IM-TWIN: from Intrinsic Motivations to Transitional Wearable INtelligent companions for autism spectrum disorder

a European funded project

PlusMe AI-augmented behaviour and IM-TWIN 1

Deliverable 3.3

This project has received funding
from the European Union's Horizon 2020
Research and Innovation Program
Under Grant Agreement No 952095.

Project duration 24 months (November 2020, October 2022),
Consortium: Consiglio Nazionale delle Ricerche (ITA),
Universiteit Utrecht (NLD), Centre de Recherches
Interdisciplinaires (FRA), Università degli Studi di Roma
La Sapienza (ITA), Plux-Wireless Biosignals S.A. (PRT).

Deliverable data

Work Package:	3 Affect Classification, PlusMe AI and IM-TWIN integration
Work Package leader:	CNR-ISTC
Deliverable beneficiary:	CNR-ISTC
Dissemination level:	Public
Due date:	28th February 2022 (Month 16)
Type:	Demonstrator
Revision:	2 (April 2023)
Authors:	Francesco Montedori, Francesca Romana Mattei, Massimiliano Schembri, Valerio Sperati, Beste Özcan, Gianluca Baldassarre

Acronyms of partners

CNR-ISTC	Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie della Cognizione (Italy)
UU	Universiteit Utrecht (The Netherlands)
CRI	Centre de Recherches Interdisciplinaires (France)
LA SAPIENZA	Università degli Studi di Roma La Sapienza (Italy)
PLUX	Plux - Wireless Biosignals S.A. (Portugal)

Table of contents

1. Overview of the deliverable

This deliverable reports a new *PlusMe* software feature, based on an additional Computer Vision module, which potentially allows the toy to trigger rewarding patterns in response to specific child's facial expressions, relevant for social interaction, like smiles or eye contacts towards the therapist[1,2]. In possible, future experimental trials (which are not part of the current project[3]), this novel toy augmented behaviour could improve the child's engagement and the social interaction during the play activities with the caregiver.

The new software feature relies on several hardware components and software modules which constitute the main core of the IM-TWIN system, as described in the project proposal (see Fig. 1). Main goal of the deliverable is then to assess the reliability of such a system, where several wirelessly connected components exchange information in real-time.

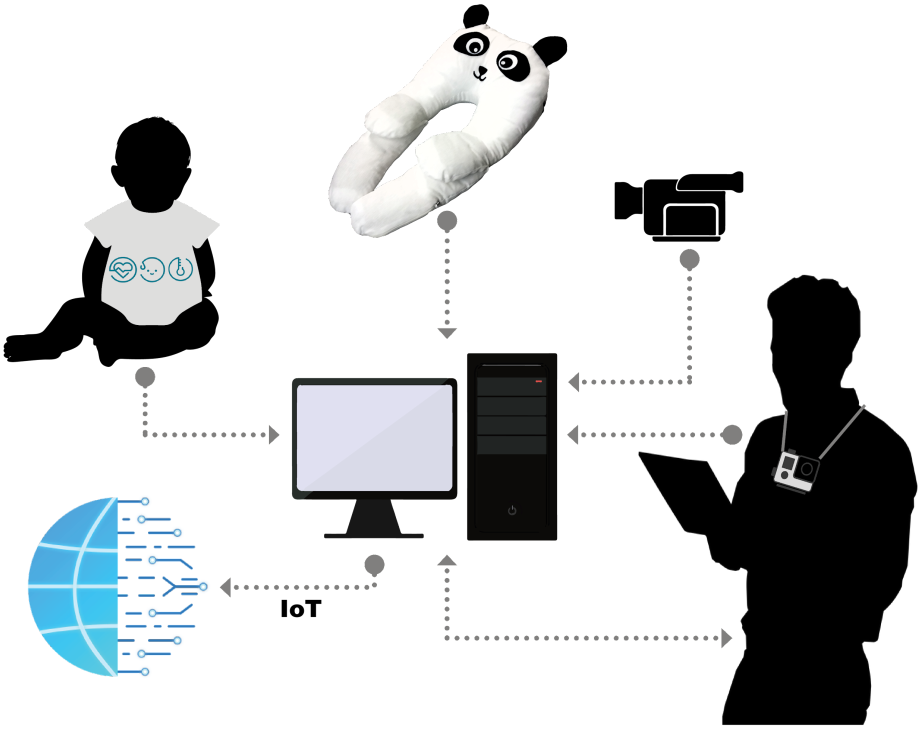

Figure 1: *the IM-TWIN system architecture, as described in the project proposal. The present PlusMe AI-augmented behaviour is based on a first implementation of such architecture, with the exclusion of the module in charge to process the sensorised t-shirt data, and the module for database collection/query.*

