Towards a taxonomy of quantitative emotion measurement using wearables for inclusive user experience research

Marc Salvador Fuentes Bongenaar

Bachelor Graduation Project

Creative Technology

Faculty of EEMCS

Supervisor: Dr. Champika M. Epa Ranasinghe

Critical observer: Dr. Andreas Kamilaris

September 2021-February 2022

Abstract

User experience research is important to improve the quality and usability of products and applications. The current methods of measuring emotions are mostly subjective, qualitative, and use techniques that are not inclusive to all potential user groups. Quantitative emotion measurement using wearables can offer new possibilities for user experience research because it is inclusive, objective, and does not require cognitive input from the user. This paper presents a taxonomy of the emotion measurement domain which was created with a systematic literature review. 38 papers about emotion measurement were reviewed to increase understanding of what sensors are used, what emotions can be measured, how the data is handled, what accuracies can be achieved, and if the user group influences emotion measurement. The resulting taxonomy can provide other researchers an overview of the current state of emotion measurement and help in determining what sensor and emotion combinations fit their needs to create their own inclusive emotion measurement system.

Acknowledgement

Firstly, I would like to thank my supervisors Champika and Andreas for the discussion we had, their suggestions, and guidance throughout the process of creating my taxonomy. I would also like to thank my friends and family for supporting and encouraging me, and always being open to discuss my graduation project.

Table of contents

List of figures

List of Tables

Chapter 1 - Introduction

Human Computer Interaction (HCI) research has become an important part of the product design cycle, HCI improves the User Experience (UX) which ultimately dictates the user satisfaction [1]. The goal of UX research is to interpret or measure experiential qualities that a product evokes in a user such as happiness, sadness, or disgust [2]. Knowing how a product or prototype makes a user feel is meaningful information that can provide guidance in the design process of a product. HCI, the field UX is rooted in, has had a historical focus on improving the quality of life of the user [3]. This can be expressed in many ways: “What menu design is best suited to the task?” or “Does the application suit the user’s needs?” [3]. A large focus is put on improving the efficiency and speed of the user using the new interface or product. While this may improve ones quality of life it might not be a good experience for the user. An example of this is the DVORAK versus QWERTY keyboard debate. Using the DVORAK can improve typing speeds compared to QWERTY by up to 74% but less than 1% of typists actually use the keyboard [4]. If a laptop manufacturer creating a laptop geared towards office workers only looks at aspects improving the efficiency of the user, they can implement the DVORAK keyboard. That manufacturer will most likely be disappointed by the sales of the laptop, even though their laptop has the potential to be more efficient, the potential users do not like the experience of using the laptop. This is why UX research is important, and why it is important to look at many aspects regarding a product instead of focusing on one alone. To interpret or measure the experiential qualities evoked by a product, several methods exist.

Subjective methods to determine UX are most popular in UX research hitherto [5]. The most commonly used methods in user experience evaluation are questionnaires, interviews, and observations [2][5]. The foremost evaluated aspects of UX are generic user experience, aesthetics, and pragmatic and hedonic quality [5]. The quantitative measurement of evaluation of emotion, stress, engagement, or frustration are less prevalent. These aspects are an integral part of the user experience and are part of the short-term affective response (STAR) and are a debated topic within the HCI community on whether they can be measured or not [2]. If these aspects are included in a user experience evaluation, they are often done through the common methods of questionnaires, interviews and observations. While these methods offer insight in the user experience, they are not objective and are susceptible to user errors, and not all target groups are able to self-report and observations can be misguided [6] [7]. An objective emotion measurement method can offer insight into the user experience without suffering from the setbacks existing in self-reporting methods. Further, the invasiveness of certain objective and subjective measurement techniques can influence the results of the measurement itself [8]. Because of this, the measurement method should be easy to use and small, portable.

A way to achieve wearable emotion measurement is using wearable sensors. To create an emotion measurement system using wearables some decisions about sensors and emotions have to be made. The goal of this paper is to provide a clear overview of wearable emotion measurement and to help guide UX researchers into finding sensors and emotions that apply to their use case. The following research question was determined:

1 - What are the possibilities of quantitative emotion measurements using wearable physiological sensors for inclusive UX research?

1.1 Which sensors are used?

1.2 Which emotions can be measured?

1.3 How can we guide a UX researcher in the selection of sensors and emotions?

This taxonomy about quantitative user experience can be a worthwhile research topic and a methodological and theoretical contribution to the field of Human Computer Interaction research [9].

Chapter 2 - Background and Related Works

In this chapter the current state of UX measurement will be assessed. Then, broader emotion measurement methods and standards will be presented, whereafter similar works to this taxonomy will be identified.

2.1 - User Experience measurement

The topic of UX measurement is a common area of discussion within the HCI community; this is highlighted in the research conducted by Law, van Schaik, and Roto regarding the attitude of the HCI community towards UX measurement [2]. In their research they conducted surveys and interviews with many UX researchers from many different fields. There is a divide within the community about the plausibility and usability of UX measurements. This divide exists because the multidisciplinary nature of the community. The methods within these fields value different aspects more than others, and this is reflected within the HCI community. This attitude is well demonstrated by a quote from Roto (2010): "No generally accepted overall measure of UX exists, but UX can be made assessable in many different ways." (p. 8) [11]. A survey conducted by Law and her colleagues showed that STAR EQ's are viewed by many in the HCI community as something that can be potentially measured which can be seen in figure 2.1 [2]. Therefore, objective emotion measurement using wearables could be a useful measurement tool for UX research.

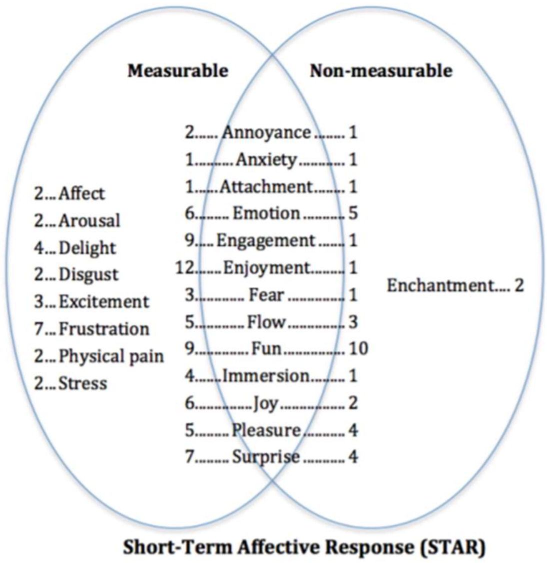


Figure 2.1: Measurability of STAR attributes, [2]

2.2 - Self reporting

Self-reporting methods in user experience research can be influenced by many factors and are not possible in some instances. The answers given to a questionnaire regarding the emotional experience of using a product might not be fully truthful, the user might be inclined to answer more positively or negatively because of the presence of the researcher. It can also be the case that the user cannot fully recall their emotions at specific moments during the user test and therefore cannot answer all questions in the questionnaire [6] [7]. Furthermore, not all potential users are able to self-report. For example, during a user test for a product geared towards people with Alzheimers, Down-syndrome, or a form of autism, self-reporting cannot always be relied on [12]. The solution of some researchers in this case has been to ask caretakers to share their observations with the researchers, but this is not objective and suffers from the same problems present with users that are able to self-report [12]. Similarly, user tests involving children cannot always rely on self-reporting. While parents or guardians can offer their observations, this method still has the same problems present in user self-reporting or caretaker observations [12][13]. Objective emotional measurement has been used in some user experience evaluations although the methods vary, and the usage is still sporadic [5]. A concrete subjective emotion measurement method can offer additional insight into the user experience and experiential qualities while not being affected by the same issues as self-reporting.

2.3 - Invasiveness

During user evaluation, the invasiveness of the measurement method can influence the experience of the user. This is already integrated in many research methods for UX research; in a user test with multiple groups, it is usually standard to conduct the user tests in the same environment [8][12]. This shows that although the environment does not directly affect the product/prototype that is being tested, it has enough impact to be a factor that has to be controlled. The same is true for the invasiveness of the measurement tool used for research methods. A person will act differently on camera than they do without being recorded. The same is true for more “extreme” examples, like electroencephalography (EEG) for emotion measurement. When the user has to go through a long set up stage to install an EEG sensor helmet before they can go through the user test, the presence of the helmet on their head can influence results. Physio-psychological and biosignal measurements are the most commonly used measurements for emotion measurement. The measurements are almost impossible to be done through a non-invasive method, but by making the measurement tool as non-invasive as possible the impact of the measurement tool on the user can be decreased. This is already possible for many biosignals because the sensors are small enough to be placed in a small device on the body. Large scale implementation of this is already present in our society through the form of smartwatches. A wearable emotion measurement device can offer an objective measurement method with minimal invasiveness which can provide new insights in user experience research.

2.4 - Present methods

Up until now most of the user experience evaluations (which included emotional experience) were done through subjective research methods [5]. In general, the methods used for general UX research varied, but as can be seen in table 2.1 only the second to last item, using physio-physical signals in the method, can be seen as truly objective. This table is referenced from research on the usage of triangulation in UX research methods. Triangulation of research methods is the practice of combining multiple research methods to reach a conclusion instead of just one. According to Pettersson, triangulation of research methods is used in more than two thirds of the 100 studies they reviewed. Regardless, table 1 still shows that even if multiple research methods were used, they were most likely both subjective method types that utilize self-reporting. According to [5] a questionnaire is the most commonly used method for evaluating emotional experience during a user test. [2] also proposes a general method for emotion evaluation for UX research by establishing base emotions for emotional questionnaire items. On a broader scale emotion measurement is used in other fields than UX but while objective methods were used, the models used and the factors measured varied.

Method Type**	%*
Self-Developed Questionnaire	53
Semi-Structured Interviews	46
Activity Logging	31
Standardized Questionnaires	26
Live User Observation	19
Videorecording	16
Free Interview	9
Think Aloud Feedback	6
Diaries	6
Focus Groups	5
Online Feedback	3
Probes	3
Physio-psychological	2
Others	5

Table 2.1: Usage of methods in UX research, [5]

2.5 - Emotional models

Within the studies focusing on emotion measurement there are two ways the measurements are translated into usable data for the research. One of these is the discrete emotional model. This model contains specific emotions and/or experiences that are measured. This could be a binary model which only checks if a user is happy or not happy, but it can also distinguish between 5 or more determined emotions. The second model is the dimensional model. These models do not name a specific emotion but instead plot the measurement on a graph [14][15][16]. The most common axes for the graph are valence and arousal. Valence is the measured positivity or negativity of the emotion; arousal is the measured intensity of the emotion which is being experienced [14][16]. Together they form a graph which can be seen in figure 2.2 [17]. In discrete emotional models valence and arousal are also often used, they then get assigned a discrete emotional state to the measured value. Not only emotions in the traditional sense of the word are measured, stress, tiredness, and engagement experienced by the test subject can be measured using similar methods to that of valence and arousal. Interestingly, engagement is a specific measurement that is disproportionately conducted more through objective methods using biosignals, facial recognition, movement, etc. as compared to other EQ's, such as emotions. The reason for this can be that engagement is a universal design goal for most products, and therefore a greater interest and focus on this topic exists [18]. The measurement of those additional non-emotional factors can be seen as adhering to the discrete model because they only focus on one aspect. Both discrete and dimensional emotional models are valid, the areas where they can be applied best need to be determined.

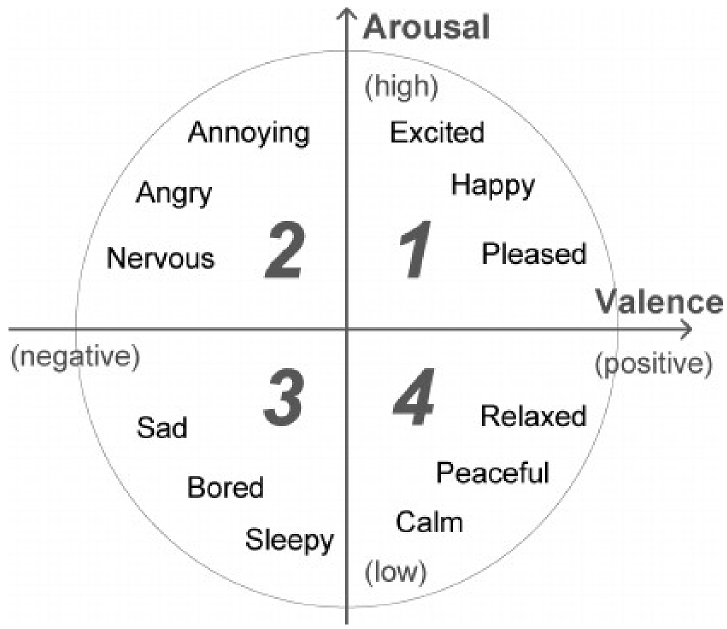


Figure 2.2: Showcase of the emotional models [17]

2.6 - Emotion evocation and datasets

To calibrate and determine the accuracy of the emotion measurement system, emotion evocation methods can be used. These datasets can also be used to train algorithms for emotion measurement. These evocation methods consist of short videos, photos, or sounds that evoke a certain emotion to the person perceiving them [19][20]. Several emotion evocation databases exist, each with different focuses. One of the databases is geared towards HCI research, the MAHNOB-HCI database. The MAHNOB database contains a laughter database and a HCI tagging database. The tagging database includes videos of spontaneously elicited emotions filmed from multiple angles. It also includes EEG measurements synchronized to the video as well as eye tracking data [20].

2.7 - Existing techniques

Wearable sensor devices which can be used for emotion measurement already exist, although there are no devices specifically made for UX research. Within HCI research objective emotion measurement is used in a few instances. There is one tool called "Vempathy", which is made for UX research that uses artificial intelligence to detect emotions from facial expressions [21]. Only one case study using this tool can be found which states that the tool is useful but is best used with triangulation of methods, a common practice in UX research [22][5]. Emotion measurement using facial recognition is a legitimate way to measure emotions, but it cannot always be applicable for the user group [24]. In some experiments a constant camera stream of the user might not be possible and in some cases the users might not be able to perform facial expression due to pre-existing conditions. It can also be debated that this method is still somewhat subjective because the subject can choose to hide or alter their facial expression. The same is true for other observation methods such as speech recognition and behavior observation [24][25]. These aspects are often observed during an interview or general observation session but in these cases the subject is still able to alter their speech patterns or behaviors which can lead to skewed results. Physiological signals are more subjective in this regard because of the versatility of options. Multiple factors can be measured, most of which are impossible or very hard to control. Therefore, this can offer a more subjective manner of emotion measurement [26].

Within physio-psychological methods, specifically electroencephalography (EEG) is most popular in UX according to several review papers [5]. This method can provide many useful insights into the user experience because it directly measures the brain activity of the user. One setback of this method is the invasiveness of the EEG measurement tool, which can influence the experience of the user as can be seen in figure 2.3 [27]. Outside of UX attempts at wearable emotion measurement have been made. Many different methods and tools have been used with varying goals and performances. The papers detailing these methods and tools will form the basis for the research for the taxonomy.

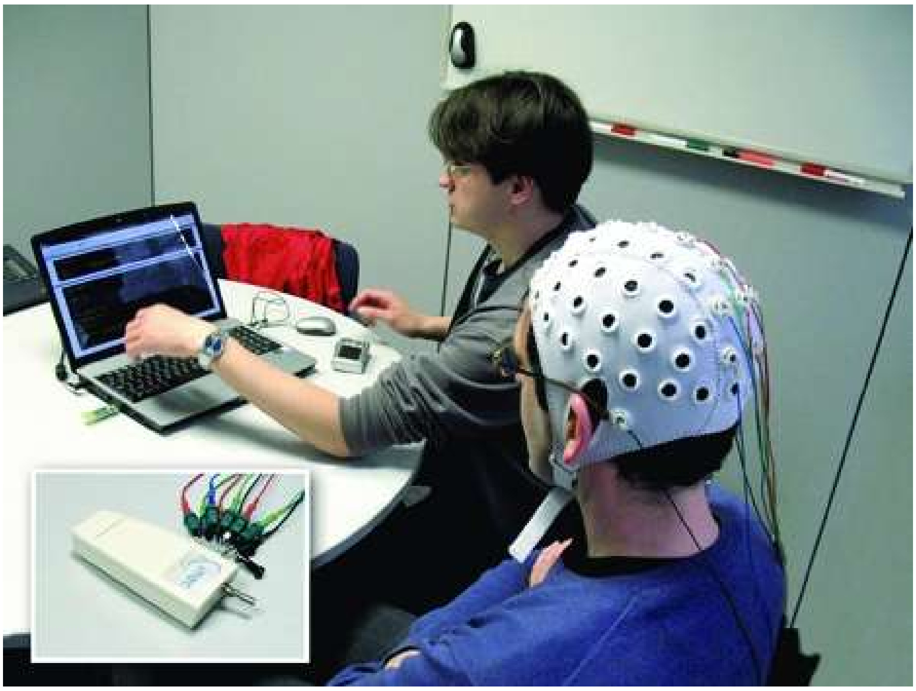


Figure 2.3: EEG setup [27]

2.8 - Existing literature

There exist multiple literature reviews focusing on the topic of emotion measurement for HCI. A paper by J. Zhang and her colleagues discusses the usability of emotion measurement for HCI and affective computing and provide several examples of emotion measurement using physiological signals [15]. The paper further goes in depth about EEG and techniques of using EEG for emotion measurement. While the paper does include a short review of existing techniques it mostly focuses on EEG. This is in contrast to the review by S. Saganowski et al. who performed a concise systematic literature review of wearable emotion measurement resulting in 35 reviewed papers. They provide numerous recommendations for further research within the topic regarding everyday emotion measurement and the data processing methods. One limitation of their research is regarding the exclusion criteria of excluding papers that only consider one emotion [14]. In [28] facial emotion measurement in UX evaluation is reviewed. Only a small number of papers (14) were reviewed. They conclude that the practice is still quite novel for UX evaluation specifically and stress the importance of standardization of tools and techniques to make facial emotion detection more accessible. A review about emotion measurement through EEG used 285 articles and concluded that emotion measurement through EEG is a growing research topic and that recent research is moving towards wireless application of EEG emotion measurement. Interestingly, most research on EEG brain computer interfaces recently has been for application in virtual reality apps and games [29]. Lastly, Landowska reviewed the usage of emotion measurement for improvement of IT usability evaluation. The paper does not go in depth about the topics and only includes one paper regarding the usage of physiological signals for emotion measurement. They conclude that emotion measurement using physiological signals provides high accuracy, is medium to highly robust against outside disturbances, is highly independent against human will, and has a medium to high possibility of interfering with usability testing procedures [30]. While reviews on the topic of (wearable) emotion measurement have been done, sometimes with a focus on HCI and UX, none of them provide a fully comprehensive review of all aspects regarding wearable emotion measurement for UX evaluation. Most of the review papers call for standardization of techniques and further research of the topic. Further, in a Scopus library search about wearable emotion measurement 19 papers from 2021 can be found. This is information that is missing from the other review papers because they were all written in 2020 or earlier. This is why a deductive taxonomy of wearable emotion measurement can fill a gap in HCI research that has not been previously addressed.

