

Person-Independent Multimodal Emotion Detection for Children with High-Functioning Autism

Annanda Sousa¹, Mathieu d'Aquin¹, Manel Zarrouk² and Jennifer Holloway³

¹Data Science Institute - National University of Ireland - Galway

²Institut Galilée - Université Paris 13

³School of Psychology - National University of Ireland - Galway

{a.defreitassousa1, mathieu.daquin, jennifer.holloway}@nuigalway.ie, zarrouk@lipn.univ-paris13.fr

Abstract

The use of affect-sensitive interfaces carries the promise of enhancing human-computer interaction by delivering a system capable of identifying a user's emotions and adapt its content accordingly. Today's technology shows great potential to support children with autism, for example by using computer systems to improve their social skills. Generally, however, this technology does not encompass the potential of affect-sensitive interfaces. This is mainly due to Emotion Detection (ED) models built for the general population usually not performing well when applied to children with autism, who express emotions differently. The aim of this project is therefore to build a person-independent Multimodal Emotion Detection system tailored for children with high-functioning autism for the ultimate goal of applying it to design affect-sensitive interfaces dedicated to children with autism. This is a work in progress and the project expects to build upon the current body of knowledge on methods to apply ED systems to this specific subset of the general population. We expect to apply the overall theoretical and practical design perspectives that arise from this research investigation (e.g. analysis of modalities and features extraction, behavioural cues based features, fusion layers and classifier techniques) to propose a guiding framework for future studies.

1 Introduction

Automatic Emotion Detection (ED) aims to automatically identify people's cognitive states or emotions, e.g. happiness, anger, fear using different types of media inputs such as texts, video, audio and sensor signals. When combining more than one type of data, they are called *Multimodal* Emotion Detection systems and usually outperform unimodal systems.

Automatic ED is advancing to become an important component of Human-Computer Interaction (HCI) through affect-sensitive systems. An affect-sensitive system detects the user's emotions and automatically adapts its interaction with the human based on their emotions. This kind of features has the potential to enhance HCI, creating an individualised

experience for the user in a more human-to-human-like interaction.

Even with all the advancement of ED for users with typical neurological development, usually referred to as neurotypical, when applying those systems to children with autism they do not perform well, mainly because of this particular population's way to express emotions [Liu *et al.*, 2008], motivating the need to develop ED systems specifically tailored for children with autism. Autism Spectrum Disorder (ASD) is a developmental disorder with spectrum manifestation of traits, characterised by impairments in social interaction, communication and repetitive patterns of behaviour and interests. High-Functioning Autism (HFASD) is defined as ASD without significant cognitive and language impairments [Gaus, 2011].

Among the results of a recent meta-analysis [Trevisan *et al.*, 2018] that compared the facial expression production between a typical development (TD) population and people with ASD, we can find evidence that people with ASD display facial expressions less often and less frequently than people with TD. Also, their expressions are found to be lower in quality and less accurate. In the work of [Grossard *et al.*, 2020], the results show that a Random Forest model needs more facial landmarks to classify facial expressions from children with ASD than it needs from children with typical development. Providing more evidence that ED systems developed for children with typical development do not perform well when applied to children with ASD.

Nowadays, the development of computer-based interventions tools for the treatment of children with autism has increased, turning technology to an important ally when it comes to teaching those children abilities they lack in social and emotional areas [Frauenberger *et al.*, 2012]. There are several examples of computer systems [Hopkins *et al.*, 2011], virtual reality (VR) environments [Boyd *et al.*, 2018], tablets and mobile applications [Hourcade *et al.*, 2012], and even robotic agents that interact with children with ASD as intervention tools [Rudovic *et al.*, 2018; Marinoiu *et al.*, 2018]. Studies have shown evidence demonstrating the effectiveness of such tools to support ASD [Ma *et al.*, 2019]. Additionally, new methods are emerging on the use of technology to support people on the autism spectrum beyond assistive and intervention tools, shifting the focus from just "fixing the problem" to a more holistic approach [Frauenberger *et al.*,

2016]. This includes, for instance, investigating ways to design technologies to support children with autism considering their special interests and strengths.