[1] Keating CT, Cook JL. Facial Expression Production and Recognition in Autism Spectrum Disorders: A Shifting Landscape. Psychiatr Clin North Am. 2021 Mar;44(1):125-139. doi: 10.1016/j.psc.2020.11.010. PMID: 33526234.
[2] Trevisan DA, Hoskyn M, Birmingham E. Facial Expression Production in Autism: A Meta-Analysis. Autism Res. 2018 Dec;11(12):1586-1601. doi: 10.1002/aur.2037. Epub 2018 Nov 4. PMID: 30393953.
[3] The further development of AI-based PlusMe behaviour, originally planned in the Task 4.4 "Pilot tests of PlusMe-AI", has been set aside in the project amendment AMD-952095-7.

2. System architecture

Here we describe the system setup we have designed in order to make the *PlusMe* potentially responsive (i.e. able to trigger rewarding patterns) to the changes in child's facial expressions. The current implementation focuses on the technical feasibility of the proposed system – namely it does not provide for experimental trials with children – therefore the roles of child and therapist are played by the researchers.

The novel Computer Vision module is based on the following IM-TWIN system components:

- a webcam wired to a self-powered *Raspberry Pi 4* board, whose goal is to record the user's face; the camera will be worn by the therapist in the next experiments to capture the child's face (see Fig. 2).
- a central computer, in charge of receiving the video stream from the webcam and performing in real-time the *Facial Expression Recognition* (FER) algorithm;
- the *PlusMe* toy and its control tablet, the latter receiving the outcome of the video processing and to trigger the rewarding patterns according to the FER outcomes.

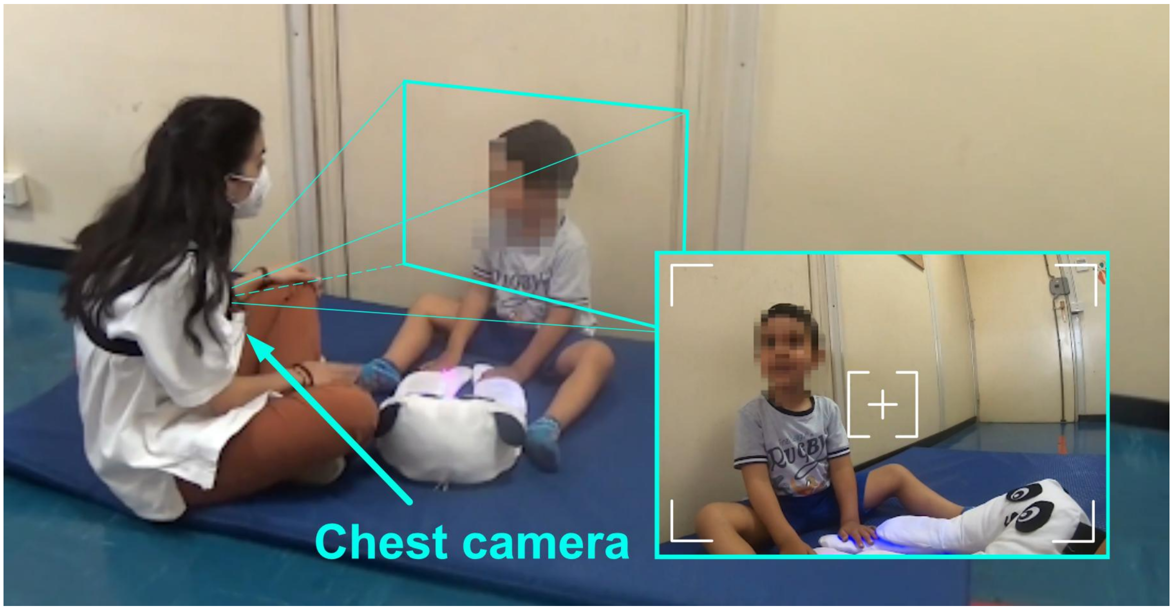

Figure 2: *a test showing how a camera worn by the therapist on her chest provides a good view of the child's face.*

Since the system should be usable in dynamic scenarios, as those involved in play-based therapy (see Fig. 2), the interaction between all the components is based on wireless communication, to ensure the maximum freedom of movement for the therapist.