Chapter 3 - Methods and Techniques

The goal of this paper is to provide a foundation for inclusive, quantitative emotion measurement to further the possibilities in UX research. As discussed in the background research there exist several papers and some reviews regarding emotion measurement. While some papers aimed towards UX research could be found, these were limited in scope, and none of the reviews discussed inclusivity of the methods reviewed. This leaves this paper with some possibilities; a paper could be made discussing the creation of an inclusive emotion measurement for UX research, or a review or classification of the topic in its current state can be created to guide other researchers in the creation of their systems. This paper denotes the latter, a taxonomy, or classification of a topic hitherto, which can be a tool that provides the foundation for based future research. While creating a prototype and testing it can provide a valuable insight into what is possible, doing so while not aware of all of the possibilities to do so limits the research unnecessarily. To create this taxonomy an adaptation of the method described by Nickerson was used [31].

3.1 - Method

A deductive approach using a systematic literature review was used. The method of taxonomy development proposed by Nickerson was chosen, more specifically, the approach seen in figure 3.1.

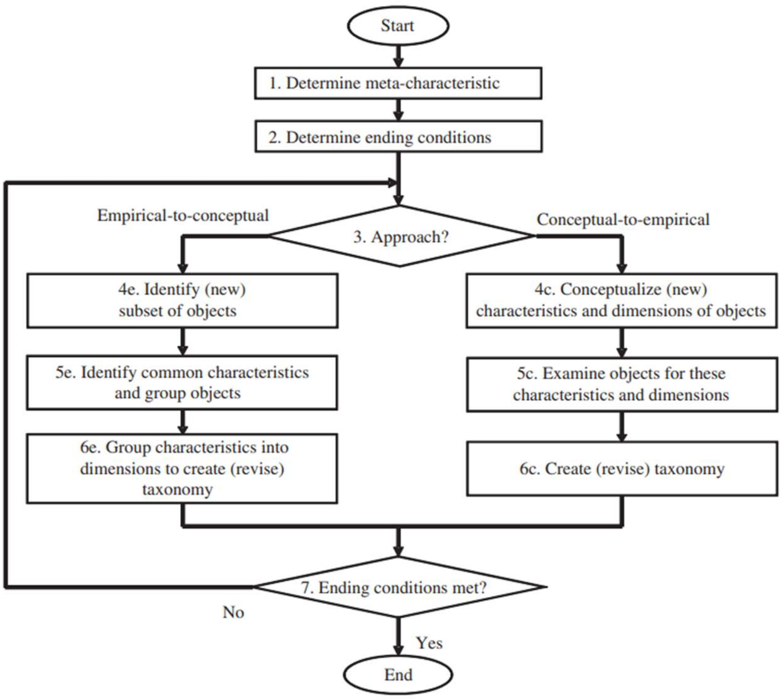


Figure 3.1: Nickerson approach [31]

Nickersons method was chosen because it offered flexibility through the iterative approach although it was altered to fit this project. Because of the iteration more characteristics and dimensions can be added during the process which can offer valuable insights which otherwise could have been missed. The Nickerson method utilizes dimensions, characteristics, and meta-characteristics. Firstly, the dimension is the general topic of the taxonomy. For this taxonomy, we will utilize one dimension: emotion measurement using wearable sensors. The meta-characteristics are the broad topics that need to be researched, and the characteristics are smaller parts within this larger meta-characteristic. The Nickerson method utilizes iteration during which the (meta-) characteristics can be expanded upon.

Starting out sensor types, emotional model, accuracy, and algorithm used were chosen as (meta-) characteristics. These characteristics were chosen because of the reviews read in the background research [29] [28] [32]. During the iterations more characteristics were identified. Two deviations from the Nickerson method were made. Firstly, the possibility of multimodality in the specific characteristics. This was done because some papers used varying methods which fell into multiple characteristics within the meta-characteristics [33]. Secondly, the iteration steps proposed by Nickerson were not followed. These iteration steps can be very useful when a large number of papers need to be reviewed, the scope of this project allows for a less restrictive approach. The aspect of the method that applies well to this research is the option to iterate through (meta-)characteristics. This aspect was kept in a different way; if another characteristic was identified during the systematic literature review it was included in the list of characteristics and all previous papers were examined again for this characteristic. This is not a very efficient method, but because of the relatively small pool of papers it was possible.

3.2 - Data-acquisition

The data to create the taxonomy was acquired with a systematic literature review. The decision of library, keywords/search query, and inclusion/exclusion criteria was dependent on a few factors:

- The library needs to include both technical and health science papers to reflect the emotion measurement field accurately
- The number of papers to review need to be large enough to draw valid conclusions from, but small enough to fit the scope of the graduation semester

The Scopus library was chosen as the place to obtain the articles. Scopus was chosen because of its affiliation with the health sciences and the biomedical field [34]. The keywords to create the search query consist of four categories: emotion, measurement, biometric, wearable.

The first three categories consist of three keywords and the last as a single keyword. The specific keywords were:

- Emotion, stress, feeling
- Measurement, recognition, detection
- Biosensing, sensor, biometrics
- Wearable

See figure 3.2 for clarification.

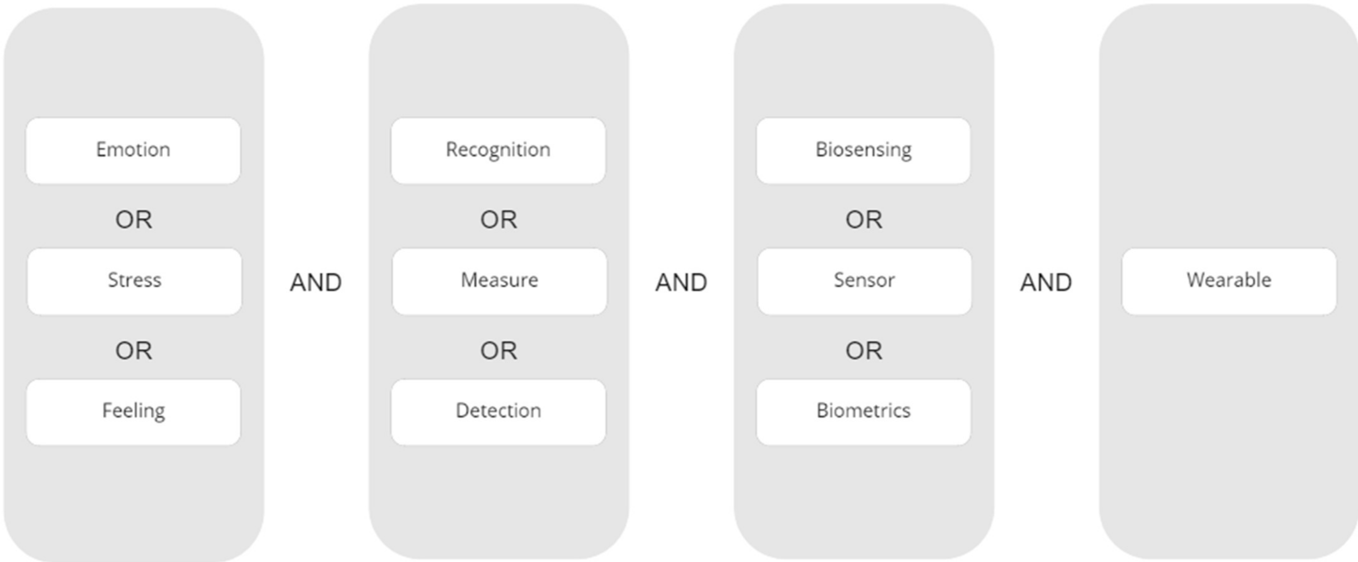


Figure 3.2: Search query

Lastly, the inclusion and exclusion criteria need to be determined. For the publication year, all relevant publications after 2018 were included in the list of articles. Further, reviews and full conference papers were excluded and lastly papers which were not relevant were excluded during the review process. The final search queries were:

- TITLE-ABS(Emotion OR Stress OR Feeling AND Recognition OR Detection Or Measure AND Biosensing OR Sensor OR Biometrics AND Wearable) PUBYEAR > 2018

- TITLE-ABS(Emotion recognition OR Stress recognition OR Feeling recognition AND Biosensing OR Sensor OR Biometrics AND Wearable) PUBYEAR > 2018

- TITLE-ABS(Emotion measure OR Stress measure OR Feeling measure AND Biosensing OR Sensor OR Biometrics AND Wearable) PUBYEAR > 2018

- TITLE-ABS(Emotion detection OR Stress detection OR Feeling detection AND Biosensing OR Sensor OR Biometrics AND Wearable) PUBYEAR > 2018

This search yielded 83 papers, of which 45 were excluded because they did not pass the inclusion/exclusion criteria. The list of all the papers can be found in appendix A.

Chapter 4 - Taxonomy

This chapter presents the analysis of the literature, the choice and explanation of all (meta-)characteristics, observations, and the resulting taxonomy.

4.1 - Literature analysis

As stated, the initial (meta-)characteristics were sensor types, emotional model, accuracy, and algorithm used. These were an adequate set of initial characteristics, but during the process of reading the selected literature 6 more (meta-)characteristics emerged to create the final list of 10. All included papers were evaluated for these characteristics. As stated earlier, the decision was made to include multimodality within meta characteristics because some papers fell into multiple characteristics. The (meta-)characteristics:

1. Sensor types
   The type(s) of sensors used in the article. Everything which was measured in a quantitative way was included here. Notably, observations and activity logging were excluded. Acoustics was included on a case-by-case basis, were automated emotion measurement through audio analysis was included but manual detection was excluded.
2. Sensor fusion
   This was a binary characteristic. Either sensor fusion was applied or not. This characteristic was added because in the future it can make finding papers detailing it easier to find for other researchers. Not all papers utilizing multiple sensors used sensor fusion. Sensor fusion was attributed to measurement methods that combined the data from multiple sensors to derive the emotion measurement.
3. Emotional model
   As discussed in the background research, either a dimensional or discrete emotional model is used in an emotion measurement method. In some cases, both were used.
4. Emotional categories
   This characteristic identifies the dimensions (either valence, arousal, or both) and the specific emotions that were measured. These were recorded on a Miro board by placing the articles on their position according to the proposed wheel of emotions by Plutchik [35][36]. The wheel of emotions can be found in figure 4.1.
5. Quadrant/octant
   The proposed 'wheel of emotions' by Plutchik includes quadrants/octants [35][36]. The quadrant/octant each paper measured was identified.

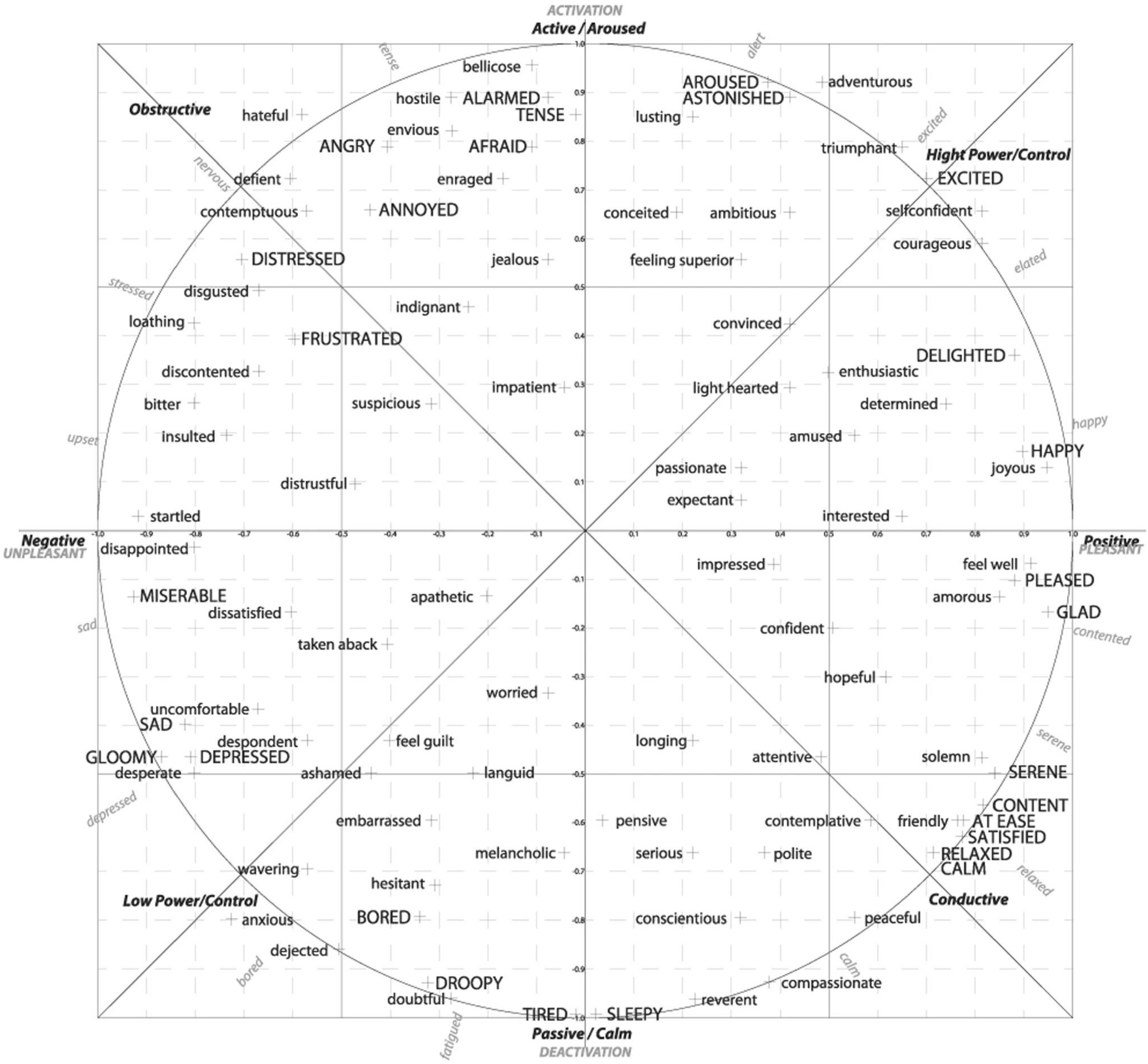


Figure 4.1: Wheel of Emotions [35][36]

6. Algorithm

   Emotion measurement can be done through various algorithms, the specific algorithm or method of detection was recorded. Some papers did not specify what was used.

7. Valence accuracy

   The reported accuracy of valence measurement, if applicable.

8. Arousal accuracy

   The reported accuracy of arousal measurement, if applicable.

9. Accuracy

   The reported accuracy of specific emotion measurement in case the paper adhered to the discrete emotional model.

10. Participants

    Number of participants in the study, unique characteristics like gender, age, and ethnic background was included if reported.

An overview of all the (meta-)characteristics can be found in table 4.1.

	Sensor		Emotional model				Accuracy			
Emotion measurement	Sensors used	Sensor fusion	Emotional model used	Emotional categories	Quadrant/octant	Algorithm used	Valence acc.	Arousal acc.	Discrete acc.	Participants

Table 4.1: Taxonomy dimension, meta characteristics, and characteristics

The full list of articles with their identified (meta-)characteristics can be found in appendix B. The article names and outlets were removed from the table to make it easier to read. The numbers in the table correspond to the article list in appendix A. After characterizing all papers some trends could be identified, these will be presented by meta-characteristics.

4.2 - Sensors

In total 23 different sensors were used. Some sensors were referred to using a different name but referred to the same sensor. This was the case of GSR (galvanic skin response) and EDA (electrodermal activity), these two were grouped together. Generally, the specific name for the type of sensor used was recorded, not the exact measurement done using the sensors. This was not the case in the instance of PPG (photoplethysmogram) and BVP (blood volume pulse). BVP is measured using a PPG sensor, but a PPG can provide a wide variety of other measurements too [37]. In the cases where the article specifically named using only BVP from the PPG sensor, BVP was recorded. If the article specifically named BVP and PPG (alternate measurements), both were recorded. This approach was chosen because many articles utilized and named BVP specifically. Furthermore, sensor fusion was recorded as "applicable" to the included papers. Some papers reported different accuracies dependent on early/late state fusion or included sensor specific accuracy too, this was recorded in the accuracy meta-characteristic. The influence of sensor fusion was not investigated, although it is possible to make some preliminary conclusions with the data recorded in appendix B.

Table 4.2 shows a list of all sensors that were used, the papers that used them and basic information about the sensor.