Being able to automatically identify emotions from children with autism can represent an important role in enhancing and individualising HCI between children with ASD and computer interfaces specially designed to support their needs and particularities [Sharmin *et al.*, 2018]. Regardless, most technological tools that have been developed to support children on the autism spectrum do not use automatic ED which could be of great relevance to turn them into significant supplementary support to classic interventions that are usually expensive and very much dependent on human presence. Another point of importance is that creating ED systems tailored for children with autism is another step towards inclusion: Tools based on ED are being developed focusing on neurotypical people, which will not be usable by children with ASD if not adapted to their ways of expressing emotions. Some examples of ED application areas are Gaming, Health and Mental Health, which currently does not include the population on the autism spectrum.

In the field of Emotion Detection, creating a person-independent model is one of many well-known challenges [Cambria *et al.*, 2017]. This challenge refers to building a model that performs on identifying emotions from people which data were not present on the model training dataset. At a high level, it is related to the fact that people express emotions in an individualised manner. General patterns on expressing emotions are typically applicable for most people, e.g. smiling usually means happiness, however only considering general patterns is not enough to build an Emotion Detection (ED) system that takes into considerations individual and specific cues to express emotions.

This fact is still true for people inside the Autism Spectrum Disorder. On the one hand, people with ASD do not express emotions in a similar way to people with typical development. On the other hand, as for the general population, there is not a uniform way how people with autism express their emotions. As a consequence, creating a person-independent ED systems that models and reflects how this specific population expresses emotions is needed.

2 Related Work

Previous studies have developed ED systems tailored to children with autism. The studies created their ED models envisioning different applications: to allow the creation of affect-sensitive computer-based intervention tools [Liu *et al.*, 2008] and affect-sensitive e-learning platforms for children with ASD [Dawood *et al.*, 2018; Chu *et al.*, 2018], to generate knowledge to support the assessment of autism [Samad *et al.*, 2018], to support the treatment of anxiety, a common co-occurring condition in people with ASD [Kushki *et al.*, 2015], and also to allow the creation of a VR-based platform as intervention tool [Bekele *et al.*, 2016; Saadatzi *et al.*, 2013].

They all have in common that they did not focus on identifying the 7 basic emotions (i.e. fear, happiness, sadness, anger, disgust, surprise and contempt) because they argued that, although the ED field focuses on those basic

emotions, they are not the most relevant in the context of autism. Therefore, they chose to target different emotion states more suitable to the autism context, e.g. liking, anxiety, engagement [Liu *et al.*, 2008] and calmness [Chu *et al.*, 2018]. Another common characteristic is the fact that they only used one modality of data input for emotion identification: physiological signals (e.g. heart rate, skin conductivity) [Liu *et al.*, 2008; Bekele *et al.*, 2016; Kushki *et al.*, 2015; Sarabadani *et al.*, 2018] and video media input (e.g. facial expressions, eye gaze, head movement) [Dawood *et al.*, 2018; Chu *et al.*, 2018; Ahmed and Goodwin, 2017]. Also, they all used machine learning techniques to create the classifier model, which is the state-of-the-art of general Emotion Detection models (i.e. models for the neurotypical population). One more common point is that all of them needed to develop and conduct an experiment to elicit emotions from children with autism in order to create an annotated dataset. Despite that, none of the datasets were made available for the research community mostly due to privacy issues.

Together these studies provide important evidences to show that it is viable to model and automatically identify emotions of children with ASD. However, such studies remain limited when considering two points: input multimodalities and generalisability of the model. To the best of our knowledge, none of their models used multimodal input data for emotion identification and most of the works created models that are individual-specific.

Multimodal inputs have been used and explored in the Emotion Detection field, where studies showed that multimodal Emotion Detection usually outperform unimodal emotion detection models. Regarding individual-specific approach, it means that the ED model created was trained separately for each individual child, becoming very good at identify emotions from that specific child, but not performing well when applied to other children. This creates a huge impediment for using person-specific Emotion Detection systems in realistic conditions because every time a new child would use the system, it would require the model to be trained on their annotated data.