The communication protocols and the main contents of data exchange are listed below (see Fig. 3):

- The Raspberry board sends the video captured by the webcam to the central computer, using the ZeroMQ TCP/IP protocol which is a common protocol relying on a wireless Internet connection. The video stream was set to provide a frame resolution of 480x480 pixels, and a speed of 14 Frame Per Second (FPS);
- The central computer, when a face is detected, performs FER on the received image and categorises the user's expression in 3 classes (namely Negative, Positive, Neutral). Then, it sends the detected class information to the *PlusMe* control tablet, via BLE protocol.
- The control tablet acts then as a bridge and broadcasts the information to the *PlusMe* device, which produces a rewarding pattern in case of detection of a positive facial expression. For this purpose, both the control App and the toy firmware have been partially rewritten.

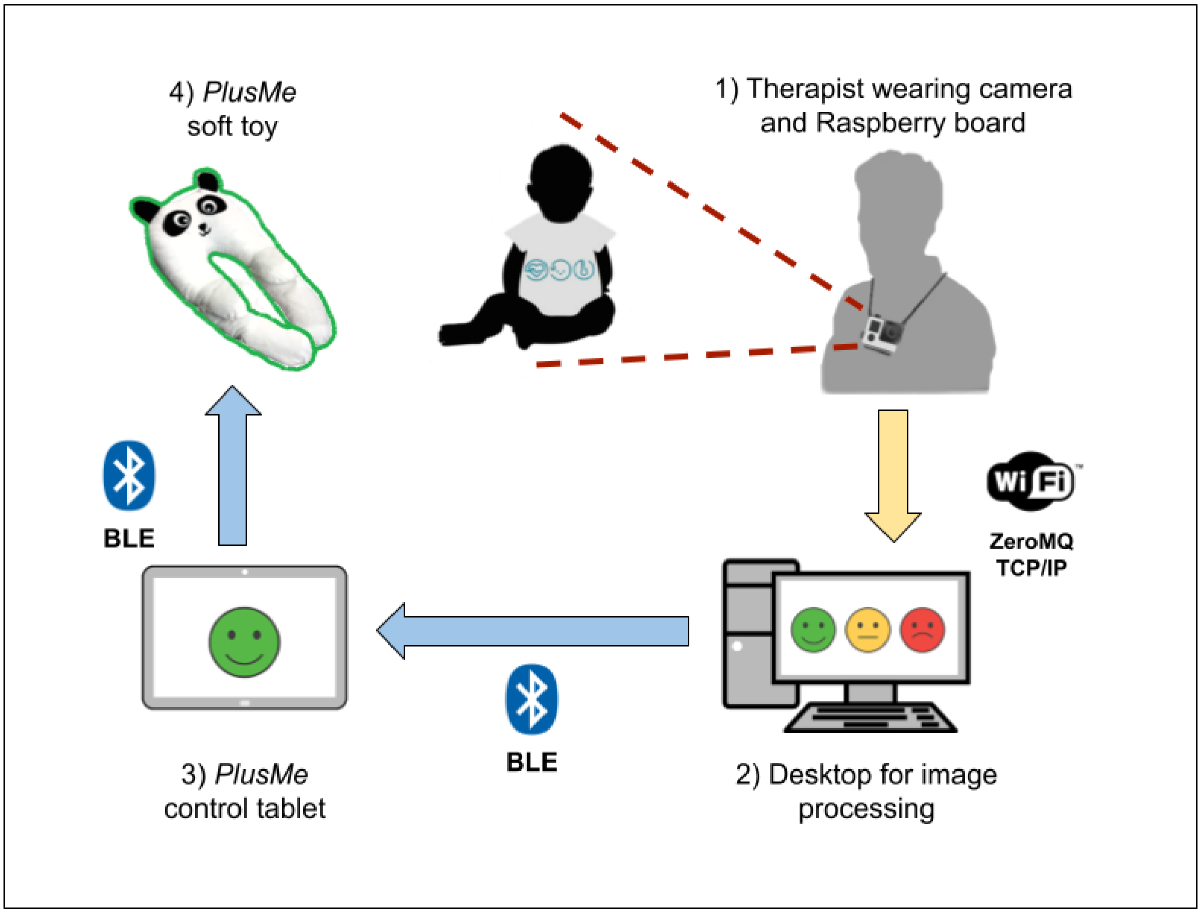

Figure 3: *PlusMe* augmented behaviour relies on a first implementation of the IM-TWIN core system: the child's facial expression (1) is processed by a central computer (2) which broadcasts the information to the tablet (3) in charge to trigger the PlusMe rewarding output (4).

