecture of the INSENSION platform for personalized assistance of non-symbolic interaction of people with PIMD

4.2 Data Acquisition

The contextualized recognition of non-symbolic behaviors of people with PIMD requires the acquisition of two types of data: a) multimedia (video and audio) streams collected from the primary users and their environment, and b) data from ambient

sensors installed in the living environment of the primary user, used to monitor specific parameters of the environment, such as temperature or illuminance, capable of generating events related to the measurements of these parameters.

The multimedia streams are intended to serve as data sources for recognizers. The video streams feed facial expression recognizers, gesture recognizers and video-based physiological state recognizer. Audio streams are processed by vocalization recognizers. Multimedia streams are transported in a local network over RTP, especially suited for real-time multimedia transport over IP networks. The transmission is established and torn down by RTSP. Video streams are encoded using the widespread H.264/AVC encoder. This encoder allows transmission of high quality, high resolution video streams in a reasonably small bandwidth portion. The audio streams may be encoded using a variety of audio codecs, however considering our assumption of using hardware such as Odroid devices, we foresee to use AAC LD or Opus.

The data acquisition stack for the measurement of ambient parameters related to the interaction context is based on existing, relatively low-level Internet of Things platforms, which are designed to be a backbone for physical sensors systems. Based on our previous experiences we assumed to use the Node-RED IoT platform [34]. This is an open source framework that allows easy adaptation to specific needs of a give IoT project. Moreover, its software components do not consume lots of resources, making it available for deployment on relatively cheap minicomputers. Node-RED reads data from a microcontroller device that handles physical sensors, and processes them to provide meaningful measurements of specific parameters. This microcontroller device is logically part of the Sensor Data Acquisition Layer. Initially, we included sensors capable of measuring air temperature, air humidity and luminosity level.

As it was already mentioned, one of the important mechanisms is synchronization of time within all the data acquisition and processing components. A simple solution is using the NTP service, as it allows to synchronize clocks within 1 ms in local networks. However we are also considering using 1588-2008 - IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems in case more precise synchronization is needed.

4.3 Integration of assistive applications

To adequately assist the primary users in interacting with other people and their ambient, the INSENSION platform should allow connecting many assistive applications at the same time. In case the Interaction Decision Support Service does not interpret the intent precisely enough, the intent could trigger multiple applications (e.g. the primary user's need – if interpreted only generally – could trigger playing music and turning up the lights). To avoid this, applications have priorities, which define which of them receive intents first. If there are many applications running at the same time and registered to receive intents of the same type but with different priorities, the INSENSION platform sends the primary user intent message to an application with the highest priority. If this application decides not to process this user intent, the message will be passed to the next one. We also propose one default application that can process any detected user intent. It is foreseen that this default

application is an application allowing the primary user to communicate their needs to other people.

5 Conclusions and Future Work

In the current paper we presented the initial design of the INSENSION platform that aims to provide functionality of recognizing meaningful non-symbolic behaviors of people with profound intellectual and multiple disabilities in order to assist these individuals in interacting with others and their ambient. The presented work was conducted during the first year of the INSENSION H2020 project and allowed us to understand whether the initial assumption that it is possible to construct a working system in question is valid. To this end, we elaborated requirements of such a system and collected video and audio material containing non-symbolic behaviors of people with PIMD. This allowed us to define use cases and analyze the collected data in order to identify the types of behaviors in question as well as the range of circumstances that influence them.

In the next step we performed early experiments concerning methods to be used for developing the intelligence of this system. In these experiments we applied known methods of video and audio analysis, as well as pattern recognition, to select the most suitable methods for future work. The components built on top of these methods are required to recognize facial expressions, gestures, vocalizations and psychophysiological state in video and audio streams acquired from the primary user with PIMD, and to collect important information on the context of the recognized behaviors in video, audio and sensor data streams acquired from the ambient of the user. They should be integrated into a comprehensive ambient intelligence system that is capable of assisting non-symbolic interaction of people with PIMD in real-life scenarios.