Sensor/measure name	Information	Use in paper #1
PPG/photoplethysmogram	The PPG sensor uses the absorption of infrared light in tissue, more specifically, blood vessels, to measure (most commonly) blood volume pulse and heart rate. The sensor is usually placed on the fingertips, although it can also be used on the wrist[37].	3, 4, 7, 10, 24, 25, 37, 42, 45, 60 ,62
EEG/electroencephalogram	The EEG sensor can measure brain waves using nodes attached to the head. While commonly used with a large number of nodes, it can be used with a small number for specific brainwaves. Can offer a detailed response to stimuli but is quite intrusive [38] [39].	2, 3, 13, 38, 45, 47, 52, 82
fNIRS/functional near infrared spectroscopy	A technique to measure hemoglobin levels in the brain using infrared, similarly to PPG. Can be used to identify cortical responses to stimuli quite accurately [40].	58
Eyebrow position	The measurement of the eyebrows through imaging. Is useful for expressive emotions such as disgust [41].	3
Pupillometry	Measures the pupil's response to stimuli. Pupil response has been linked to liking and concentration [42].	3, 17, 65
Face muscle response	The response of the face muscles can be	3

	used to identify facial expression. Generally done through stretchable sensors [41].	
Saccadic eye response	The saccadic movement of the eyes (simultaneous movement) can be used to identify points of focus and engagement. Useful for measuring response to video [43].	17
FEA/facial expression analysis	FEA is done through processing video with the computer. It is useful for expressive emotions [44].	10, 52, 75
FTT/fingertip temperature	Temperature response in the fingertips. Used in one instance.	1
BP/blood pressure	Blood pressure can be measured using a blood pressure sensor. It is related to heartbeat and commonly used in conjunction with other cardiac or blood related measurements.	41, 54
SP20/oximeter	SP20 can be used to measure the oxygen saturation in the blood.	41
BVP/blood volume pulse	A commonly used measurement acquired with a PPG sensor. Measures the change in blood volume moving through veins.	10, 17, 47, 72, 77
Glucose	Glucose measurement of the blood. Glucose saturation is linked to diet, so it can offer an interesting link between differentiating emotional response and diet [45].	41, 54
HR/heart rate	Measurement of the heart rate. Can be	17, 44, 53, 54, 71, 73

	done through many different ways.	
ECG/EKG//electrocardiogram	The electrocardiogram measures the electrical activity of the heart. Heart rate can be derived from an ECG.	8, 17, 19, 24, 41, 45, 49, 72, 74, 81
RESP/respiratory response	RESP is the respiratory response to stimuli.	17, 42, 53, 72
TEMP/temperature	The temperature of the skin can change slightly during emotional response [46].	8, 17, 37, 41, 53, 71, 72, 73
GSR/galvanic skin response	GSR is the most commonly used sensor for emotion measurement. GSR measures the electrical skin response of the body. GSR seems to be fairly accurate in the arousal dimension and mostly used for that purpose. Generally measured on the wrist of fingers [47].	8, 10, 17, 25, 30, 37, 41, 49 52, 53, 54, 58, 60, 65, 71, 72, 73, 74, 77, 81
EMG/electromyography	EMG can be measured with small electrodes to measure contraction of specific muscles depending on placement.	17, 44, 81
RFID/radio frequency identification	RFID is used for experimental emotion measurement in one instance. RFID can be used for contactless monitoring of several biosignals [48].	22
IMU/inertial measurement unit	An IMU can be used to measure movement of the user. Used in one instance in combination with other sensors. Can contribute additional situational movement data to emotion measurements.	40
Typing	The analysis of typing behavior to measure emotion.	10
Acoustic	Analysis of sound recordings of the	10, 11

	participant. Analyzed for peaks at certain frequencies or sound patterns.	

Table 4.2: Table of sensors used in reviewed papers

1 - The paper number is referring to the number on the list of articles in appendix A

4.3 - Emotional model

Within the emotion meta-characteristic several aspects were recorded. Firstly, the emotional model used in the paper was noted. This could be either the dimensional emotional model or a discrete emotional model. The dimensional emotional model measures emotions along two axes and plots the measurement as a point on a plane [35] [36]. While there are general conventions for which positions on the plane correspond to which specific emotions, this is not what is being recorded. The emotional plane is made up of the valence and arousal axes. Valence is the positivity of the emotion; arousal is the intensity of the emotion [35]. The discrete emotional model measures discrete emotions such as joy, sadness, or anger. The list of which emotions are measured is chosen by the creator of the method but there are some common discrete emotion choices. These generally align with established ideas in psychology of a list of basic emotions [49] [50]. There are two general approaches to measuring discrete emotions, a translation from the dimensional plane to a discrete emotion can be made or a dataset is created/used to which measured signals are compared.

Secondly, the emotional categories were recorded. In the dimensional emotional model this was arousal, valence, or both. In some cases, like the Spiders+ setup, valence and arousal were measured in stages instead of with numerical measurement [41]. This shows the intent of the dimensional model well, where the important aspect of the measurement is the general location of the measurement in terms of quadrant. The emotional categories for the discrete model were the individual measured emotions. Using Miro, these papers were also placed upon a dimensional/discrete combinational chart (figure 4.1) to visualize the spread of measured emotions. The Miro chart can be found in appendix C.

This visualization effort flows into the last characteristic within this meta-characteristic; the octant/quadrant that the paper measures. The quadrant and octant a paper belongs to was decided according to the proposed wheel of emotion from figure 4.1 [35][36]. In the eventual taxonomy, only the recorded quadrant was used to create figures. In appendix B and C, the octants for each paper can be found.

4.4 - Algorithm used

To achieve accurate emotion measurement algorithms are needed that dissect the incoming sensor signals and analyze them to measure the emotions of the user. In general, two approaches are used. The first approach is seeing what fits best, this is mostly used in the discrete emotional model. These methods measure the incoming signal and assign one emotion from the list of discrete emotions the algorithm was trained to recognize. In some of these cases, there is a “neutral” emotion [44] [51]. A second approach is a system that will measure the incoming data and not report anything until a specific set of signals corresponds to the threshold held by the algorithm for a specific emotion. The data gathered about algorithms was not investigated further in combination with sensors and emotions, but some observations could be made during the literature analysis process.

Machine learning is the most commonly used method of processing data, and it generally provides the most accurate results among the used methods. The other method used was “basic signal analysis” which includes among others correlation-based feature extraction, peak detection, and frequency analysis [52] [53]. The reason why machine learning algorithms (ML) work well is because they can grow over time and learn more from data it acquires. Usually, the ML algorithms were trained using pre-existing datasets like the DEAP dataset [26] [54] [55]. There were several papers that offered valuable research regarding the efficiency and accuracy of algorithms they compared [54] [56].

4.5 - Accuracy

The reported accuracy of the papers varies wildly. Some papers reached a very high accuracy number, up to 96% in one case [47]. The lowest reported accuracy was 34% [57]. The accuracies in between averaged around the 70-80% mark, depending on method, emotional model, etc. Comparing the accuracies between articles is possible but the results it would provide are not very useful, this is because all of the papers utilize different standards and hardware. For example, the 96% accuracy paper measures arousal and uses a GSR sensor to achieve this. The GSR sensor seems to be very accurate for measuring arousal, but it is not very good at measuring valence. The sensor that they use in this paper is the SHIMMER GSR+ unit which costs around 514 euro [58]. The paper that reported 34% accuracy uses ECG and GSR with BiTalino (r)evolution Kit which costs around 149 euro [59].

The accuracy was found to be dependent on two factors: the hardware and the software.The hardware dictates a hard limit to the possible accuracy of a method, a barrier that cannot be broken. Low-cost sensors just do not provide the accuracy a high-end sensor does. This can be aided by good signal filtering and processing, as shown by [60] where the same BiTalino kit is used as in the 34% paper but where a 85% accuracy is achieved; this also showcases the second factor, the impact of the software.

Another observation about the accuracy is the subject dependent and subject independent accuracy. Not all papers reported this, but for the papers that did the subject independent accuracy was always lower [46]. This is something to keep in mind when creating an emotion measurement method. If you know who your participants in a user test are going to be beforehand and you can train a machine learning algorithm using their data, your accuracy is going to be higher.

4.6 - Participants

The literature analysis process shows that the number of participants is generally quite low. The high end of experiment participants is about 30 people with one outlier having 70 participants. A few papers focused on groups that cannot access general qualitative emotion measurement methods. These groups were children and people with ASD [44] [46] [61]. The results shown in those papers are promising, falling in line with the other papers. This showcases the potential of physiological signals for inclusive quantitative emotion measurement.

One other observation that could be made from the participant numbers is the common occurrence of 32 participants. This is because of a simple reason, the DEAP dataset has datasets of 32 people and these papers utilized the dataset in their experiment instead of “real people” [51] [62]. These papers reported accuracies that were in line with the other papers, but the validity of the methods these papers present can be questioned. Because the algorithm is trained using the dataset and then the performance is measured using the same dataset, it is hard to gauge real-world performance.

4.7 - Taxonomy

The literature analysis culminated in the table which can be found in appendix B. This table shows all of the (meta-)characteristics from table 4.1 as well as the publication year and notes. The article name and outlet can be found in appendix A. The notes are especially useful for some of the articles which were in the literature list but were excluded because they were not relevant to the research. A number of these papers reviewed algorithm performance which can be very useful for creating an emotion measurement application.

When we think about the third sub-research question “How can we guide a UX researcher in the selection of sensors and emotions?” the table in appendix B might not apply very well. While it shows a compendium of a large number of papers relevant to selecting sensors and emotions, it takes a long time to grasp all of the data. This is why some data visualizations and derivative tables were created containing the most important data. Firstly, the visualizations and tables will be introduced and motivated thereafter they will be presented on the following pages.

To determine what visualizations and tables can be useful, we have to keep in mind what goal they serve. A UX researcher needs to be able to find out what emotions have been measured before and what sensors have been used to achieve that. From this information, the researcher needs to be able to find the articles relevant to their choice of emotions and/or sensors. When the researcher has the information on what papers are relevant, they can look in appendix B to find any other additional information they might deem useful.

To achieve this, a number of Sankey diagrams have been created using Sankeymatic [63]. Firstly, an all-encompassing diagram which shows all collected data (figure 4.2). This diagram shows the large variations in the field well but is quite cluttered and can be difficult to use. That is why a second Sankey diagram was made to declutter the image somewhat (figure 4.3). This diagram shows the same data but removes any emotion or sensor that was only used in one or two instances. This also makes motivating choices in a measurement design easier because there you are assured of multiple papers discussing the emotion or sensor. An important aspect of the emotion measurement field is the choice of emotional model, and the previous Sankey diagrams combined the data of both. Five additional Sankey diagrams were created to separate these two models. There is one diagram showing all instances of valence and arousal measurement (figure 4.4) and there are four Sankey diagrams showing discrete emotions for all four quadrants which can be found in appendix D.

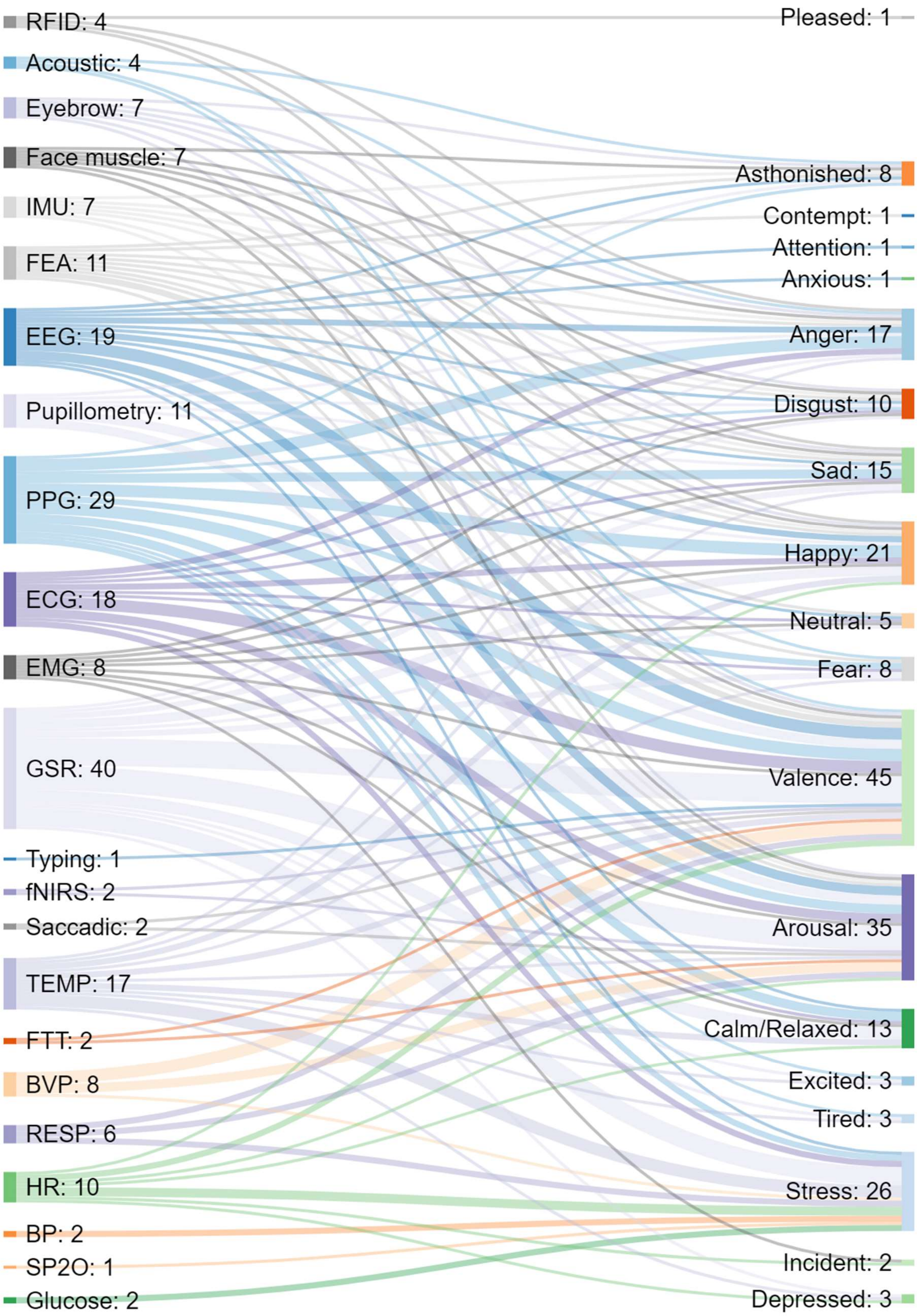


Figure 4.2: Sankey diagram of all emotions and sensors

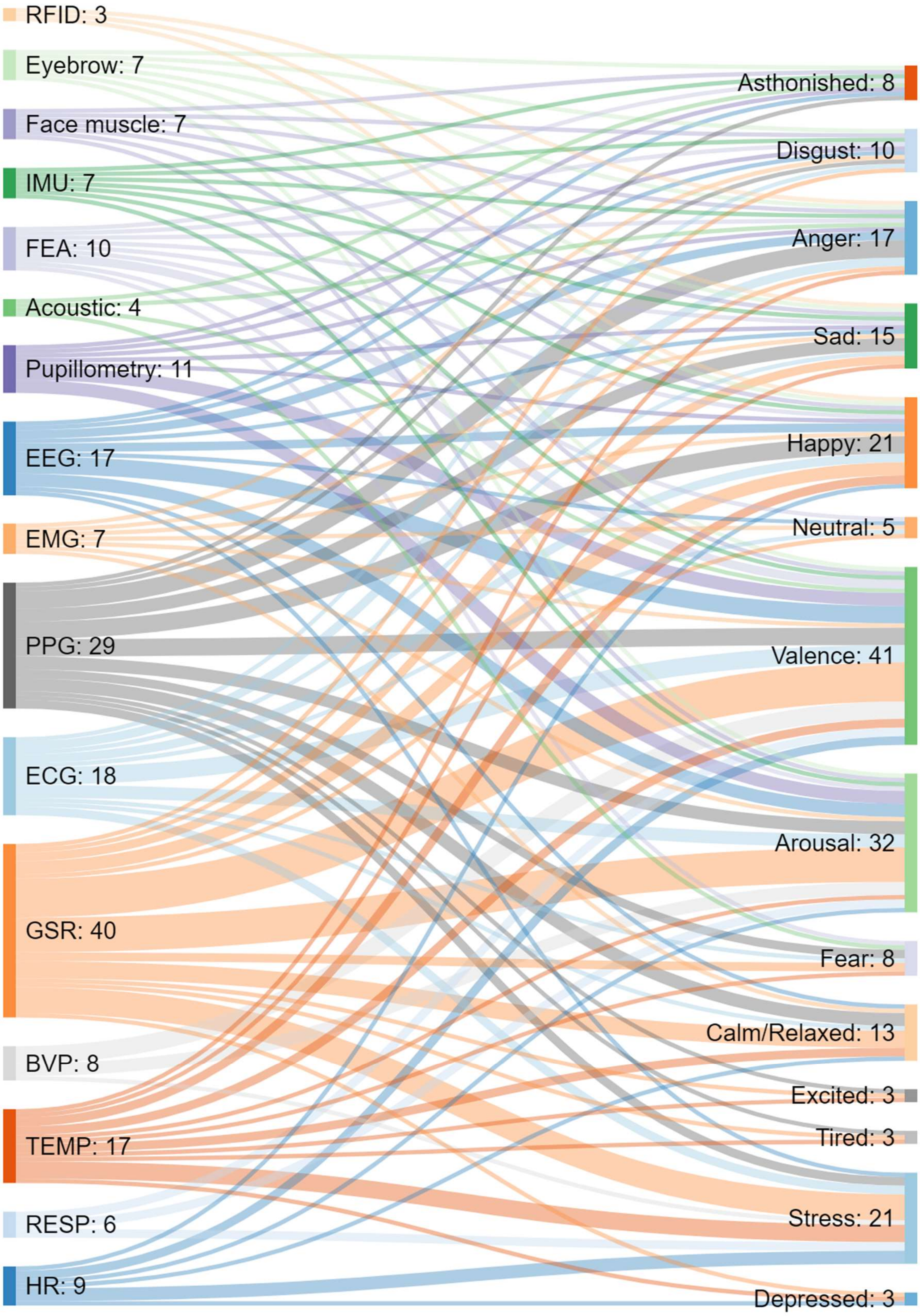


Figure 4.3: Sankey diagram of all emotions and sensors used in two instances or more

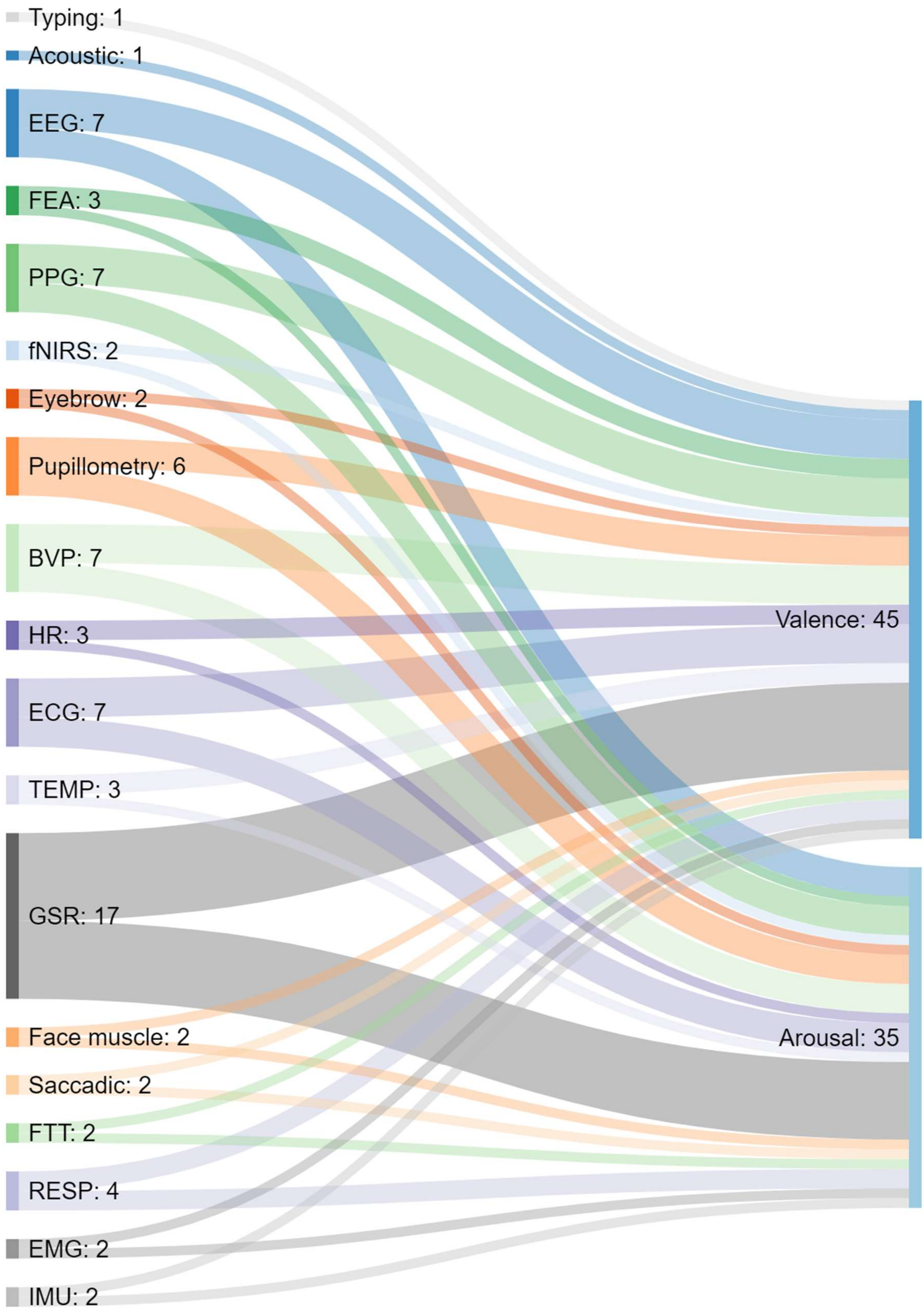


Figure 4.4: Sankey diagram of valence and arousal and sensors used to measure them

A number of tables have been created in addition to the diagrams:

- Table showing the emotions used in each paper
- Table showing the sensors used in each paper
- Table combining the previous two tables
- Table showing the numerical data which was used to create the Sankey diagram

These tables can be found in appendix E.