3 Research Objectives

This research seeks to advance ED systems tailored to children with autism by exploring ways to design and develop a person-independent Multimodal Emotion Detection system to be used by children with autism. The ultimate goal of this research is to enable the benefits of ED on HCI for children with ASD. Hence, during this project, we aim to answer the following Research Question:

RQ1: How to create a multimodal Emotion Detection system which:

- i) Is tailored to how children with high-functioning autism express emotions;
- ii) Is person-independent, i.e. reach an equivalent or higher accuracy than state-of-the-art person-independent ED systems for the neurotypical population, when applied to children with ASD not involved in training the model.

To be able to answer **RQ1**, we further need to explore an-

swers to the following research question:

RQ2: How to build a ground truth dataset annotated with the emotion states we aim to identify?

RQ3: Which modality input(s) and features are more relevant for cues extraction in the context of multimodal Emotion Detection for children with autism?

RQ4: Which data fusion methods work better in the context of multimodal Emotion Detection for children with autism?

4 The proposed system

Considering the objectives stated above, the proposed multimodal ED system tailored for children with high-functioning autism will involve four input modalities: video, audio, text and physiological signals (i.e. heart rate measure). These four modalities were selected for the feasibility of achieving data acquisition by the user’s family and are widely used on the field of ED. Based on those input, our model will use features extracted from: facial expressions, body movements, the words content of the speech, the tone of the voice and the heart rate values. All of them are broadly used in the ED field.

Following the previous related works, we will not focus on identifying the 7 basic emotions, i.e. surprise, happiness, anger, disgust, contempt, sadness and fear. Instead, we will use a framework of emotion zones for regulation [Kuypers, 2013] that is extensively used in psychology to help children with ASD to learn emotion regulation. It is common for children with ASD to present impairments in emotion regulation that is manifested by they finding it hard both to understand their emotions and to calm down after they leave a calming state [Scarpa and Reyes, 2011]. A child with ASD necessitates being in a calming emotional state to be able to listen, to interact and to learn.

The zones of regulation framework has 4 different zones, that are represented by colours (See Figure 1). One of the emotions zones is the **calming zone**, represented by the green colour. This is the ideal state, where the child is calm, relaxed and ready to work, to listen and to interact. Another emotion zone is the **warning zone** (yellow colour). In this state, the child is presenting signals of agitation or excitement. This state can originate from both positive and negative emotions. It can start from intense happiness or excitement, and also from frustration. The following zone is the **high-agitation zone**, with a red colour. In here the child is really upset or angry, presenting serious difficulties in keeping control of their emotions. The last zone is the **slowing zone**, with a blue colour, in which the child is on low energy and showing emotional signals of being sad, tired, sick or bored. The child, here, might move slower than usual, stop speaking or show delays in responding to interaction.

This project is developing a classifier able to identify which of the four emotions zones a child with HFASD is engaged with using multimodal inputs of data. By choosing to use this emotions zones’ framework we obtain some benefits: firstly, the framework additionally includes guidelines on activities to lead children back to the calming zone, making it easy to incorporate such activities within an affect-sensitive interface. Second, parents of children with ASD are more likely to

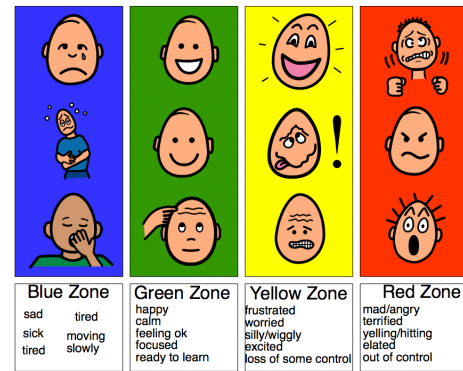


Figure 1: The four emotions zones for regulation framework.

be familiarised with this framework because it is commonly used in the context of autism, hence making the tagging task by the parents more comfortable. Thirdly, considering the children’s well-being, it is less harmful for the emotional comfort of children with HFASD, during the emotion elicitation experiment, to elicit the four emotion zones than other strong negative emotions, e.g. fear, anxiety (more about data collection in Section 6).