Upon conducting the work that we describe herewith it can be concluded with a high level of certainty that such a system can be designed and developed. Nevertheless, there are some areas which need special attention in our further work on this platform. First of all, we need to carefully address the process of configuring the system to work for a given individual with PIMD. Due to the fact that the non-symbolic behaviors cannot be generalized and are highly personal, we foresee a training phase when preparing the system to work for that given individual. Further on, as we have expected, the development of components for gesture recognition, facial expression recognition and vocalization recognition requires additional data collected from the end users. Nevertheless, the selected methods to be used for developing those components are promising. We encountered a different situation with the video-based physiological state recognition. The methods described in the literature do not work in real-life scenarios. Therefore, we need to assume a more complex approach to attempt building a working component providing this type of functionality. Due to the fact that we are unable to confirm it is possible to implement such a component by the end of the INSENSION project, we decided to evaluate a fallback solution using a wristband. A simple acceptance test confirmed that the end users are fine with wearing such devices. We foresee using the Empatica E4 device for this purpose, if needed.

The further work that has been planned for the nearest future considers collection of further data from the representative group of the end users, usage of these data to develop final versions of intelligent components, and full development of the whole system that incorporates these components in real-life scenarios. This development is planned to be conducted using an iterative approach with the direct participation of end users. Starting with the mockup of the whole system that is capable of demonstrating the whole concept to the end users, we plan on iteratively verifying the current concept and working components, and on developing new versions of system components using the feedback collected from the current interaction. Once we achieve a fully working prototype of the whole INSENSION system, we aim to perform a real-life validation of its functionality.

6 Acknowledgements

The research presented herewith has been conducted within the INSENSION project which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 780819.

References

1. Berndt, R-D., Takenga, M.C., Kuehn, S., Preik, P., Berndt, S., Brandstoetter, M., Planinc, R., Kampel, M.,: An Assisted Living System for the Elderly – FEARLESS Concept, In: Proceedings of the IADIS Multi Conference on Computer Science and Information Systems eHealth2012, pp. 131-138, Lisbon (2012).
2. Azevedo, C., Chesta, C., Coelho, J., Dimola, D., Duarte, C., Manca, M., ... & Santoro, C.: Towards a Platform for Persuading Older Adults to Adopt Healthy Behaviors. In: Orji, R., Reisinger, M., Busch, M. Dijkstra, A., Kaptein, M., Mattheiss, E. (eds.): Proceedings of the Personalization in Persuasive Technology Workshop, Persuasive Technology 2017, Amsterdam (2017).
3. Cvetković, B., Gjoreski, M., Šorn, J., Frešer, M., Bogdański, M., Jackowska, K., Kosiedowski, M., Luštrek, M.: Management of Physical, Mental and Environmental Stress at the Workplace. In: International Conference on Intelligent Environments (IE), pp. 76-83, IEEE (2017).
4. Andò, B., Baglio, S., La Malfa, S., Marletta, V.: Innovative Smart Sensing Solutions for the Visually Impaired. In: Pereira, J. (ed.) Handbook of Research on Personal Autonomy Technologies and Disability Informatics, pp. 60-74, IGI Global, Hershey (2011).
5. Praveen, N., Karanth, N., Megha, M. S.: Sign language interpreter using a smart glove. In: International Conference on Advances in Electronics, Computers and Communications (ICAECC 2014), pp. 1-5, IEEE (2014).
6. Atkin, K., Lorch, M. P.: An ecological method for the sampling of nonverbal signalling behaviours of young children with profound and multiple learning disabilities (PMLD). Developmental neurorehabilitation 19(4), 211-225 (2016).
7. Rotter, B., Kane, G., Gallé, B.: Nichtsprachliche Kommunikation: Erfassung und Förderung. Geistige Behinderung 31, 1–26 (1992).