Lastly, two visualizations showing the popularity of emotions and sensors were created. The popularity of emotions was visualized using a scatter plot where the position of the dots was chosen according to [35]. This plot can be found in figure 4.5. The intensity of the dot shows the popularity of measuring that emotion. The bar chart of all used sensors shows which sensors are most popular and which are the least popular (figure 4.6).

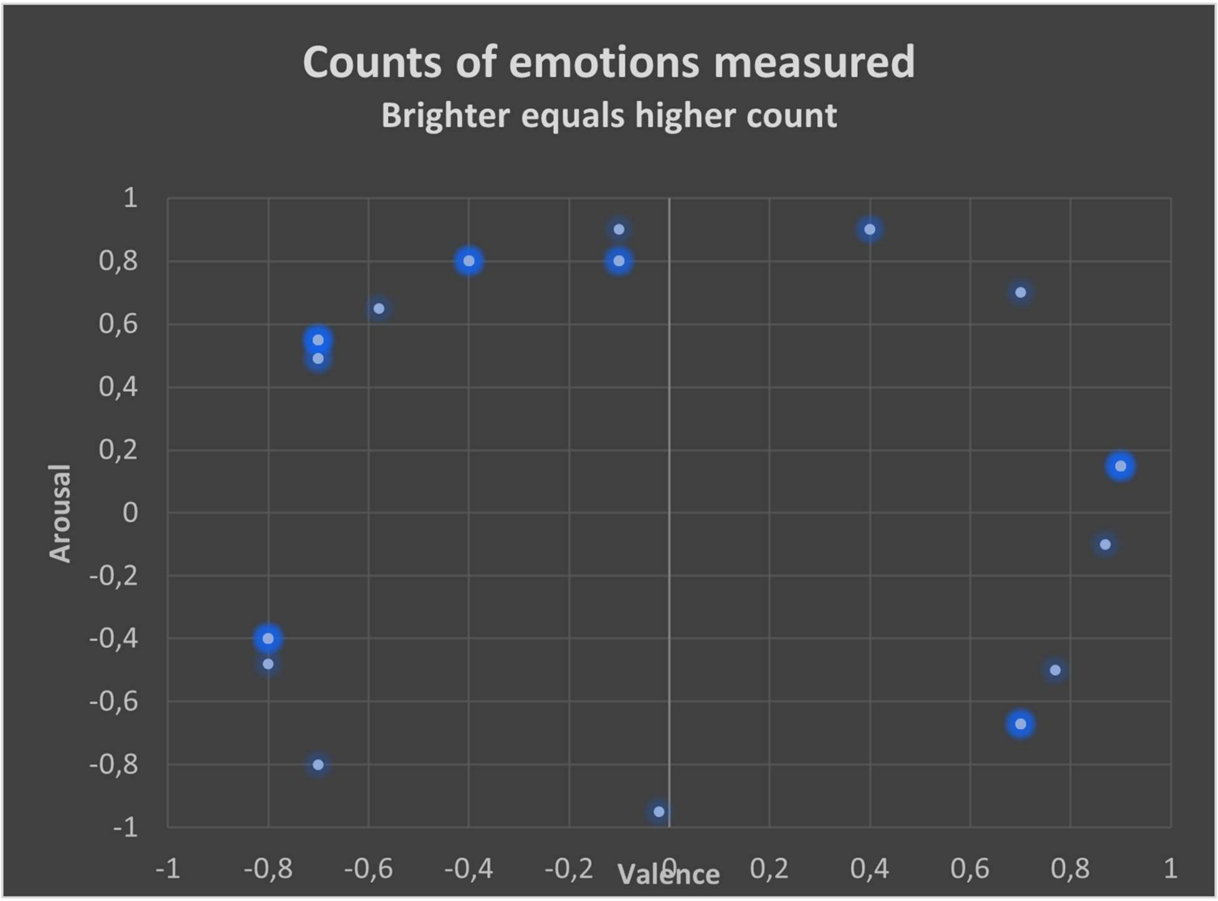


Figure 4.5: Scatterplot of emotions measured

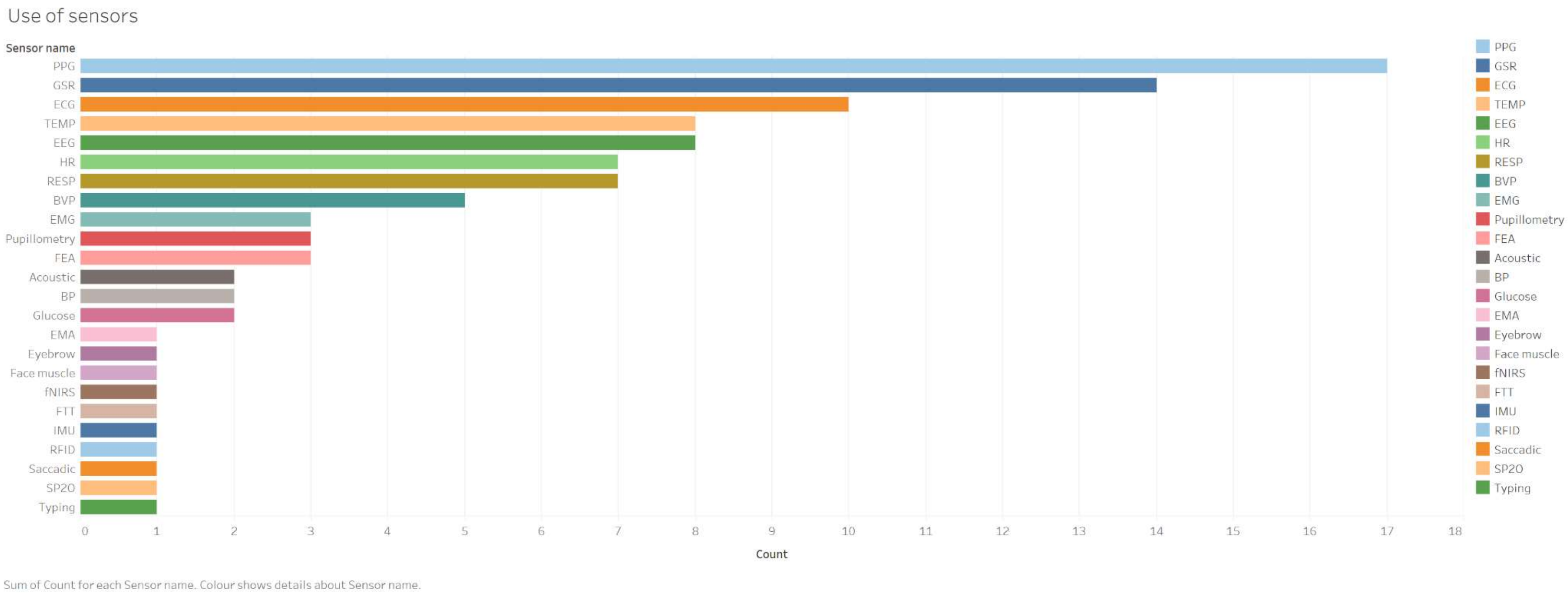


Figure 4.6: Bar chart showing counts of sensor usage

Chapter 5 - Evaluation

There are various methods of evaluating a taxonomy. This taxonomy will be evaluated using two evaluation techniques which can be found in “Because your taxonomy is worth it: Towards a framework for taxonomy evaluation” [64]. Both the illustrative scenario evaluation method and the logical argument evaluation method will be conducted. These methods were chosen because they complement each other well. With the illustrative scenario the theoretical usefulness of the taxonomy can be determined and with the logical argument the validity of the taxonomy and its expansion of existing knowledge can be demonstrated.

5.1 - Illustrative scenario

Szopinski and his colleagues present the illustrative scenario as a way of determining the usefulness of a taxonomy with a real-world or synthetic scenario [64]. If we want to evaluate this taxonomy using an illustrative scenario, we first have to determine a typical real-world use case. There are three use cases that will be evaluated:

- A UX researcher who wants to use quantitative emotion measurement because their user group cannot be assessed using surveys/observation. They have a list of needs for the measurement method.
- A hardware engineer who wants to create hardware to improve the field of quantitative emotion measurement by making measurement devices. The engineer wants to know what the most often used sensors are and make a device that combines these.
- A creative technologist is wondering if some less-often used sensors might be useful for measuring valence. They want to know how valence is currently measured to compare their research.

5.1.1 - UX researcher

The UX researcher wants to measure if their educational game is evoking the right experience for the target group. The researcher is worried that the game might make the users stressed and therefore wants to measure if the participants of the user tests are stressed. The researcher also wants to know if the participants of the user test are happy during the gameplay. This is also very important to the researcher because the game is aimed towards children, and he wants to make sure they associate learning with fun. The researcher is hesitant to use surveys because he knows that the test group does not want to make the researcher sad, so they always tell him they liked the game. He also noticed that some of the younger participants seemed to not fully understand the questions he was asking. The researcher might want to use emotion measurement in addition to surveys, but he has some requirements. Firstly, he does not want the sensors to distract the children because he wants them to focus on the game. Secondly, the researcher would preferably measure stress and happiness but if he can use less intrusive sensors if he switches to the dimensional emotional model, he is willing to do so. Lastly, he wants to have an accuracy of at least 70%.

The researcher, Jim, uses this taxonomy to make his choices. Firstly, he reads the table of sensors to get an idea of what is available (table 4.2). He reads that pupillometry can be used to determine liking which satisfies one of his needs. This is also a method that does not need physical contact with the participant, which satisfies another need. Jim finds measuring stress a bit harder. The most often used sensors are PPG and GSR which are quite intrusive (figure 4.2). These sensors generally mount at the wrist or fingers which can really distract the young participants. After a bit of research, he finds that GSR can be mounted under the earlobe too [65]. While this is still intrusive, it will not be on the hands or peripheral vision of the participants, which Jim is okay with. From table 3 in appendix E Jim can find the papers that can inform him about existing methods using the sensors he chose. From the list of papers, he finds some mentioning a reported accuracy of over 70% which is according to Jim's needs. With this information, Jim knows what he needs to measure what he wants according to his requirements.

5.1.2 - Hardware engineer

The hardware engineer, Jane, wants to create a wrist mounted sensor band which can be used for emotion measurement. Jane does not care about the emotions that can be measured; she just wants to create a low cost band that can improve the availability of emotion measurement devices. Jane can see from figure 4.6 that PPG, GSR, and ECG are the most popular methods. She also finds the Empatica E4 band which combines PPG, GSR, IMU, and temperature [66]. Jane decides to expand on this by creating a sensor band that does not have an IMU but includes other sensors. Jane removes the IMU because it is not often used. She includes a microphone and a modular EMG unit. She also decides to do some signal processing on the band, to derive an ECG from the PPG signal. While this ECG will not be as accurate as chest mounted ones, it will open the door to more emotion measurement capabilities.

5.1.3 - Creative technologist

The creative technologist, Susanne, finds a few less often used sensors from figure 4.6. She chooses to research the usefulness of blood pressure, heart rate, temperature, and EMG for measuring valence. After conducting her research, she wants to compare her results to the most often used methods. Using figure 4.2 she finds that EEG, PPG, and GSR are some of the most often used methods. In table 3 in appendix E she can find the names and numbers of the papers using this method and can find the reported accuracy in the full taxonomy in appendix B. If she wants to learn more about a paper, she can find the full name in appendix A and take a closer look at it herself.

5.2 - Logical argument

Evaluating a taxonomy by logical argument is subjective, but it can show the place of the taxonomy within existing research well. As stated in chapter 2, there is no taxonomy about this topic yet. There are, however, multiple literature reviews about emotion measurement using biosignals. Two of these reviews will be compared to this taxonomy, to see where this taxonomy compares or exceeds existing work, and where this taxonomy lacks depth compared to the other two. The two review that will be compared to this taxonomy are:

5.2.1 - A Survey of Emotion Recognition using Physiological Signal in Wearable Devices

This review performed a SLR with 26 papers [36]. The inclusion/exclusion criteria are remarkably similar to those used in this paper, except for the participant requirement. [36] requires studies to have at least 10 participants to be included, a requirement that was not made for this taxonomy. It also does not take the term sensor as broadly as this taxonomy does, [36] only includes physiological signals, so FEA or pupillometry are not included. Wijasena and colleagues included a few interesting characteristics which were not included in this paper: stimulus and environment [36]. It also misses a few characteristics which were included in this taxonomy: sensor fusion, participants, and not all characteristics of the emotional model. The review presents a more in-depth view of algorithms and filtering, something that misses from this taxonomy. It misses the figures and charts present in this taxonomy such as the Sankey charts and the table with dots for occurrence, which limits its potential for advising emotions and sensors. Overall, “A Survey of Emotion Recognition using Physiological Signal in Wearable Devices” offers a good insight into the field of emotion measurement with more insight into the algorithms and filtering aspect than what this taxonomy offers. It does lack the depth of connection between sensors and emotions which this taxonomy has, making it less applicable to motivate sensor/emotion choices for other researchers.

5.2.2 -Emotion Recognition Using Wearables: A Systematic Literature Review – Work-in-progress

This review includes a SLR which yielded 27 papers [67]. It has similar inclusion/exclusion criteria to this paper except for requiring at least 5 participants. This paper is similar to [36] in that they only review physiological signals. It presents similar findings as [36] with even more emphasis on the algorithms, feature extraction, and filtering of the signal. It does not present any connection between the used sensors and the emotion they were measuring, which makes it less usable than [36] and this taxonomy for motivating emotion and sensor choices. It does present algorithms and signal processing in more depth than [36] and this taxonomy.

5.3 - Evaluation

If we look at both the illustrative scenarios and the logical argument, we see that this taxonomy has a valid use case in the real world. We also see that this taxonomy is lacking in some respects, especially the depth in which algorithms, feature extraction, and filtering are discussed and included. Considering the research questions and sub-questions, this taxonomy fulfilled its goal. It presents sensor usage and emotion measurement in a quantitative way and can help guide researchers to find combinations that work for their use case. Expansion of this taxonomy can make it even more useful; this could be done by combining reported accuracy with algorithms and filtering methods for example. In conclusion, this taxonomy seems to be a valid tool which can be used to motivate a part of the creation of emotion measurement methods, but it does not provide enough information to advise choices in the whole process.

Chapter 6 - Discussion

The results shown in chapter 4 show an interesting picture of the current state of emotion measurement using physiological sensors. Relevant, recent papers have been systematically reviewed and the findings categorized. Major trends can be seen in figure 4.5 and 4.6 and a visualization of sensor usage pertaining to emotions offers an opportunity to other researchers to quickly find combinations suitable to their needs as shown in the evaluation chapter (figure 4.2 and other Sankey diagrams). In this chapter, some reasoning for why the results is what they are will be discussed after which limitations to this research and avenues for future research will be presented.

6.1 - Interpretation

One key thing this taxonomy has shown is that the distribution of usage among sensors and emotions is not even. We can see that the usage of PPG and GSR is unusually high. This is because of many reasons, but a few could be identified. Firstly, these sensors are very useful. [47] suggests that GSR is very useful for arousal measurement and [55] showcases the usability of PPG for valence well. This makes the combination of the two very logical, so it is not strange that these are the most used. However, this cannot be the only reason. While PPG and GSR are very useful, other sensors are too. Research that does not use one of these two sensors achieves similar results with different sensors, so why are PPG and GSR still so popular? The hardware used in the papers that were reviewed varies, but one piece of hardware was mentioned quite often. This was the Empatica E4 band. This band offers PPG, GSR, temperature, and movement [66]. The availability of this well-made well tested band which has been used in similar research previously makes it an attractive choice. Another reason is the limitations of datasets. Most papers that utilized machine learning to recognize emotions from the signal utilized public datasets to train their algorithms. While this is great for creating well trained algorithms, it limits the possible sensor that can be used in an emotion measurement system because these datasets only record a select number of sensors.