5 Methodology

The methodology’s pipeline to address our research goal is depicted on Figure 2, encompassing five different stages.

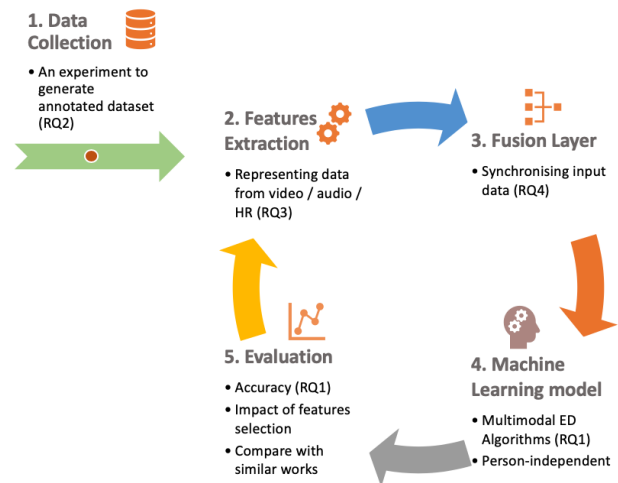


Figure 2: The pipeline of this research project.

In order to answer our Research Questions, we will follow the methodology:

- to conduct a study with human participants to elicit, capture and tag different emotion zones expressions, for dataset creation (RQ2);
- to use the standard ED methodology to build a multimodal ED classifier by iterating over the steps:
 - features extraction (RQ3);

- (b) fusion information layer design (RQ4);
- (c) training and testing of machine learning models using annotated dataset (RQ1);
- (d) evaluation by designing experiments to analyse the relation between the type of data input, features and data fusion techniques, and the accuracy of the model and compare to previous works.

Therefore, our first challenge to address is to obtain the ground truth dataset (phase 1). To achieve this, we have finished the design and planning of the experiment for data collection (See Section 6). For the subsequent phases of this research, we plan to follow the general approach of investigating the state-of-the-art methods applied for the population without ASD, evaluate their performance within our dataset and propose on how we can extend those methods to the population of children with HFASD.

6 Data collection

As a required component for meeting the aim of this research, we have to create an annotated dataset featuring children with ASD expressing emotions because previous related works did not make available any working dataset. To do so, we need to conduct a behavioural experiment with human participants to elicit, capture and tag emotions.

There is no way to directly observe an emotion because it is an internal experience of an individual, what we can do is to define and capture behavioural indices of the presence of a given emotion. Also, emotions do not just appear out of nowhere, they are usually an individual response of a physical or mental event, i.e. an event in the real world or thought, thus we need to evoke them.

During the experiment, we intend to collect the behavioural indices that children with ASD engage with when they are in different emotion zones, together with the measurement of their heart rate. Examples of behavioural indices are facial expressions and body movements such as smiling, flapping hands, head movement. We will ask the participants to perform tasks expected to evoke the emotion zones while we capture the participant's behaviour using different data inputs, i.e., video, audio and heart rate. We will extract features from these data to train a multimodal emotion detection system to identify the four emotion zones from a child with ASD.

For the study, we will recruit 12 children of the age of 8 to 12 years old and their parents/guardians as participants¹. The aimed participant number is an average of the number of subjects selected by the related studies (See Section 2). These works reported that it was challenging to recruit participants, and they had to operate with a small number of subjects for their models. To be considered part of this study, the child must 1) have a previous history of diagnosis of ASD, 2) not have a history of language or intellectual disability, 3) have their parents or guardian consent to the participation of the study. Participation in the study involves performing emotion eliciting tasks during three different sessions. Each session is expected to last around 30 minutes. We will use a

computer-based task environment, i.e. the child will interact with a computer for the task's execution.

We developed web-based software to serve as a task environment. During the experiment, the child will interact with a computer using the task environment interface. This software is a sequence of tasks expected to elicit each of the emotion zones. Between each zone elicitation, we will add calming content to help the child to calm down between emotion zone's tasks, to both minimise any stress and set a baseline of emotions between the elicitation part. Also, it finalises the session with calming content. The tasks for eliciting each zone are as follow video content for zones green, blue and calming activity, a game for zone yellow and a set of Math questions for the red zone. We selected the eliciting tasks with the input of psychologists with vast experience on work with children with ASD.