3. Implementation

The real-time FER algorithm, processing the webcam images, is performed using a Neural Network (NN), trained with the RAF-DB dataset[4], whose output is a vector that describes the probability associated with each one of the six facial emotions (happy, sad, surprise, disgust, fear, anger), plus a neutral state.

Specifically we used a customised convolutional multilayer network[5], featuring a new type of architecture called *Amending Representation Module* to deal with the padding erosion problem as a substitute for the pooling layer. This network has been trained with square cropped images that show only the face of the subjects in order to ignore the background and reduce the noise.

Before running the above-mentioned network, we use a standard OpenCV face detection algorithm[6] to identify the face inside the image, extract it (ignoring background) and send it to the FER neural network. In this way, we send to the FER neural network only cropped faces, increasing classification accuracy.

In order to ensure that FER is performed exclusively when the user's face is frontal to the camera – it is assumed it happens when the child looks towards the therapist, a relevant event for the social interaction – we decided to compute face landmarks on the image using a pre-trained neural network[7]. We then calculate the distance between specific landmarks and compute the ratio of these lengths: these values are used to decide whether a frame should be processed (i.e. the user's face is frontal) or not (i.e. the user's face is not frontal).

The speed of the system depends both on the quality of the Internet connection and on the specifics of the machine where the algorithms run. We achieved a speed of 14 FPS using an Asus gaming laptop (Intel core i7 10th generation and NVIDIA RTX 2080 Super GPU) and being connected to the GARR national network[8]. Our test shows how this performance is indeed enough to make the toy respond in real-time to facial expressions (Fig. 4).

A brief video demo is available at the link https://bit.ly/35o02Nq. For sake of clarity, in the video the toy changes colours according to all three detected emotions. Since our main purpose is to potentially detect the child's positive engagement, it is enough to identify whether the user's expression corresponds to a positive (happy, surprise), negative (fear, sad, disgust, angry), or neutral emotion; thus we grouped the six emotions in these three classes. Also, in the video the webcam is fixed on a steady support, simulating the therapist sitting in front of the child.