The trends in emotions are shown well in the scatter plot in figure 4.5. This scatter plot shows a circular pattern in the valence/arousal chart. This is not too strange; the emotions close to the middle of the graph are not that intense so measuring them is harder than measuring emotions closer to the edge. Still, the scatter plot does not show an even distribution along the circle border, and some emotions are measured in way higher proportion than others. For some emotions this is logical. Stress, for example, is a quite popular research topic in general, so it is not strange that this is well represented in the emotion measurement field. The motivation for some other emotions is not that obvious. Two other reasons have been identified. Firstly, a lot of papers utilizing the discrete emotional model use a common theory of basic emotion to motivate their choice of emotions to measure, like Ekman's proposal [49]. This is why those emotions occur more often than others. Secondly, the datasets cause trends in emotion selection, just as in sensor usage. This is for the same reason, not all emotions can be recorded in the datasets, so researchers that use them are limited to those included in the dataset.

6.2 - Limitations

In the evaluation chapter some limitations to this taxonomy came to light. These will be expended upon. The inclusion/exclusion criteria limited the search to papers published after/during 2019. The three-year period of papers yielded results that seem useful, but a larger period could have produced more insights, especially pertaining to the testing of usability of sensors for emotion measurement instead of implementation. In the evaluation chapter some characteristics came to light that were not included in this taxonomy that would have been useful. These are emotion evocation methods, experiment setting, and signal processing. Signal processing and algorithms in general were not expanded on in this taxonomy, which limits the scope of usability.

One major issue with this taxonomy, and reviews about emotion measurement in general, is the interchangeability of datasets and emotional models. The datasets that are used all differ in sensors that are used and what they use for emotional evocation. Therefore, comparing results of papers that use different datasets is not really fair. Further, the emotional models in general are hard to combine. In this taxonomy the model proposed by [35] [36] was used, but this is not the model that all papers included in this paper have used themselves. This makes comparing them objectionable because they might not be talking about measuring the same affect. This is especially true for accuracy comparisons, the showcase of sensor usage in combination with emotions which is shown in this paper is not impacted as badly by this because no quantitative comparisons are made. This problem is quite hard to solve because the field is not large enough to make large taxonomies or reviews about papers only using certain datasets, sensors, or emotional models. When the field grows this might be a good direction to go in to but for now, it is a necessary evil.

Lastly, there is a discussion to be had about the validity of emotion measurement in general. Law and colleagues already showed hesitance to UX measurement in general, and emotion measurement is especially hard to prove [2]. The way emotion measurement works in its current state is usually through training an algorithm using datasets and using that algorithm to measure new signals. It can be argued that you are not measuring emotions at all, just measuring similar reactions to the emotional evocation methods used to create the dataset. This is a limitation which is hard to solve and something that can benefit from discussing it more. An ethical reflection about this topic can be found in appendix F.

6.3 - Future research

There are many possibilities for future research in this field. Further research into the algorithms used for emotion measurement in combination with the sensors used and emotions measured can be very valuable. This taxonomy only notes down the used algorithms but does not go into depth about their effectiveness. Furthermore, as discussed in the paragraph interpreting the results the possibilities of emotion measurement are very dependent on datasets. The creation of more datasets with different emotions and sensors can be a great research contribution that can help expand the knowledge and possibilities in the field tremendously. Similarly, hardware combining different sensors into one package can achieve the same result, improving the possibility of researching different sensor combinations and making emotion measurement more accessible in general.

Chapter 7 - Conclusion

The goal of this graduation project was to provide an overview of the possibilities of emotion measurement and a way to advise future researchers in their choice of emotion to measure and sensors to use. The taxonomy presented in the report achieves this, showing trends in the field and an easy way to find similar research to aid in the creation of new emotion measurement methods. The reason why emotion measurement is important is because of its applicability to user groups that could not be previously assessed. The research shows that this seems to hold up, although these user groups were not often used in papers. Emotion measurement seems to be possible in many ways with many possible combinations of sensors and emotions that have been and can be tried. The results presented by the papers so far are promising and streamlining and expanding emotion measurement research can solidify its place in UX research.

Reference list

[1] E. L.-C. Law and S. Abrahão, "Interplay between User Experience (UX) evaluation and system development," International Journal of Human-Computer Studies, vol. 72, no. 6, pp. 523–525, Jun. 2014, doi: 10.1016/j.ijhcs.2014.03.003.

[2] E. L.-C. Law, P. van Schaik, and V. Roto, "Attitudes towards user experience (UX) measurement," International Journal of Human-Computer Studies, vol. 72, no. 6, pp. 526–541, Jun. 2014, doi: 10.1016/j.ijhcs.2013.09.006.

[3] H. Hochheiser and J. Lazar, "HCI and Societal Issues: A Framework for Engagement," International Journal of Human-Computer Interaction, vol. 23, no. 3, pp. 339–374, Dec. 2007, doi: 10.1080/10447310701702717.

[4] "The Dvorak Keyboard," mit.edu, 2021. https://www-mit-edu/~jcb/Dvorak/ (accessed Nov. 25, 2021).

[5] I. Pettersson, F. Lachner, A.-K. Frison, A. Riener, and A. Butz, "A Bermuda Triangle?," Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Apr. 2018, doi: 10.1145/3173574.3174035.

[6] R. W. Robins, R Chris Fraley, and R. F. Krueger, Handbook of research methods in personality psychology. New York ; London: Guilford, 2010.

[7] G. T. Fujiura, "Self-Reported Health of People with Intellectual Disability," Intellectual and Developmental Disabilities, vol. 50, no. 4, pp. 352–369, Aug. 2012, doi: 10.1352/1934-9556-50.4.352.

[8] Faruk Arici and Rabia Meryem Yılmaz, "The effect of laboratory experiment and interactive simulation use on academic achievement in teaching...," ResearchGate, Mar. 15, 2020. https://www.researchgate.net/publication/339612585_The_effect_of_laboratory_experiment_and_interactive_simulation_use_on_academic_achievement_in_teaching_secondary_school_force_and_movement_unit (accessed Dec. 01, 2021).

[9] J. O. Wobbrock and J. A. Kientz, "Research contribution in human-computer interaction," interactions, vol. 23, no. 3, pp. 38–44, Apr. 2016, doi: 10.1145/2907069.

[10] Iho Kalmyk, “AI emotion detection for user research — a UX case study,” Medium, Oct. 14, 2019. https://uxdesign.cc/ai-emotion-detection-for-user-research-a-ux-case-study-be6bc485058d (accessed Dec. 01, 2021).

[11] V. Roto, E. L.-C. Law, A. Vermeeren, en J. Hoonhout, “Demarcating User Experience Dagstuhl Seminar”, 2011.

[12] J. Lazar, Jinjuan Heidi Feng, and H. Hochheiser, Research methods in human computer interaction. Cambridge, Ma: Morgan Kaufmann Publishers, An Imprint Of Elsevier, 2017.

[13] J. Hernandez, I. Riobo, A. Rozga, G. D. Abowd, and R. W. Picard, “Using electrodermal activity to recognize ease of engagement in children during social interactions,” *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Sep. 2014, doi: 10.1145/2632048.2636065.

[14] S. Saganowski et al., “Emotion Recognition Using Wearables: A Systematic Literature Review - Work-in-progress,” 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Mar. 2020, doi: 10.1109/percomworkshops48775.2020.9156096.

[15] J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” Information Fusion, vol. 59, pp. 103–126, Jul. 2020, doi: 10.1016/j.inffus.2020.01.011.

[16] S. P S en M. G S, “Emotion Models: A Review”, International Journal of Control Theory and Applications, vol 10, bll 651–657, 01 2017.

[17] N. Du et al., “Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving,” Transportation Research Part C: Emerging Technologies, vol. 112, pp. 78–87, Mar. 2020, doi: 10.1016/j.trc.2020.01.006.

[18] K. Doherty and G. Doherty, “Engagement in HCI,” ACM Computing Surveys, vol. 51, no. 5, pp. 1–39, Jan. 2019, doi: 10.1145/3234149.

[19] “FilmStim (Schaefer & al.),” Uclouvain.be, 2016. https://sites.uclouvain.be/ipsp/FilmStim/ (accessed Oct. 27, 2021).

[20] Sander Koelstra, “HCI Tagging Database - Home,” Mahnob-db.eu, 2021. https://mahnob-db.eu/hci-tagging/ (accessed Oct. 27, 2021).

[21] "Vempathy," CapSource, 2018. https://capsource.io/companies/vempathy/ (accessed Nov. 28, 2021).

[22] Iho Kalmyk, "AI emotion detection for user research — a UX case study," Medium, Oct. 14, 2019. https://uxdesign.cc/ai-emotion-detection-for-user-research-a-ux-case-study-be6bc485058d (accessed Nov. 29, 2021).

[23] M. Pantie and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424–1445, 2000, doi: 10.1109/34.895976.

[24] V.A. Petrushin "Emotion in speech: recognition and application to call centers" Proceeding of the 1999 Conference on Artificial Neural Networks in Engineering (1999), pp. 7-10

[25] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), vol. 36, no. 1, pp. 96–105, Feb. 2006, doi: 10.1109/tsmcb.2005.854502.

[26] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," Computer Methods and Programs in Biomedicine, vol. 140, pp. 93–110, Mar. 2017, doi: 10.1016/j.cmpb.2016.12.005.

[27] The Engineer, "EEG device spells words and phrases | The Engineer," The Engineer, Apr. 09,2010. https://www.theengineer.co.uk/eeg-device-spells-words-and-phrases/ (accessed Dec. 01, 2021).

[28] E. de Souza Veriscimo, J. L. Bernardes Júnior, and L. A. Digiampietri, "Facial Emotion Recognition in UX Evaluation: A Systematic Review," Human-Computer Interaction. Theory, Methods and Tools, pp. 521–534, 2021, doi: 10.1007/978-3-030-78462-1_40.

[29] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review," Applied Sciences, vol. 7, no. 12, p. 1239, Dec. 2017, doi: 10.3390/app7121239.

[30] A. Landowska, "Towards Emotion Acquisition in IT Usability Evaluation Context," Proceedings of the Mulitimedia, Interaction, Design and Innnovation on ZZZ - MIDI '15, 2015, doi: 10.1145/2814464.2814470.

[31] R. C. Nickerson, U. Varshney, and J. Muntermann, “A method for taxonomy development and its application in information systems,” European Journal of Information Systems, vol. 22, no. 3, pp. 336–359, May 2013, doi: 10.1057/ejis.2012.26.

[32] H. Z. Wijasena, R. Ferdiana, and S. Wibirama, “A Survey of Emotion Recognition using Physiological Signal in Wearable Devices,” 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), Apr. 2021, doi: 10.1109/aims52415.2021.9466092.

[33] J. Hertel, S. Karaosmanoglu, S. Schmidt, J. Braker, M. Semmann, and F. Steinicke, “A Taxonomy of Interaction Techniques for Immersive Augmented Reality based on an Iterative Literature Review,” 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Oct. 2021, doi: 10.1109/ismar52148.2021.00060.

[34] Elsevier, “About Scopus - Abstract and citation database | Elsevier,”https://www-elsevier-com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030 (accessed Feb. 10, 2022).

[35] L. Camras and R. Plutchik, “Emotion: A Psychoevolutionary Synthesis,” The American Journal of Psychology, vol. 93, no. 4, p. 751, Dec. 1980, doi: 10.2307/1422394.

[36] D. Kollias et al., “Deep Affect Prediction in-the-Wild: Aff-Wild Database and Challenge, Deep Architectures, and Beyond,” International Journal of Computer Vision, vol. 127, no. 6–7, pp. 907–929, Feb. 2019, doi: 10.1007/s11263-019-01158-4.

[37] M. Ghamari, “A review on wearable photoplethysmography sensors and their potential future applications in health care,” International Journal of Biosensors & Bioelectronics, vol. 4, no. 4, 2018, doi: 10.15406/ijbsbe.2018.04.00125.

[38] “EEG (electroencephalogram) - Mayo Clinic,” Mayoclinic.org, 2020. https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875 (accessed Feb. 10, 2022).

[39] H. Lu, M. Wang, and A. K. Sangaiah, “Human Emotion Recognition Using an EEG Cloud Computing Platform,” Mobile Networks and Applications, Sep. 2018, doi: 10.1007/s11036-018-1120-1.

[40] T. Wilcox and M. Biondi, “fNIRS in the developmental sciences,” WIREs Cognitive Science, vol. 6, no. 3, pp. 263–283, Feb. 2015, doi: 10.1002/wcs.1343.

[41] J. Nie et al., “SPIDERS+: A light-weight, wireless, and low-cost glasses-based wearable platform for emotion sensing and bio-signal acquisition,” Pervasive and Mobile Computing, vol. 75, p. 101424, Aug. 2021, doi: 10.1016/j.pmcj.2021.101424.

[42] M. B. Winn, D. Wendt, T. Koelewijn, and S. E. Kuchinsky, “Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started,” Trends in Hearing, vol. 22, p. 233121651880086, Jan. 2018, doi: 10.1177/2331216518800869.

[43] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar, “CorrNet: Fine-Grained Emotion Recognition for Video Watching Using Wearable Physiological Sensors,” Sensors, vol. 21, no. 1, p. 52, Dec. 2020, doi: 10.3390/s21010052.

[44] H. Kalantarian, P. Washington, J. Schwartz, J. Daniels, N. Haber, and D. P. Wall, “Guess What?,” Journal of Healthcare Informatics Research, vol. 3, no. 1, pp. 43–66, Oct. 2018, doi: 10.1007/s41666-018-0034-9.

[45] K. Niven, P. Totterdell, E. Miles, T. L. Webb, and P. Sheeran, “Achieving the same for less: Improving mood depletes blood glucose for people with poor (but not good) emotion control,” Cognition & Emotion, vol. 27, no. 1, pp. 133–140, Jan. 2013, doi: 10.1080/02699931.2012.679916.

[46] C. B. Redd et al., “Physiological Signal Monitoring for Identification of Emotional Dysregulation in Children,” 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jul. 2020, doi: 10.1109/embc44109.2020.9176506.

[47] E. Fortune, Y. Yusuf, and R. Blocker, “Measuring Arousal and Emotion in Healthcare Employees Using Novel Devices,” 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Oct. 2020, doi: 10.1109/bibe50027.2020.00142.

[48] Q. Xu, X. Liu, J. Luo, and Z. Tang, “Emotion monitoring with RFID: an experimental study,” CCF Transactions on Pervasive Computing and Interaction, vol. 2, no. 4, pp. 299–313, Oct. 2020, doi: 10.1007/s42486-020-00043-1.

[49] P. Ekman, “An argument for basic emotions,” Cognition and Emotion, vol. 6, no. 3–4, pp. 169–200, May 1992, doi: 10.1080/02699939208411068.

[50] M. D, E. Watkins, and E. Harmon-Jones, Handbook of cognition and emotion. New York: The Guilford Press, 2013.

[51] Md. G. R. Alam, S. F. Abedin, S. I. Moon, A. Talukder, and C. S. Hong, "Healthcare IoT-Based Affective State Mining Using a Deep Convolutional Neural Network," IEEE Access, vol. 7, pp. 75189–75202, 2019, doi: 10.1109/access.2019.2919995.

[52] J. Watson et al., "Using fNIRS and EDA to Investigate the Effects of Messaging Related to a Dimensional Theory of Emotion," Advances in Neuroergonomics and Cognitive Engineering, pp. 59–67, Jun. 2019, doi: 10.1007/978-3-030-20473-0_6.

[53] T. Zhang, A. El Ali, C. Wang, X. Zhu, and P. Cesar, "CorrFeat: Correlation-based Feature Extraction Algorithm using Skin Conductance and Pupil Diameter for Emotion Recognition," 2019 International Conference on Multimodal Interaction, Oct. 2019, doi: 10.1145/3340555.3353716.

[54] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems," Journal of Medical and Biological Engineering, vol. 40, no. 2, pp. 149–157, Jan. 2020, doi: 10.1007/s40846-019-00505-7.

[55] A. S. Ravindran, S. Nakagome, D. S. Wickramasuriya, J. L. Contreras-Vidal, and R. T. Faghih, "Emotion Recognition by Point Process Characterization of Heartbeat Dynamics," 2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT), Nov. 2019, doi: 10.1109/hi-poct45284.2019.8962886.

[56]J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutierrez, M. P. Garcia, and C. Lopez-Ongil, "Toward Fear Detection using Affect Recognition," 2019 XXXIV Conference on Design of Circuits and Integrated Systems (DCIS), Nov. 2019, doi: 10.1109/dcis201949030.2019.8959852.

[57] G. J. Nalepa, K. Kutt, B. Giżycka, P. Jemioło, and S. Bobek, "Analysis and Use of the Emotional Context with Wearable Devices for Games and Intelligent Assistants," Sensors, vol. 19, no. 11, p. 2509, May 2019, doi: 10.3390/s19112509.

[58] "Shimmer3 GSR+ Unit," Shimmer Wearable Sensor Technology, Oct. 13, 2021. https://shimmersensing.com/product/shimmer3-gsr-unit/ (accessed Feb. 14, 2022).

[59] "Bitalino," Bitalino.com, 2022. https://bitalino.com/products/board-kit-bt (accessed Feb. 15, 2022).

[60] R. Francese, M. Risi, and G. Tortora, “A user-centered approach for detecting emotions with low-cost sensors,” Multimedia Tools and Applications, vol. 79, no. 47–48, pp. 35885–35907, Nov. 2020, doi: 10.1007/s11042-020-09576-0.

[61] A. Bagirathan, J. Selvaraj, A. Gurusamy, and H. Das, “Recognition of positive and negative valence states in children with autism spectrum disorder (ASD) using discrete wavelet transform (DWT) analysis of electrocardiogram signals (ECG),” Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 1, pp. 405–416, May 2020, doi: 10.1007/s12652-020-01985-1.

[62] S. Jiang, P. Zhou, Z. Li, and M. Li, “Memento: An Emotion Driven Lifelogging System with Wearables,” 2017 26th International Conference on Computer Communication and Networks (ICCCN), Jul. 2017, doi: 10.1109/icccn.2017.8038411.

[63] “SankeyMATIC: Build a Sankey Diagram,” Sankeymatic.com, 2022. https://sankeymatic.com/build/ (accessed Feb. 10, 2022).

[64] Szopinski, D., Schoormann, T., Kundisch, D.: Because your taxonomy is worth it: towards a framework for taxonomy evaluation. In: Proceedings of the Twenty-Seventh European Conference on Information Systems (ECIS) (2019)

[65] A. Cruz, G. Pires, A. C. Lopes, and U. J. Nunes, “Detection of Stressful Situations Using GSR While Driving a BCI-controlled Wheelchair,” 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2019, doi: 10.1109/embc.2019.8857748.

[66] E. Temperature , “E4 wristband | Real-time physiological signals | Wearable PPG, EDA, Temperature, Motion sensors,” Empatica, 2020. https://www.empatica.com/research/e4/ (accessed Feb. 20, 2022).