We decided the emotion zones' elicitation order by considering first the participant's well-being. So, the green zone starts the session to be sure we will not cause any negative emotion in the beginning, scaring the child. We then create a crescendo of emotion zones by eliciting the yellow zone followed by the red zone. This way, by asking the child to solve a demanding worksheet (red zone task elicitation), they will already be over-excited by having played the game before (yellow zone task elicitation). The blue zone was selected to be the last because, by the end of the session, it is expected that the participant will already show signs of being tired, therefore becoming easier to elicit the blue zone. To be the most effective in eliciting the four emotion zones, before the session we will ask the parents to answer a questionnaire to outline examples of content that usually makes their child move to a certain emotion zone. Based on the content of this questionnaire we will adapt the task environment's content to be individualised for each child.

We will annotate the data collected into four different categories, each of them representing one of the emotion zones. The annotation will also include behavioural markers, we will require from the annotator to select which behaviour they observed that supported their selection of a given emotion zone. None of the previous works used the emotion zone's framework as target emotions to identify, hence comparing results with their works will not be straightforward. To minimise this gap and have some measure of comparison we will in parallel annotate the dataset to include the happiness/unhappiness/neutral emotion. Happiness/unhappiness is a measure of Quality of Life (QoL) [Ramey *et al.*, 2019], and have being used as independent variable to analyse the effectiveness of interventions in Psychology. We decided to not only target the happiness/unhappiness emotion for this project because this emotion alone does not have the power to represent if a child with ASD is in an optimal state for learning. A child with ASD can become overexcited and agitated because of happiness and not being able to stay still for learning until they calm down, for example.

The children's parents or guardians will perform the annotation task after the eliciting sessions. We will also recruit psychology students to act as blind annotators. It is part of our future work to define which agreement measure we will use to annotate the dataset. In this case, the parents are the

¹<http://emotion-asd.datascienceinstitute.ie/>

specialists of identifying their child’s emotions because they know them, but parents also can have biases that an annotator who does not know the child would not present. Thus, it is important to define metrics of which annotation has more weight in case of disagreement.

We have developed web-based software to support the annotation task. The annotator will watch the video record from the study session and will have four different clickable buttons on the screen representing each emotion zone. We will instruct the annotator to click on the button to select an emotion zone, as soon as they identify the emotion zone in question. When they select a zone, the system asks the annotator to indicate which behavioural indices were present that guided them on their emotion zone decision. Some examples of behavioural indices they can point are a body movement, a facial expression, hands’ movements, a word said, etc.

To create the multimodal annotated dataset, we will follow the methodology used by the authors of the RECOLA dataset [Ringeval *et al.*, 2013]. RECOLA is a multimodal annotated dataset that has the same modalities we intend to include in this study, i.e. video, audio and physiological signals, and it was used as a benchmark dataset for several multimodal emotion detection challenges. They divided the sessions’ records into videos of 5 minutes and annotated fixed time windows of 400 milliseconds. They also balanced the training, validation and test datasets according to the annotation distribution.

Before running the experiments, we are going to conduct pilot sessions with the participation of children with typical development from the same age range of 8-12 years. By running a pilot session, we intend to test the experiment protocol, data collection, data synchronisation and data analysis steps. We expect to verify if the format of the data we will collect can be used within the data transformation, analysis and creation of a multimodal emotion detection system. We will also test the task environment software and the annotation software. With the information collected during the pilot sessions, we will iterate over the experiment protocol, to add any needed improvement identified during pilots. The collected pilot data will not be published and will be dealt with the same planned measures for data protection and privacy as the data from the posterior study. The results from the pilot will not be included in the project’s results.

This research was reviewed by the Institution’s Research Ethics Committee and the Data Protection Office at NUI Galway and obtained full approval.