[4] http://www.whdeng.cn/raf/model1.html

[5] https://github.com/JiaweiShiCV/Amend-Representation-Module

[6] https://docs.opencv.org/4.x/d2/d58/tutorial_table_of_content_dnn.html

[7] https://github.com/yinguobing/cnn-facial-landmark

[8] https://www.garr.it/it/

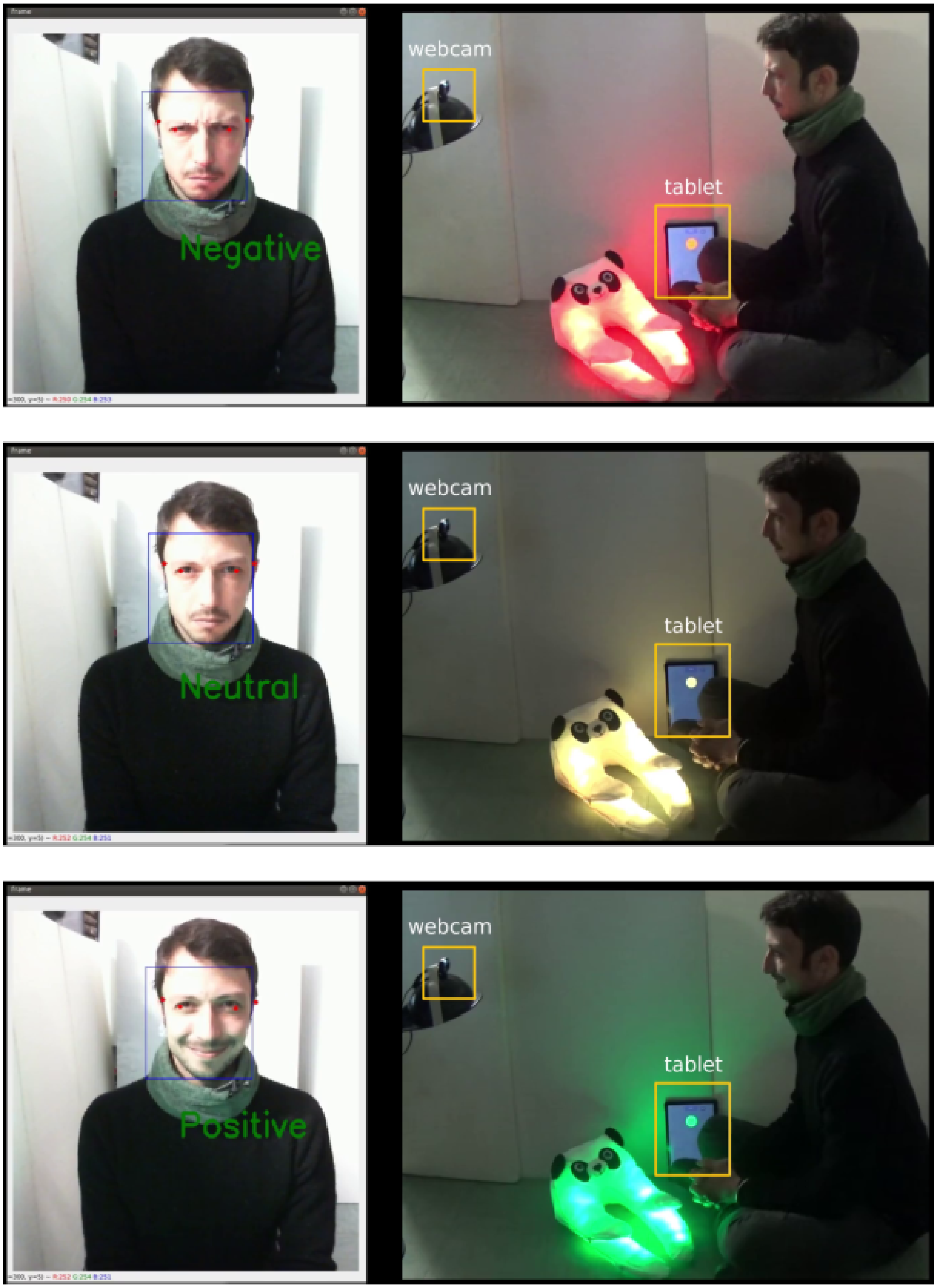

Figure 4: Three frames extracted from the demonstration video. For each row, the left panel shows the image captured by a frontal fixed webcam (to be worn on the therapist's chest in the future implementation); the panel also displays the detected bounding box, the selected facial landmarks and the detected facial expression, as processed by FER algorithm. The right panel presents the experimental setup; for sake of clarity, PlusMe changes colour according to all three detected emotions (negative, neutral and positive facial expressions).

4. Eye Contact detector tool

As shown in deliverable D2.1 “*Processing of physiological signals, visual info, and PlusMe interaction, first version”* (section 3 “*Processing of visual information*”)[9], CNR-ISTC developed a reliable tool (henceforth *camera glasses*) to detect the eye contact between child and therapist[10] (fig 5). The tool, based on a recent algorithm[11], was realised to collect behavioural data during the play activities with TWC devices (fig. 6).