[67] S. Saganowski et al., “Emotion Recognition Using Wearables: A Systematic Literature Review - Work-in-progress,” 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Mar. 2020, doi: 10.1109/percomworkshops48775.2020.9156096.

Appendix A

1. Banos, O., Castro, L.A., Villalonga, C. Ubiquitous technologies for emotion recognition (2021) 11 (15), art. no. 7019, .

2. Mai, N.-D., Lee, B.-G., Chung, W.-Y. Affective computing on machine learning-based emotion recognition using a self-made eeg device (2021) 21 (15), art. no. 5135, .

3. Nie, J., Liu, Y., Hu, Y., Wang, Y., Xia, S., Preindl, M., Jiang, X. SPIDERS+: A light-weight, wireless, and low-cost glasses-based wearable platform for emotion sensing and bio-signal acquisition (2021) 75, art. no. 101424, .

4. Chaowadee, N., Lertsiriyothin, P., Phuangkhemkhao, T., Chanwimalueang, T. Reinforced Learning in Children through a Stress Warning Unit (2021) art. no. 9441507, pp. 386-389.

5. Wijasena, H.Z., Ferdiana, R., Wibirama, S. A Survey of Emotion Recognition using Physiological Signal in Wearable Devices (2021) art. no. 9466092, .

6. Buda, T.S., Khwaja, M., Matic, A. Outliers in Smartphone Sensor Data Reveal Outliers in Daily Happiness (2021) 5 (1), art. no. 3448095, .

7. Preethi, M., Nagaraj, S., Madhan Mohan, P. Emotion based Media Playback System using PPG Signal (2021) art. no. 9419475, pp. 426-430.

8. Miranda, J.A., Canabal, M.F., Gutiérrez-Martín, L., Lanza-Gutierrez, J.M., Portela-García, M., López-Ongil, C. Fear recognition for women using a reduced set of physiological signals (2021) 21 (5), art. no. 1587, pp. 1-31.

9. Rumiantcev, M. Emotions and activity recognition system using wearable device sensors (2021) 2021-January, art. no. 9347652, .

10. Yang, K., Wang, C., Gu, Y., Sarsenbayeva, Z., Tag, B., Dingler, T., Wadley, G., Goncalves, J. Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition (2021) .

11. Yoo, H., Baek, J.-W., Chung, K. Cnn-based voice emotion classification model for risk detection (2021) 29 (2), pp. 329-334. Activity and Behavior Computing conference, ABC 2020 (2021) 204, 306 p.

12. Activity and Behavior Computing conference, ABC 2020

13. M. Hassib, M. Braun, B. Pfleging, and F. Alt, "Detecting and Influencing Driver Emotions Using Psycho-Physiological Sensors and Ambient Light," *Human-Computer Interaction – INTERACT 2019*, pp. 721–742, 2019, doi: 10.1007/978-3-030-29381-9_43.

14. Arora, A., Chakraborty, P., Bhatia, M.P.S. Problematic Use of Digital Technologies and Its Impact on Mental Health During COVID-19 Pandemic: Assessment Using Machine Learning (2021) 348, pp. 197-221.

15. 12th International Conference on Intelligent Human Computer Interaction, IHCI 2020 (2021) 12615 LNCS, 942 p.

16. 12th International Conference on Intelligent Human Computer Interaction, IHCI 2020 (2021) 12616 LNCS, 942 p.

17. Zhang, T., Ali, A.E., Wang, C., Hanjalic, A., Cesar, P. Cornet: Fine-grained emotion recognition for video watching using wearable physiological sensors (2021) 21 (1), art. no. 52, pp. 1-25.

18. Santamaria-Granados, L., Mendoza-Moreno, J.F., Ramirez-Gonzalez, G. Tourist recommender systems based on emotion recognition—a scientometric review (2021) 13 (1), art. no. 2, pp. 1-38.

19. Bagirathan, A., Selvaraj, J., Gurusamy, A., Das, H. Recognition of positive and negative valence states in children with autism spectrum disorder (ASD) using discrete wavelet transform (DWT) analysis of electrocardiogram signals (ECG) (2021) 12 (1), pp. 405-416.

20. Alam, M.A.U., Roy, N., Holmes, S., Gangopadhyay, A., Galik, E. AutoCogniSys: IoT assisted context-aware automatic cognitive health assessment (2020) pp. 184-195.

21. Saganowski, S., Kazienko, P., Dziezyc, M., Jakimow, P., Komoszynska, J., Michalska, W., Dutkowiak, A., Polak, A., Dziadek, A., Ujma, M. Consumer wearables and affective computing for wellbeing support (2020) pp. 482-487.

22. Xu, Q., Liu, X., Luo, J., Tang, Z. Emotion monitoring with RFID: an experimental study (2020) 2 (4), pp. 299-313.

23. Li, W., Rovetta, A., DIng, X., Chen, L., Han, Y. Virtual Multi-Interaction for Rehabilitation Robotics (2020) art. no. 9195384, pp. 204-208.

24. Fairclough, S.H., Dobbins, C. Personal informatics and negative emotions during commuter driving: Effects of data visualization on cardiovascular reactivity & mood (2020) 144, art. no. 102499, .

25. Gasparini, F., Giltri, M., Bandini, S. Discriminating affective state intensity using physiological responses (2020) 79 (47-48), pp. 35845-35865.

26. Dzieżyc, M., Gjoreski, M., Kazienko, P., Saganowski, S., Gams, M. Can we ditch feature engineering? End-to-end deep learning for affect recognition from physiological sensor data (2020) 20 (22), art. no. 6535, pp. 1-21.

27. Dong, W., Cheng, X., Yao, D. Soft body sensor networks: Architectures and Applications (2020) art. no. 9238051, .

28. Li, Y., Xia, X., Jiang, D., Sahli, H., Jain, R. MEMOS: A multi-modal emotion stream database for temporal spontaneous emotional state detection (2020) pp. 370-378.

29. Priebe, M., Foo, E., Holschuh, B. Shape Memory Alloy Haptic Compression Garment for Media Augmentation in Virtual Reality Environment (2020) pp. 34-36.

30. Fortune, E., Yusuf, Y., Blocker, R. Measuring Arousal and Emotion in Healthcare Employees Using Novel Devices (2020) art. no. 9287986, pp. 835-838.

31. Fujii, A., Murao, K., Matsuhisa, N. Disp2ppg: Pulse Wave Generation to PPG Sensor using Display (2020) pp. 119-123.

32. Bauer, J., Hechtel, M., Holzwarth, M., Sessner, J., Franke, J., Wittenberg, T., Waldhör, K., Schena, A., Falanga, A., Ristok, B., Böhle, F. A Concept for Context Awareness in Smart Environments (2020) 6 (3), art. no. 20203098, .

33. Cecchi, S., Piersanti, A., Poli, A., Spinsante, S. Physical Stimuli and Emotions: EDA Features Analysis from a Wrist-Worn Measurement Sensor (2020) 2020-September, art. no. 9209307, .

34. Li, F., Shirahama, K., Nisar, M.A., Huang, X., Grzegorzek, M. Deep transfer learning for time series data based on sensor modality classification (2020) 20 (15), art. no. 4271, pp. 1-25.

35. Antoniou, P.E., Arfaras, G., Pandria, N., Athanasiou, A., Ntakakis, G., Babatsikos, E., Nigdelis, V., Bamidis, P. Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: Cohort study (2020) 8 (3), art. no. e17823, .

36. Ramli, N.A., Nordin, A.N., Azlan, N.Z. Review on strain sensors for detection of human facial expressions recognition systems (2020) 13 (3), pp. 445-472.

37. Redd, C.B., Silvera-Tawil, D., Hopp, D., Zandberg, D., Martiniuk, A., Dietrich, C., Karunanithi, M.K. Physiological Signal Monitoring for Identification of Emotional Dysregulation in Children (2020) 2020-July, art. no. 9176506, pp. 4273-4277.

38. Lu, H., Wang, M., Sangaiah, A.K. Human Emotion Recognition Using an EEG Cloud Computing Platform (2020) 25 (3), pp. 1023-1032.

39. M Saad, W.H., Rahman, N.A., Karis, M.S., Chia, S.L., A Karim, S.A., Talib, M.H., Razak, M.S.J.A. Analysis on continuous wearable device for blood glucose detection using GSR sensor (2020) 13 (Special Issue ISSTE 2019), pp. 9-16.

40. Nie, J., Hu, Y., Wang, Y., Xia, S., Jiang, X. SPIDERS: Low-cost wireless glasses for continuous in-situ bio-signal acquisition and emotion recognition (2020) art. no. 9097620, pp. 27-39.

41. Patlar Akbulut, F., Ikitimur, B., Akan, A. Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome (2020) 104, art. no. 101824, .

42. Ayata, D., Yaslan, Y., Kamasak, M.E. Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems (2020) 40 (2), pp. 149-157.

43. Saganowski, S., Dutkowiak, A., Dziadek, A., Dziezyc, M., Komoszynska, J., Michalska, W., Polak, A., Ujma, M., Kazienko, P. Emotion Recognition Using Wearables: A Systematic Literature Review-Work-in-progress (2020) art. no. 9156096, .

44. Sugimoto, M., Hamasaki, S., Yajima, R., Yamakawa, H., Takakusaki, K., Nagatani, K., Yamashita, A., Asama, H. Incident detection at construction sites via heart-rate and emg signal of facial muscle (2020) pp. 886-891.

45. Yang, C.-J., Fahier, N., He, C.-Y., Li, W.-C., Fang, W.-C. An AI-edge platform with multimodal wearable physiological signals monitoring sensors for affective computing applications (2020) 2020-October, art. no. 9180909, .

46. D'Arco, L., Zheng, H., Wang, H. SenseBot: A wearable sensor enabled robotic system to support health and well-being (2020) 2815, pp. 15-29.

47. Nakisa, B., Rastgoo, M.N., Rakotonirainy, A., Maire, F., Chandran, V. Automatic Emotion Recognition Using Temporal Multimodal Deep Learning (2020) 8, art. no. 9206543, pp. 225463-225474.

48. 21th Annual Conference on Towards Autonomous Robotics, TAROS 20120 (2020) 12228 LNAI, 412 p.

49. Yun, H., Fortenbacher, A., Helbig, R., Geißler, S., Pinkwart, N. Emotion Recognition from Physiological Sensor Data to Support Self-regulated Learning (2020) 1220, pp. 155-173.

50. Antoniou, P., Arfaras, G., Pandria, N., Ntakakis, G., Bambatsikos, E., Athanasiou, A. Real-Time Affective Measurements in Medical Education, Using Virtual and Mixed Reality (2020) 12462 LNAI, pp. 87-95.

51. Emsawas, T., Kimura, T., Fukui, K.-I., Numao, M. Comparative Study of Wet and Dry Systems on EEG-Based Cognitive Tasks (2020) 12241 LNAI, pp. 309-318.

52. Fortune, E., Yusuf, Y., Zornes, S., Lopez, J.L., Blocker, R. Assessing induced emotions in employees in a workplace setting using wearable devices (2020) .

53. Bara, C.P., Papakostas, M., Mihalcea, R. A deep learning approach towards multimodal stress detection (2020) 2614, pp. 67-81.

54. AbdElnapi, N.M., Omran, N.F., Ali, A.A., Omara, F.A. A framework for stroke prevention using IoT healthcare sensors (2020) 114, pp. 175-186.

55. McStay, A. Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy (2020) 7 (1), .

56. Roshini, A., Kiran, K.V.D. Challenges in physiological signal extraction from cognitive radio wireless body area networks for emotion recognition (2020) 1108 AISC, pp. 504-513.

57. AHFE International Conference on Human Factors and Wearable Technologies, 2019 and the AHFE International Conference on Game Design and Virtual Environments, 2019 (2020) 973, 426 p.

58. Watson, J., Sargent, A., Topoglu, Y., Ye, H., Zhong, W., Suri, R., Ayaz, H. Using fnirs and EDA to investigate the effects of messaging related to a dimensional theory of emotion (2020) 953, pp. 59-67.

59. Aich, S., Chakraborty, S., Sim, J.-S., Jang, D.-J., Kim, H.-C. The design of an automated system for the analysis of the activity and emotional patterns of dogs with wearable sensors using machine learning (2019) 9 (22), art. no. 4938, .

60. Miranda, J.A., Canabal, M.F., Lanza-Gutierrez, J.M., Garcia, M.P., Lopez-Ongil, C. Toward Fear Detection using Affect Recognition (2019) art. no. 8959852, .

61. Zhai, D., Schiavone, G., De Raedt, W., Van Hoof, C. Investigation of Heart Rate Changes before and during/after Smoking Events in Free Living Conditions (2019) art. no. 8953091, pp. 137-140.

62. Ravindran, A.S., Nakagome, S., Wickramasuriya, D.S., Contreras-Vidal, J.L., Faghih, R.T. Emotion Recognition by Point Process Characterization of Heartbeat Dynamics (2019) art. no. 8962886, pp. 13-16.

63. Shu, J., Chiu, M., Hui, P. Emotion Sensing for Mobile Computing (2019) 57 (11), art. no. 8908557, pp. 84-90.

64. Zhang, T. Multi-modal fusion methods for robust emotion recognition using body-worn physiological sensors in mobile environments (2019) pp. 463-467.

65. Zhang, T., Ali, A.E., Wang, C., Zhu, X., Cesar, P. CorrFeat: Correlation-based feature extraction algorithm using skin conductance and pupil diameter for emotion recognition (2019) pp. 404-408.

66. Miao, Y., Wu, G., Liu, C., Hossain, M.S., Muhammad, G. Green Cognitive Body Sensor Network: Architecture, Energy Harvesting, and Smart Clothing-Based Applications (2019) 19 (19), art. no. 8466664, pp. 8371-8378.

67. Larradet, F., Niewiadomski, R., Barresi, G., Mattos, L.S. Appraisal theory-based mobile app for physiological data collection and labelling in the wild (2019) pp. 752-756.

68. Rathi, S.R., Deshpande, Y.D. Embedding affect awareness into online learning environment using deep neural network (2019) art. no. 9128811, .

69. Zainudin, M.N.S., Kee, Y.J., Idris, M.I., Kamaruddin, M.R., Ramlee, R.H. Recognizing the activity daily living (ADL) for subject independent (2019) 8 (3), pp. 5422-5427.

70. Beckmann, N., Viga, R., Dogangun, A., Grabmaier, A. Measurement and analysis of local pulse transit time for emotion recognition (2019) 19 (17), art. no. 8709953, pp. 7683-7692.

71. Kanjo, E., Younis, E.M.G., Ang, C.S. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection (2019) 49, pp. 46-56.

72. Montesinos, V., Dell'Agnola, F., Arza, A., Aminifar, A., Atienza, D. Multi-Modal Acute Stress Recognition Using Off-the-Shelf Wearable Devices (2019) art. no. 8857130, pp. 2196-2201.

73. Majumder, A.J.A., McWhorter, T.M., Ni, Y., Nie, H., Iarve, J., Ucci, D.R. SEmoD: A personalized emotion detection using a smart holistic embedded IoT system (2019) 1, art. no. 8754492, pp. 850-859.

74. Nalepa, G.J., Kutt, K., Zycka, B.G., Jemioło, P., Bobek, S. Analysis and use of the emotional context with wearable devices for games and intelligent assistants (2019) 19 (11), art. no. 2509, .

75. Kalantarian, H., Washington, P., Schwartz, J., Daniels, J., Haber, N., Wall, D.P. Guess What?: Towards Understanding Autism from Structured Video Using Facial Affect (2019) 3 (1), pp. 43-66.

76. Yuda, E., Tanabiki, T., Iwata, S., Abe, K., Hayano, J. Detection of daily emotions by wearable biometric sensors (2019) art. no. 8883968, pp. 286-287.

77. López-Ales, E., Herrrero, M.T., Palma, J. A machine learning approach for emotion detection through low-cost hardware (2019) 2609, .

78. 39th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, AI 2019 (2019) 11927 LNAI, 499 p.

79. Egger, M., Ley, M., Hanke, S. Emotion Recognition from Physiological Signal Analysis: A Review (2019) 343, pp. 35-55.

80. 7th International Conference on Distributed, Ambient and Pervasive Interactions, DAPI 2019, held as part of the 21st International Conference on Human-Computer Interaction, HCI International 2019 (2019) 11587 LNCS, 480 p.

81. Alam, M.G.R., Abedin, S.F., Moon, S.I., Talukder, A., Hong, C.S. Healthcare IoT-Based Affective State Mining Using a Deep Convolutional Neural Network (2019) 7, art. no. 8726298, pp. 1-15.

82. Jiang, S., Li, Z., Zhou, P., Li, M. Memento: An emotion-driven lifelogging system with wearables (2019) 15 (1), art. no. 8, .

83. Fallon, M., Spohrer, K., Heinzl, A. Wearable devices: A physiological and self-regulatory intervention for increasing attention in the workplace (2019) 29, pp. 229-238.