7 Conclusion

We presented a work in progress of an emotion detection system tailored for children with high-functioning autism. The model’s novelty involves mainly two points: the inclusion of several data input modalities, and it is a personal-independent model. The input modalities involved in the proposed model are video, audio and heart rate. The main foreseen contribution of this research work is the creation of a person-independent Multimodal Emotion Detection model to be integrated into affect-sensitive systems that support children with autism. The affect-sensitive systems, thanks to this research’s

work, will be able to identify the child’s emotional zones and suggest/present activities to bring the child back to a calming emotional state based on which emotional state the child is at the moment.

Also, it is part of this project’s scope to make the multimodal dataset available to the research community. In order to protect the data subject’s privacy rights, the dataset will be formed by the extracted features from the original raw audio/video files together with heart rate measures. Therefore, it will only contain non-identifiable data.

Finally, this project expects to build upon the current body of knowledge on methods to apply Emotion Detection systems to this specific subset of the general population. We expect to apply the overall theoretical and practical design perspectives that arise from this research investigation (e.g. analysis of modalities and features extraction, behavioural cues based features, fusion layers and classifier techniques) to propose a guiding framework for future studies.

Currently, we had to temporally pause the experiments for data collection. So, we are working on the next phases of the methodology pipeline, investigating the state-of-the-art person-independent multimodal emotion detection systems for the general population to later propose how to adapt them to the population with ASD.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2, co-funded by the European Regional Development Fund.

We are grateful to Aindrias Cullen for providing us with comprehensive advice on data protection legislation, so we could design a project that is compliant with GDPR. We thank Dr Ciara Gunning for providing us with specialised advice on how to work with children with ASD and how to design the data collection experiment, as well as her support for recruiting participants for this study.

References

- [Ahmed and Goodwin, 2017] Alex A Ahmed and Matthew S Goodwin. Automated detection of facial expressions during computer-assisted instruction in individuals on the autism spectrum. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6050–6055. ACM, 2017.
- [Bekele *et al.*, 2016] Esubalew Bekele, Joshua Wade, Dayi Bian, Jing Fan, Amy Swanson, Zachary Warren, and Nilanjan Sarkar. Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD). *Proceedings - IEEE Virtual Reality*, 2016-July:121–130, 2016.
- [Boyd *et al.*, 2018] LouAnne E Boyd, Saumya Gupta, Sagar B Vikmani, Carlos M Gutierrez, Junxiang Yang, Erik Linstead, and Gillian R Hayes. vrsocial: Toward immersive therapeutic vr systems for children with autism. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 204. ACM, 2018.