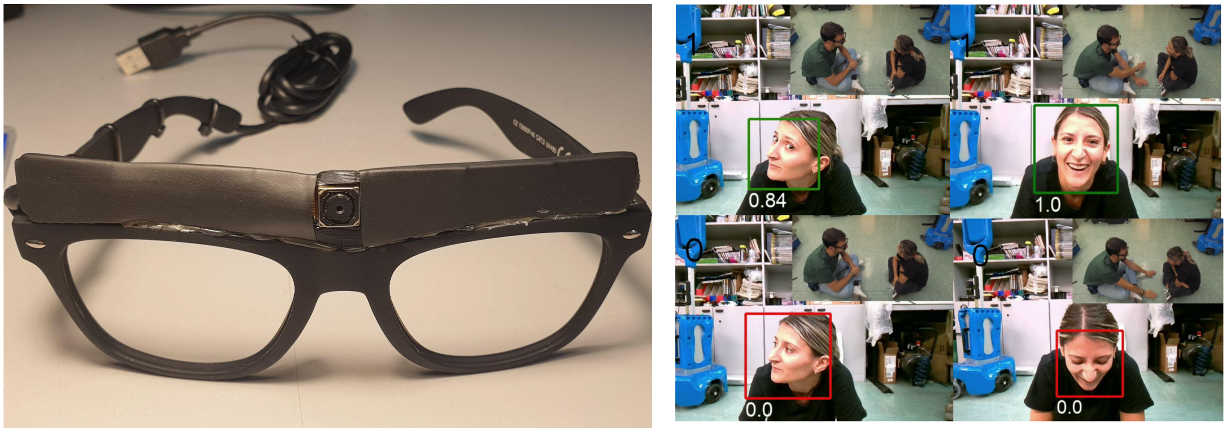

Figure 5, left: the *camera glasses* tool is realised by embedding a micro camera in a pair of customised glasses; **right:** a test involving two researchers, showing how the tool correctly detects the eye contact.

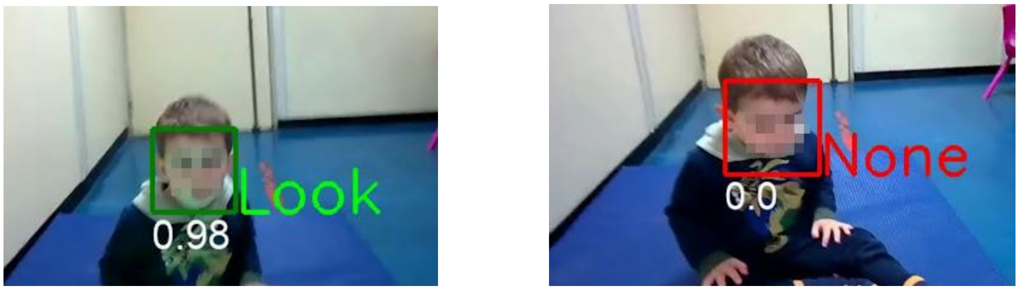

Figure 6. An example of output by camera glasses during a test with a child. **Left**: when the child looks at the therapist, the output of Artificial Intelligence indicates this event with a green box and the probability that the child is actually looking into the eyes is shown. **Right**: the AI output produces a red box when no eye contact is detected.

[9] https://im-twin.eu/wp-content/uploads/2023/02/DELIVERABLE_D2.1_Processing_of_physiological_signals_PlusMe_interaction.pdf .

[10] A video of the tool is available at https://im-twin.eu/video/#eye_contact_detector

[11] Chong, E., Clark-Whitney, E., Southerland, A. *et al.* Detection of eye contact with deep neural networks is as accurate as human experts. *Nat Commun* **11**, 6386 (2020), https://doi.org/10.1038/s41467-020-19712-x

This new, AI-based tool, in combination with the system setup described in the previous paragraph, could be used in order to make the *PlusMe* potentially responsive (i.e. able to trigger rewarding patterns) to the detection of this important social behaviour. For example, it is possible to envisage an experiment where *PlusMe* produces different colours and amazing sounds when the child looks into the therapist's eyes. This automated, plausibly enjoyable, toy behaviour could stimulate the child to increase the frequency of eye contact toward the therapist, so as to reinforce this critical social behaviour, generally atypical and impaired in ASD children.[12,13,14,15].

As described in deliverable D4.1 "*Empirical validation: PlusMe*" (section 5.1.2 "*Camera glasses*")[16], the *camera glasses* tool has been used in a pilot study involving 6 children with neurodevelopmental disorders, to test the reliability of collected data (Fig. 6). The results showed how the performance of the tool is comparable to human rating (*Inter Rater Reliability* index IRR >= 0.8).

5. Future Development

This deliverable presents the first technical implementation of the IM-TWIN system, where different components (described in fig. 3), are connected to each other and exchange information in real-time. Such implementation, based on a Computer Vision module, provides the *PlusMe* toy with a novel interactive feature which, if properly developed[17], could potentially improve the social interaction between child and therapist during play activities.

About the next, complete IM-TWIN system implementation, the current architecture will also integrate the output of the processing of physiological data collected by the sensorised t-shirt. This new source of information, along with the video information, will be integrated and will provide the therapist with the general feedback about the child's social/emotional state and engagement during the therapeutic activities, as described in the project proposal (Fig 6).