Appendix B

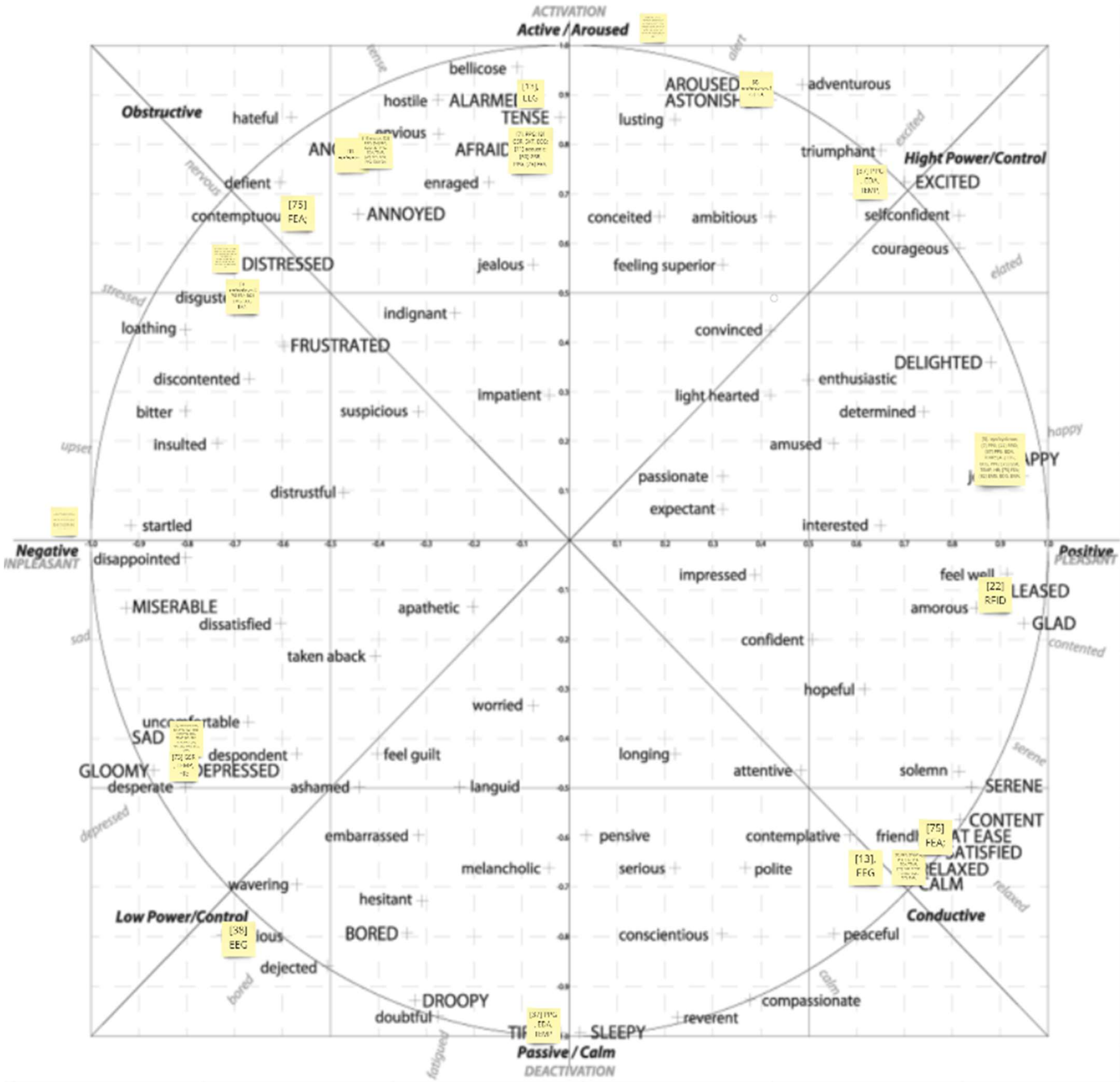
Number	Year	Sensors	Sensor fusion	Emotional model	Categories	Quadrant	Algorithm	Valence accuracy	Arousal accuracy	Accuracy	Sample size	Notes
1	2021	Review		Unspecified		N/A	Review	Review			N/A	
2	2021	EEG	N/A	Discrete	Pos, Neg, Neu	N/A	SVM, MLP, 1D-CNN			78.52%	8 (male, same ethinicity)	
3	2021	PPG, EEG, Eyebrow, Pupillometry, Zygomaticus	Applicable	Dimensional and Discrete	Val, Aro, 3 stages in both, happy, angry, disgusted, surprised, sad	Obstructive, positive, active, low power/control	Optical-flow based algorithm	49.32%	47.25%	83.87%	10 (3 different ethinicities)	
4	2021	PPG	N/A	Discrete	Stress	Obstructive	Peek detec, freq analysis					
5	2021	Review									N/A	
6	2021	Does not use sensors										
7	2021	PPG	N/A	Discrete	Calm, happy, fear, sad	conductive, postive, low power/control, obstructive/aroused	LDA			91.81%	150	
8	2021	ECG, SKT, GSR	Applicable	Discrete	Fear	obstructive/aroused/active	Frequency analysis, attribute analysis			96.33%, 85% (subj indept)	12 (all women)	
9	2021	Does not use sensors										
10	2021	acoustic, visual, typing, BVP, PPG, GSR	Applicable	Discrete	binary positive and negative	N/A	2d array			89.20%	45	
11	2021	Acoustic,	N/A	Discrete	Angry, fearful, surprised		CNN			77.20%	unknown	
12	2021											
13	2021	EEG	N/A	Discrete	Attention, stress, meditation	Active, obstructive, conductive					unknown	
14	2021	Not relevant										
15	2021	Full conference book										
15	2021	Full conference book										
17	2021	ECG, BVP, EDA, RESP, TEM	Applicable	Dimensional	Valence, arousal	N/A		76.37%/70.29% (data	74.03%/68.15%		30, 20 (between two datasets	
18	2021	Review										
19	2021	ECG	N/A	Dimensional	Valence	N/A	KNN	TD:84.7%/ASD:81%			Children, autistic (6) and reg. dev. (6)	
20	2020	PPG, EDA									22, older adults	
21	2020	Review										Comparing a lot of devices/sensors
22	2020	RFID	Applicable	Discrete	Anger, joy, sadness, pleasu	Obstructive, positive, negative, positive	RF signal filtering			83.30%	1 male, 1 female	
23	2020	Not relevant										
24	2020	ECG, PPG	Applicable	Discrete	Anger	Obstructive	freq analysis			70.48%	7 males, 7 females, 2 fell out	
25	2020	GSR, PPG	Applicable	Discrete	Relaxation, stress	Conductive, obstructive	Linear & fine Gaussian SVM			Dependant on trail: 97.9%,89.9%, 96.9%,89.2%	9 males, 9 females	
26	2020						Multimodal deep learning					Experiment comparing 10 NN methods, useful
27	2020											Review paper
28	2020											Physiological signals dataset
29	2020	Not relevant										
30	2020	GSR , FEA	N/A	Dimensional	Arousal, valence		basic signal analysis	57%	96%		6 adults	
31	2020	Not relevant										
32	2020	Not relevant										
33	2020	EDA	N/A								2 males, 2 females	
34	2020											Experiment comparing neural networks with the DEAP dataset
35	2020	HR, EDA, EEG									11 participants	Not relevant, no link to emotions
36	2020	Review paper										
37	2020	PPG, EDA, TEMP	Applicable	Discrete	Behaviours+Anger,excited,h	Obstructive, high power /control, positive, low power/control	Logistic Regression (LR), Support Vector Machine (SVM), and Decision Trees (DT)			LRSVM68%,DT63%, subj. dep model 85%	5 Children 8-12	
38	2020	EEG	Not applicable	Discrete	Anxiety	low power/control	Wavelet Db4				20 participants aged 20-25	
39	2020	Not relevant										
40	2020	IR camera, proximity sensor, IMU	Applicable	Discrete	Same as [3]		Optical-flow based algorithm	49.32%	47.26%	83.87%		
41	2020	ECG, GSR, SP2O, TEMP, Gl	Applicable	Discrete	Stress	Obstructive	Leverberg–Marquardt NN			90.50%	30 participants	
42	2020	PPG, FTT, Resp. Belt	Applicable	Dimensional	Valence, arousal	n/a	Several ML algorithms	69.53%-72.18%	69.86%-73.08%		32, DEAP dataset	Also reports non-fused performance
43	2020	review paper										Quite useful, focuses on fusion/extraction/procession and sensor types; not performance
44	2020	EMG, HR	Applicable	Discrete	"incident"	n/a	Gaussian naive Bayes classifier			0.48	3 male students, twenties	
45	2020	EEG, ECG, PPG	Applicable	Discrete	happy, angry, sad	positive, obstructive, low power/control	CNN			EEG76.94%, ECG/PPG76.8%	20 participants	Also reports non-fused performance

46	2020	HR, phone	Applicable	Discrete	anger, disgust, fear, happiness, sadness, surprise, neutral		CNN				Not provided	Questionable relevance, no reported method or accuracy of proposed system
47	2020	EEG, BVP	Applicable	Discrete	Two valence and two arous	n/a	CNN LSTM			Early fusion 71.61% late fusion70.17%,	20 subjects	Compares early and late fusion approaches
48	2020											
49	2020	EDA, ECG	Applicable	Dimensional	Valence, arousal	n/a	qualitative, quantitative, machine learning and fuzzy logic approaches	Dependent on methoc	Dependent on method		70 participants	good comparison of algorithms, not specific in measuring method
50	2020											Could not access
51	2020	not relevant										Comparison of EEG methods, small link to emotions but no mention of discrete/dimensional, specific accuracies
52	2020	(EEG or GSR) in combination	Applicable	Dimensional	Valence, arousal	n/a	basic signal analysis	GSR87%, EEG53%	Both >73%		5	List agreement, precision, recall as accuracy. Grabbed agreement as accuracy. Good comparison study showcasing the potential of GSR in arousal measurement.
53	2020	HR, TEMP, GSR, BR	Applicable	Discrete	Stress	Obstructive	Convolutional-Autoencoders and Recurrent Neural Networks			53.30%	28	Findings indicate wideangle camera is more accurate in detecting stress from participants
54	2020	Glucometer, BP, HR, GSR	n/a	Discrete	Stress	Obstructive	basic signal analysis			No reported accuracy	10	Does not report accuracy
55	2020	not relevant										Can be interesting for reflection
56	2020	not relevant										Denotes challenges in wireless transfer of physiological signals, can be useful in certain situations although well established WiFi communication between sensor nodes and computers already exist. Signal loss.
57	2020	Full conference										
58	2020	EDA, fNIRS	Applicable	Dimensional	Valence, arousal	n/a	time/freq series analysis w/ statistical analysis to determine result				24	No reported accuracy. Proof of concept for EDA arousal measurement potential.
59	2019	About dogs, not humans										
60	2019	PPG, GSR	Applicable	Discrete	Fear	Active/aroused	Comparison of different methods				32 (DEAP dataset)	paper proposes a fear recognition method with a proof of concept using the DEAP dataset. Limited to two sensors and presenting results for several algorithms.
61	2019	not relevant										
62	2019	PPG	n/a	Dimensional	Valence, arousal, dominanc	n/a	CNN	61.90%	60.10%		22 out of 32 participants, DEAP dataset	Also measures dominance (DAP framework) but does not discuss it in the discussion. Provides a good description of how to extract PPG measurement and how they can be used.
63	2019	not relevant, FEA										Paper discusses ER as a whole firstly. Provides three experiment performances relating to three different processing methods.
64	2019	not relevant										Sensor fusion paper. Can be useful to explain the specific topic but no linkage to emotion recognition
65	2019	Pupil diameter, GSR	Applicable	Dimensional	Valence, arousal	n/a	a correlation-based feature extraction algorithm (CorrFeat)	89.22%	73.12%		27 subjects mahnob database	Divide valence/arousal into #class-classification (SAM)
66	2019	ECG	n/a	Discrete	Angry, happy, sad, calm		Only measurements					No recognition, only ECG measurements in different emotional states
67	2019	Not relevant										Interesting link to appraisal theory. Uses BVP, EDA, TEMP but only to determine baselines. No detection going on.
68	2019	Not relevant										Specifies use of non-wearable sensors in abstract.
69	2019	Not relevant										Related to activity daily living (ADL) reocgnition using ML
70	2019	PPG based PTT										Describes the development of a PPG based PTT measurement method. No accuracy ratings.
71	2019	HR, GSR, TEMP	Applicable	Dimensional	Valence 5 step SAM	n/a	CNN-LSTM	95%, 87.3% for physiological only			10 (from dataset)	
72	2019	ECG, BVP, RSP, SKT, EDA	Applicable	Discrete	Stress	Obstructive	Multi modal machine learning			84.13%	30 participants aged 25-30	
73	2019	GSR, TEMP, HR	Applicable	Discrete	happy, depressed, stressed	positive, low power/control, obstructive, conductive	SVM			79%	22 participants	
74	2019	ECG, GSR	Applicable	Dimensional	3 step valence/arousal	n/a	Gradient boosting	34%	34%		132	Accuracy improved in context aware situations (game)
75	2019	FEA	n/a	Discrete	Happy, Sad, Surprised, Fea	positive, low power/control, active/aroused, obstructive,	pre-trained classifier face point recognition				8 (children, ASD)	Does not report on accuracy. Surprised = asthonsihed, neutral = calm/at ease
76	2019	HRV	n/a	n/a	n/a	n/a	complex demodulation				8	A study regarding HRV monotoring of workers, does not relate this to emotion detection or prediction
77	2019	GSR, BVP	Applicable	Dimensional	valence, arousal	n/a	random forest	85%	85%		not reported	
78	2019	full conference										
79	2019	review paper										
80	2019	full conference										
81	2019	EMG, ECG, EMA	Applicable	Discrete	happy, relaxed, disgust, sad	positive,conductive, obstructive, low power/control	CNN			87.50%	32 (DEAP dataset)	
82	2019	EEG	n/a	Discrete	General event logging	n/a	Fuzzy C-Means clustering			83%	Not reported, uses DEAp dataset (32 participants) to compare against	Sadly no reported amount of participants. Compares results against various datasets.

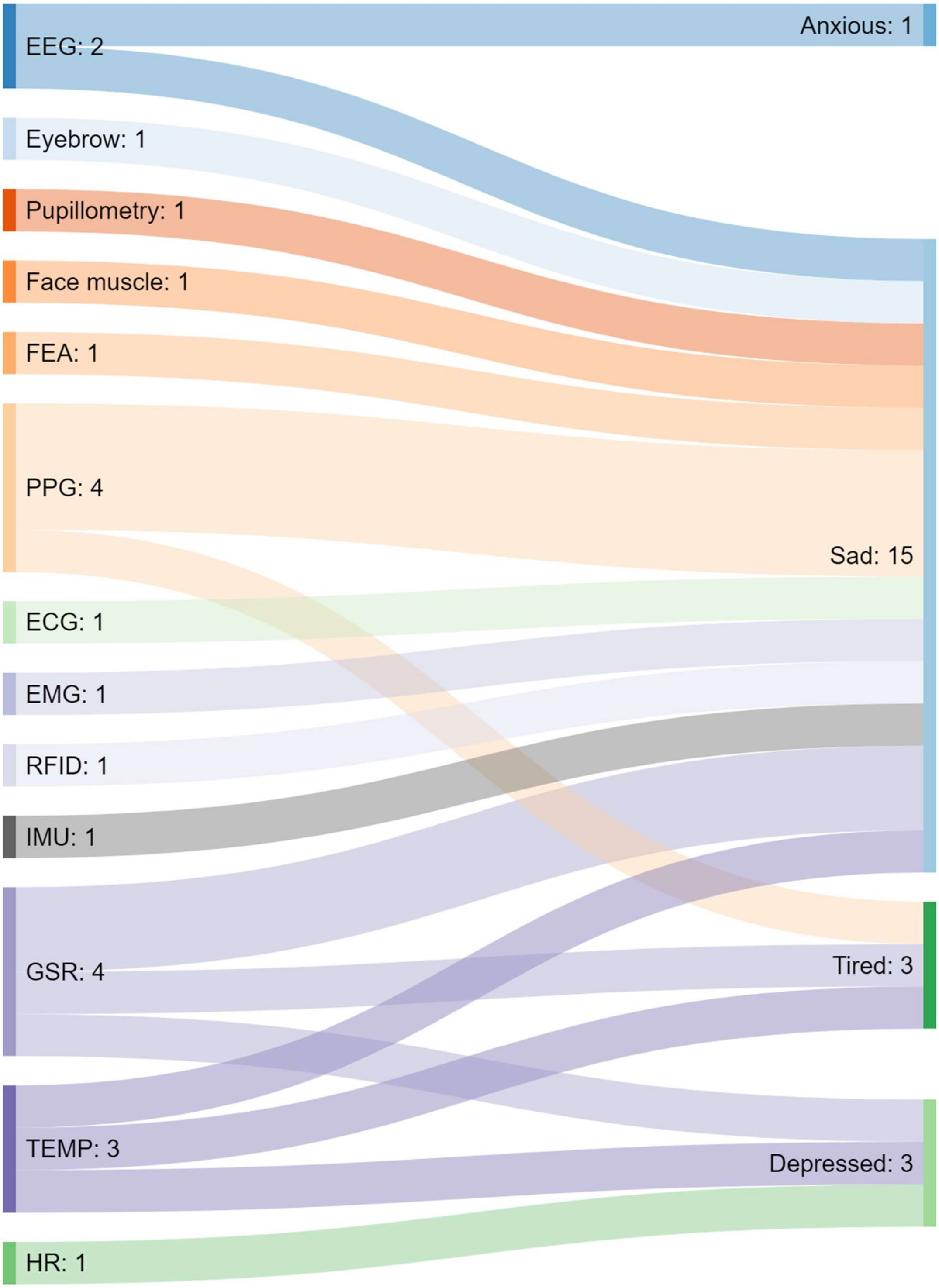
83	2019	review paper										

Appendix C

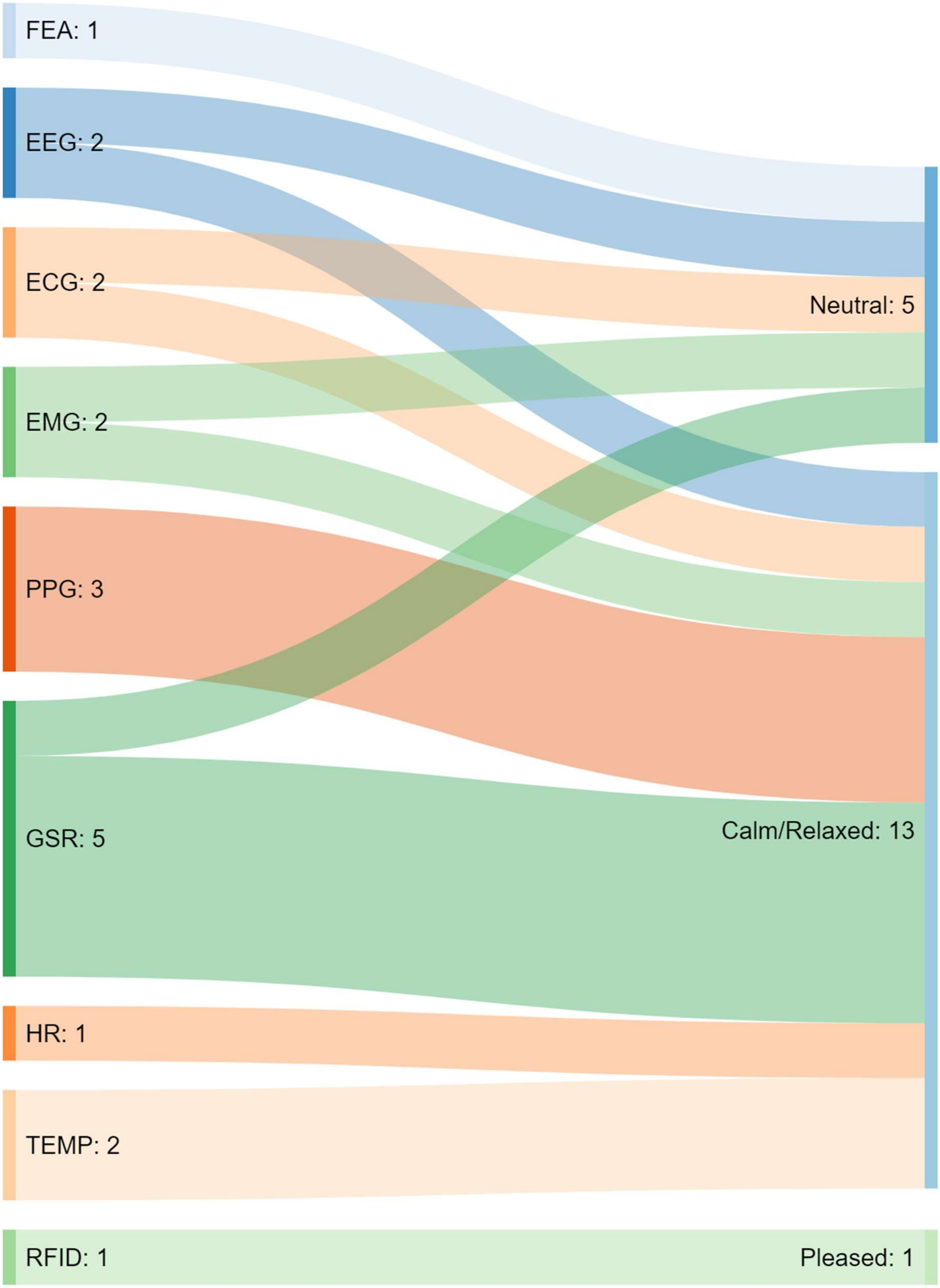
Miro board with post-its for instances of emotion measurement. Basis of the scatter plot.