- [Cambria *et al.*, 2017] Erik Cambria, Devamanyu Hazarika, Soujanya Poria, Amir Hussain, and RBV Subramanyam. Benchmarking multimodal sentiment analysis. In *International Conference on Computational Linguistics and Intelligent Text Processing*, pages 166–179. Springer, 2017.
- [Chu *et al.*, 2018] Hui Chuan Chu, William Wei Jen Tsai, Min Ju Liao, and Yuh Min Chen. Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. *Soft Computing*, 22(9):2973–2999, 2018.
- [Dawood *et al.*, 2018] Amina Dawood, Scott Turner, and Prithvi Perepa. Affective Computational Model to Extract Natural Affective States of Students with Asperger Syndrome (AS) in Computer-based Learning Environment. *IEEE Access*, 6:67026–67034, 2018.
- [Frauenberger *et al.*, 2012] Christopher Frauenberger, Judith Good, Alyssa Alcorn, and Helen Pain. Supporting the design contributions of children with autism spectrum conditions. In *Proceedings of the 11th International Conference on Interaction Design and Children*, pages 134–143. ACM, 2012.
- [Frauenberger *et al.*, 2016] Christopher Frauenberger, Judith Good, and Narcis Pares. Autism and technology: Beyond assistance & intervention. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 3373–3378. ACM, 2016.
- [Gaus, 2011] Valerie L Gaus. Cognitive behavioural therapy for adults with autism spectrum disorder. *Advances in Mental Health and Intellectual Disabilities*, 5(5):15–25, 2011.
- [Grossard *et al.*, 2020] Charline Grossard, Arnaud Dapogny, David Cohen, Sacha Bernheim, Estelle Juillet, Fanny Hamel, Stéphanie Hun, Jérémy Bourgeois, Hugues Pellerin, Sylvie Serret, Kevin Bailly, and Laurence Chaby. Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions: An experimental study using random forest classifiers. *Molecular Autism*, 11(1):1–14, 2020.
- [Hopkins *et al.*, 2011] Ingrid Maria Hopkins, Michael W Gower, Trista A Perez, Dana S Smith, Franklin R Amthor, F Casey Wimsatt, and Fred J Biasini. Avatar assistant: improving social skills in students with an asd through a computer-based intervention. *Journal of autism and developmental disorders*, 41(11):1543–1555, 2011.
- [Hourcade *et al.*, 2012] Juan Pablo Hourcade, Natasha E Bullock-Rest, and Thomas E Hansen. Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders. *Personal and ubiquitous computing*, 16(2):157–168, 2012.
- [Kushki *et al.*, 2015] Azadeh Kushki, Ajmal Khan, Jessica Brian, and Evdokia Anagnostou. A Kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder. *IEEE Transactions on Biomedical Engineering*, 62(3):990–1000, 2015.
- [Kuypers, 2013] Leah Kuypers. The zones of regulation: A framework to foster self-regulation. *Sensory Integration Special Interest Section Quarterly*, 36(4):1–4, 2013.
- [Liu *et al.*, 2008] Changchun Liu, Karla Conn, Nilanjan Sarkar, and Wendy Stone. Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. *International Journal of Human Computer Studies*, 2008.
- [Ma *et al.*, 2019] Tengmeng Ma, Hasti Sharifi, and Debaleena Chattopadhyay. Virtual humans in health-related interventions: A meta-analysis. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, page LBW1717. ACM, 2019.
- [Marinoiu *et al.*, 2018] Elisabeta Marinoiu, Mihai Zanfir, Vlad Olaru, and Cristian Sminchisescu. 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2158–2167, 2018.
- [Ramey *et al.*, 2019] Devon Ramey, Olive Healy, Russell Lang, Laura Gormley, and Nathan Pullen. Mood as a dependent variable in behavioral interventions for individuals with asd: a systematic review. *Review Journal of Autism and Developmental Disorders*, pages 1–19, 2019.
- [Ringeval *et al.*, 2013] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*, (i), 2013.
- [Rudovic *et al.*, 2018] Ognjen Rudovic, Jaeryoung Lee, Miles Dai, Björn Schuller, and Rosalind W Picard. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, 3(19), 2018.
- [Saadatzi *et al.*, 2013] Mohammad Nasser Saadatzi, Karla Conn Welch, Robert Pennington, and James Graham. Towards an affective computing feedback system to benefit underserved individuals: an example teaching social media skills. In *International Conference on Universal Access in Human-Computer Interaction*, pages 504–513. Springer, 2013.
- [Samad *et al.*, 2018] Manar D. Samad, Norou DIawara, Jonna L. Bobzien, John W. Harrington, Megan A. Witherow, and Khan M. Iftekharruddin. A Feasibility Study of Autism Behavioral Markers in Spontaneous Facial, Visual, and Hand Movement Response Data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(2):353–361, 2018.
- [Sarabadani *et al.*, 2018] Sarah Sarabadani, Larissa Christina Schudlo, Ali-Akbar Samadani, and Azadeh Kushki. Physiological detection of affective states in children with autism spectrum disorder. *IEEE Transactions on Affective Computing*, 2018.
- [Scarpa and Reyes, 2011] Angela Scarpa and Nuri M Reyes. Improving emotion regulation with cbt in young chil-

dren with high functioning autism spectrum disorders: A pilot study. *Behavioural and cognitive psychotherapy*, 39(4):495–500, 2011.

[Sharmin *et al.*, 2018] Moushumi Sharmin, Md Monsur Hossain, Abir Saha, Maitraye Das, Margot Maxwell, and Shameem Ahmed. From research to practice: Informing the design of autism support smart technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 102. ACM, 2018.

[Trevisan *et al.*, 2018] Dominic A Trevisan, Maureen Hoskyn, and Elina Birmingham. Facial expression production in autism: A meta-analysis. *Autism Research*, 11(12):1586–1601, 2018.