8. Hall, S. S., Arron, K., Sloneem, J., Oliver, C.: Health and sleep problems in Cornelia de Lange syndrome: a case control study. Journal of Intellectual Disability Research 52(5), 458-468 (2008).
9. Vos, P., Cock, P. de, Munde, V., Petry, K., van den Noortgate, W., Maes, B.: The tell-tale: what do heart rate, skin temperature and skin conductance reveal about emotions of people with severe and profound intellectual disabilities? Research in Developmental Disabilities 33(4), 1117–1127 (2012).
10. Rowland, C., Fried-Oken, M.: Communication Matrix: A clinical and research assessment tool targeting children with severe communication disorders. Journal of Pediatric Rehabilitation Medicine 3, 319–329 (2010).
11. Ross, E., Oliver, C.: Preliminary analysis of the psychometric properties of the Mood, Interest and Pleasure Questionnaire (MIPQ) for adults with severe and profound learning disabilities. British Journal of Clinical Psychology 42, 81–93 (2003).
12. Breau, L. M., McGrath, P. J., Camfield, C. S., Finley, A. G.: Psychometric properties of the non-communicating children's pain checklist-revised. Pain 99, 349–357 (2002).
13. Lotan, M., Moe-Nilssen, R., Ljunggren, A. E., Strand, L. I. Reliability of the Non-Communicating Adult Pain Checklist (NCAPC), assessed by different groups of health workers. Research in Developmental Disabilities 30, 735–745 (2009).
14. Regnard, C., Reynolds, J., Watson, B., Matthews, D., Gibson, L., Clarke, C.: Understanding distress in people with severe communication difficulties: developing and assessing the Disability Distress Assessment Tool (DisDAT). Journal of Intellectual Disability Research 51(4), 277–292 (2007).
15. ELAN (Version 5.2) [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, https://tla.mpi.nl/tools/tla-tools/elan/, last accessed 2019/01/15.
16. Brennan, Kevin, (ed.),: A Guide to the Business Analysis Body of Knowledger, Iiba (2009).
17. Ekman, P., Rosenberg, E. L. (Eds.).: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997).
18. OpenPose: Multi-Person Pose Estimation [Computer software]. Carnegie Mellon, The Robotics Institute, http://www.consortium.ri.cmu.edu/projOpenPose.php, last accessed 2019/01/15.
19. Howard, A.: Elementary Linear Algebra. John Wiley & Sons, USA (2010).
20. Hwang, B. W., Kim, S., Lee, S. W.: A full-body gesture database for automatic gesture recognition. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06), pp. 243-248, IEEE (2006).
21. Theodorou, T., Mporas I., Fakotakis N.: Audio Feature Selection for Recognition of Non-linguistic Vocalization Sounds. In: Hellenic Conference on Artificial Intelligence, pp. 395-405, Springer, Cham (2014).
22. Truong, K. P., van Leeuwen, D. A.: Automatic discrimination between laughter and speech. Speech Communication 49 (2), 144-158 (2007).
23. Weninger, F., Schuller, B., Wollmer, M., Rigoll, G.: Localization of non-linguistic events in spontaneous speech by Non-Negative Matrix Factorization and Long Short-Term Memory. In: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 5840-5843, IEEE (2011).
24. Swarnkar, V., Abeyratne, U. R., Amrulloh, Y., Hukins, C., Triasih R., Setyati, A.: Neural network based algorithm for automatic identification of cough sounds. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), p. 1764-1767, IEEE (2013).

25. Amoh, J., Odame, K.: Deep Neural Networks for Identifying Cough Sounds. IEEE Transactions on Biomedical Circuits and Systems 10(5), 1003-1011 (2016).
26. Munde, V., Vlaskamp, C., Vos, P., Maes, B., Ruijssenaars, W.: Physiological measurements as validation of alertness observations: an exploratory case study of three individuals with profound intellectual and multiple disabilities. Intellectual and developmental disabilities 50(4), 300-310 (2012).
27. Vos, P. , De Cock, P., Munde, V., Petry, K., Van Den Noortgate, W., Maes, B.: The tell-tale: What do heart rate; skin temperature and skin conductance reveal about emotions of people with severe and profound intellectual disabilities?. Research in Developmental Disabilities 33(4), 1117-1127 (2012).
28. Lima, M., Silva, K., Amaral, I., Magalhães, A., De Sousa, L.: Beyond behavioural observations: a deeper view through the sensory reactions of children with profound intellectual and multiple disabilities. Child: care, health and development, 39(3), 422-431 (2013).
29. Vos, P. , De Cock, P., Petry, K., Van Den Noortgate, W., Maes, B.: Do you know what I feel? A first step towards a physiological measure of the subjective well-being of persons with profound intellectual and multiple disabilities. Journal of Applied Research in Intellectual Disabilities 23(4), 366-378 (2010).
30. Heusch, G., Anjos, A., Marcel, S.: A Reproducible Study on Remote Heart Rate Measurement. arXiv preprint arXiv:1709.00962 (2017).
31. Balakrishnan, G., Durand, F., Guttag, J.: Detecting pulse from head motions in video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3430-3437 (2013).
32. Lucas, B. D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the International Joint Conference on Artificial Intelligence 1981, pp. 674-679 (1981).
33. Michelson, B. M.: Event-driven architecture overview. Patricia Seybold Group 2(12), 10-1571 (2006).
34. Node-RED: Flow-based programming for the Internet of Things [Computer software]. JS Foundation, https://nodered.org/, last accessed 2019/01/15.