[12] Senju A, Johnson MH. Atypical eye contact in autism: models, mechanisms and development. Neurosci Biobehav Rev. 2009 Sep;33(8):1204-14. doi: 10.1016/j.neubiorev.2009.06.001. Epub 2009 Jun 16. PMID: 19538990.

[13] Hirsch J, Zhang X, Noah JA, Dravida S, Naples A, et al. (2022) Neural correlates of eye contact and social function in autism spectrum disorder. PLOS ONE 17(11): e0265798. https://doi.org/10.1371/journal.pone.0265798

[14] Tanaka JW, Sung A. The "eye avoidance" hypothesis of autism face processing. J Autism Dev Disord. 2013;46(5):1538–52.

[15] Madipakkam AR, Rothkirch M, Dziobek I, Sterzer P. Unconscious avoidance of eye contact in autism spectrum disorder. Sci Rep. 2017 Oct 17;7(1):13378. doi: 10.1038/s41598-017-13945-5. PMID: 29042641; PMCID: PMC5645367.

[16] https://im-twin.eu/wp-content/uploads/2023/03/DELIVERABLE_D4.1_empirical_validation_PlusMe_VERSION_2.pdf

[17] The further development of AI-based PlusMe behaviour, originally planned in the project Task 4.4 "Pilot tests of PlusMe-AI", has been set aside in the project amendment AMD-952095-7

About the FER algorithm, described in sec. 3, the use of the new tool *camera glasses* during play activities with children at SAPIENZA, lets to envisage the forthcoming collection of a new dataset, useful to refine the original training set “RAF-DB”, and then improve the FER performance to acceptable results when treating data from ASD participants[18]. In this regard, the videos collected by SAPIENZA will be processed using a *Face Detection* algorithm, to extract only those frames in which a face is correctly detected; the resulting images will be then labelled by SAPIENZA specialists using the open source tool 'Label Studio'[19] to assess the correct child's facial expression; finally, the newly labelled dataset will be used to refine the training of the neural network for *facial expression recognition* of ASD participants. It is also important to note that the new dataset could be further improved, by involving in the data collection new institutes treating ASD children, which expressed an interest in joining the IM-TWIN experimental trials[20].

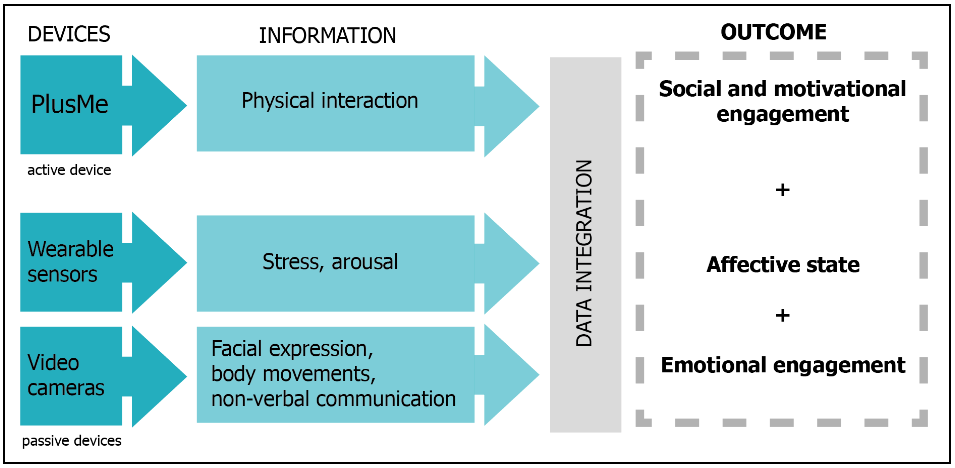

Figure 6: the different sources of information to detect the general child’s level of social engagement and affective/emotional state.

[18] At the date of April 2023, despite a formal request to Georgia Tech Institute, it was not possible to access to the *Multimodal Dyadic Behaviour* MMDB dataset, specific for ASD children, and described at the link https://cbs.ic.gatech.edu/mmdb/

[19] https://labelstud.io/

[20] See deliverable D5.8 “Identification of target groups and relevant stakeholders 2”, https://im-twin.eu/deliverables/