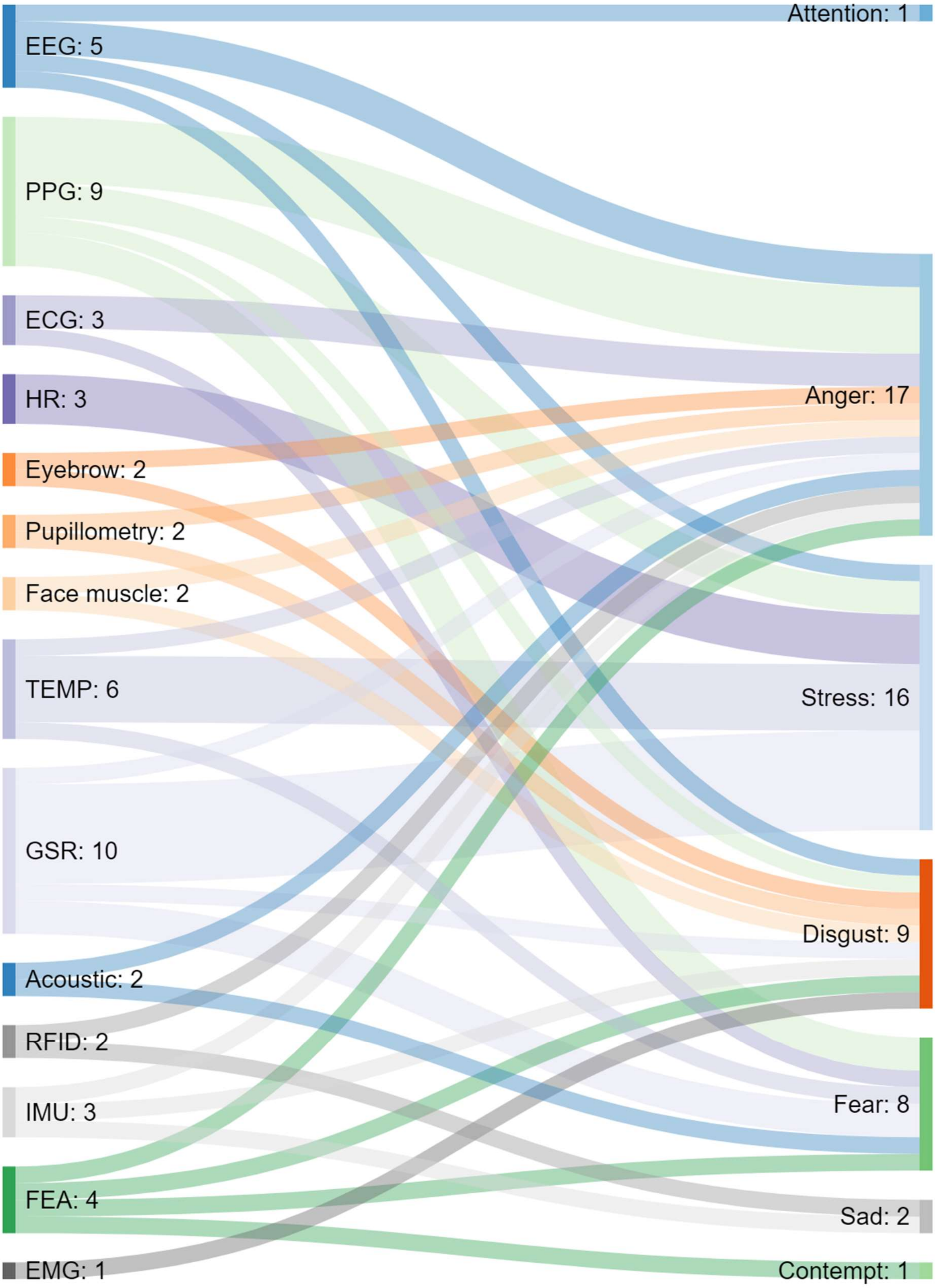
Appendix D



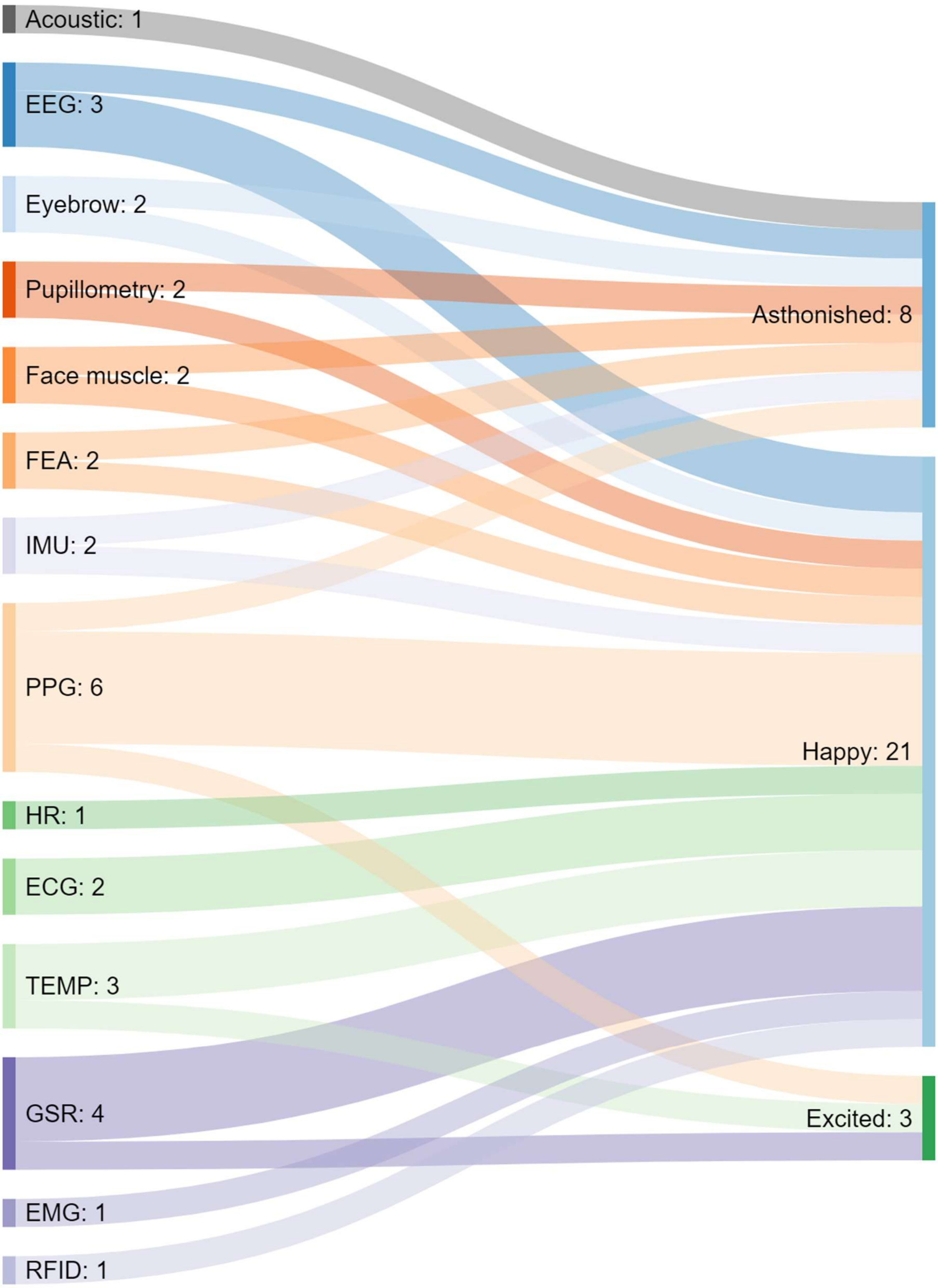
Sankey diagram showcasing the negative valence/negative arousal emotions and the sensors used to measure them.



Sankey diagram showcasing the negative valence/positive arousal emotions and the sensors used to measure them.



Sankey diagram showcasing the positive valence/negative arousal emotions and the sensors used to measure them.



Sankey diagram showcasing the positive valence/positive arousal emotions and the sensors used to measure the

Appendix E

Table 1: the included papers with dots for the sensors they use

#	Paper:	Attention	Afraid/fea	Angry	Stress	Contempt	Disgust	Surprised	Excited	Happy	Pleased	Neutral	Calm/Rela	Tired	Anxious	Depressed	Sad	Valence	Arousal	Incident	Dominance	Event logging
2	Affective	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	■	-	-	-	-
3	SPIDERS: /	-	-	■	-	-	■	■	-	■	-	-	-	-	-	-	■	■	■	-	-	-
4	Reinforce	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Emotion b	-	■	-	-	-	-	-	-	■	-	-	■	-	-	-	■	-	-	-	-	-
8	Fear recog	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	and	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-
11	Cnn-based	-	■	■	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	A Real-Tim	■	-	-	■	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-
17	Cormet: Fi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
19	on of	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-
22	Emotion m	-	-	■	-	-	-	-	-	■	■	-	-	-	-	-	■	-	-	-	-	-
24	Personal ir	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	Discrimina	-	-	-	■	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-
30	Measuring	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-
37	Physiologi	-	-	■	-	-	-	-	■	■	-	-	■	■	-	-	■	-	-	-	-	-
38	Emotion	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-
40	SPIDERS:	-	-	■	-	-	■	■	-	■	-	-	-	-	-	-	■	■	■	-	-	-
41	sensor-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
42	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
44	Incident de	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-
45	An AI-edge	-	-	■	-	-	-	-	-	■	-	-	-	-	-	-	■	-	-	-	-	-
47	Automatic	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
49	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
52	Assessing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
53	A deep lea	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
54	A framewo	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
58	Using fnirs	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
60	Toward Fe	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
62	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	■	-
65	CorrFeat: (	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
71	Deep learn	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-
72	Multi-Moda	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
73	SEmoD: A	-	-	-	■	-	-	-	-	■	-	-	■	-	-	■	-	-	-	-	-	-
74	Analysis a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
75	Guess Wh	-	■	■	-	■	■	■	-	■	-	■	-	-	-	-	■	-	-	-	-	-
77	A machine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
81	Healthcare	-	-	-	-	-	■	-	-	■	-	■	■	-	-	-	■	-	-	-	-	-
82	Memento:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■

Table 2: the included papers with dots for the sensors they use

#	Paper:	EEG	fNIRS	Eyebrow	Pupillome	Face musc	Saccadic	FEA	PPG	FTT	BP	SP2O	BVP	Glucose	HR	ECG	RESP	TEMP	GSR	EMG	RFID	IMU	Typing	Acoustic
2	Affective	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	SPIDERS: A	■	-	■	■	■	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	Reinforce	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Emotion b	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	Fear recog	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	■	■	-	-	-	-	-
10	and	-	-	-	-	-	-	■	■	-	-	-	■	-	-	-	-	-	■	-	-	-	■	■
11	Cnn-based	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■
13	A Real-Tim	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	Cornet: Fi	-	-	-	■	-	■	-	-	-	-	-	■	-	■	■	■	■	■	■	-	-	-	-
19	on of	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-
22	Emotion m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-
24	Personal ir	-	-	-	-	-	-	-	■	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-
25	Discrimina	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-
30	Measuring	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-
37	Physiologi	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	■	■	-	-	-	-	-
38	Emotion	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	SPIDERS:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-
41	sensor-	-	-	-	-	-	-	-	-	-	■	■	-	■	-	■	-	■	■	-	-	-	-	-
42	Emotion R	-	-	-	-	-	-	-	■	■	-	-	-	-	-	-	■	-	-	-	-	-	-	-
44	Incident de	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	■	-	-	-	-
45	An AI-edge	■	-	-	-	-	-	-	■	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-
47	Automatic	■	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-
49	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	-	-	-	-	-
52	Assessing	■	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-
53	A deep lea	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	■	■	■	-	-	-	-	-
54	A framewo	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	■	-	-	-	-	-
58	Using fnirs	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-
60	Toward Fe	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-
62	Emotion R	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
65	CorrFeat: (	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-
71	Deep learn	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	-	-
72	Multi-Moda	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	■	■	-	-	-	-	-
73	SEmoD: A	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	-	-
74	Analysis a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	-	-	-	-	-
75	Guess Wh	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
77	A machine	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	■	-	-	-	-	-
81	Healthcare	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	-
82	Memento:	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 3: a combination of previous charts

#	Paper:	EEG	fNIRS	Eyebrow	Pupillome	Face musc	Saccadic	FEA	PPG	FTT	BP	SP2O	BVP	Glucose	HR	ECG	RESP	TEMP	GSR	EMG	RFID	IMU	Typing	Acoustic	Attention
2	Affective	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	SPIDERS: /	■	-	■	■	■	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	Reinforce	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Emotion b	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	Fear recog	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	■	■	-	-	-	-	-	-
10	and	-	-	-	-	-	-	■	■	-	-	-	■	-	-	-	-	-	■	-	-	-	■	■	-
11	Cnn-based	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-
13	A Real-Tim	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■
17	Cornet: Fi	-	-	-	■	-	■	-	-	-	-	-	■	-	■	■	■	■	■	■	-	-	-	-	-
19	on of	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-
22	Emotion m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-
24	Personal ir	-	-	-	-	-	-	-	■	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-
25	Discrimina	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-
30	Measuring	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-
37	Physiologi	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	■	■	-	-	-	-	-	-
38	Emotion	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	SPIDERS:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-
41	sensor-	-	-	-	-	-	-	-	-	-	■	■	-	■	-	■	-	■	■	-	-	-	-	-	-
42	Emotion R	-	-	-	-	-	-	-	■	■	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-
44	Incident de	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	■	-	-	-	-	-
45	An AI-edge	■	-	-	-	-	-	-	■	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-
47	Automatic	■	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-
49	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	-	-	-	-	-	-
52	Assessing	■	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-
53	A deep lea	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	■	■	■	-	-	-	-	-	-
54	A framewo	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	■	-	-	-	-	-	-
58	Using fnirs	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-
60	Toward Fe	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-
62	Emotion R	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
65	CorrFeat: (	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-
71	Deep learn	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	-	-	-
72	Multi-Moda	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	■	■	-	-	-	-	-	-
73	SEmoD: A	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	-	-	-
74	Analysis a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	-	-	-	-	-	-
75	Guess Wh	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
77	A machine	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	■	-	-	-	-	-	-
81	Healthcare	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	■	■	-	-	-	-	-
82	Memento:	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

#	Paper:	Afraid/fea	Angry	Stress	Contempt	Disgust	Surprised	Excited	Happy	Pleased	Neutral	Calm/Rela	Tired	Anxious	Depresse	Sad	Valence	Arousal	Incident	Dominanc	Event logg
2	Affective	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	■	-	-	-	-
3	SPIDERS: /	-	■	-	-	■	■	-	■	-	-	-	-	-	-	■	■	■	-	-	-
4	Reinforce	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Emotion b	■	-	-	-	-	-	-	■	-	-	■	-	-	-	■	-	-	-	-	-
8	Fear recog	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	and	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-
11	Cnn-based	■	■	-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	A Real-Tim	-	-	■	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-
17	Cornet: Fi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
19	on of	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-
22	Emotion m	-	■	-	-	-	-	-	■	■	-	-	-	-	-	■	-	-	-	-	-
24	Personal ir	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	Discrimina	-	-	■	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-	-	-
30	Measuring	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-
37	Physiologi	-	■	-	-	-	-	■	■	-	-	■	■	-	-	■	-	-	-	-	-
38	Emotion	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-	-	-	-
40	SPIDERS:	-	■	-	-	■	■	-	■	-	-	-	-	-	-	■	■	■	-	-	-
41	sensor-	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
42	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
44	Incident de	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-
45	An AI-edge	-	■	-	-	-	-	-	■	-	-	-	-	-	-	■	-	-	-	-	-
47	Automatic	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
49	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
52	Assessing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
53	A deep lea	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
54	A framewo	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
58	Using fnirs	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
60	Toward Fe	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
62	Emotion R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	■	-
65	CorrFeat: (	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
71	Deep learn	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	-	-	-	-
72	Multi-Moda	-	-	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
73	SEmoD: A	-	-	■	-	-	-	-	■	-	-	■	-	-	■	-	-	-	-	-	-
74	Analysis a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
75	Guess Wh	■	■	-	■	■	■	-	■	-	■	-	-	-	-	■	-	-	-	-	-
77	A machine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■	■	-	-	-
81	Healthcare	-	-	-	-	■	-	-	■	-	■	■	-	-	-	■	-	-	-	-	-
82	Memento:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	■

Table 4: a table showing how many times each sensor was used to measure each emotion

Sensor	Attention	Afraid/fear	Angry	Stress	Contempt	Disgust	Aroused/asthonised	Excited	Happy	Pleased	Neutral	Calm/Relaxed	Tired	Anxious	Depressed	Sad	Valence	Arousal	Incident
EEG	1		2	1		1	1		2		1	1		1		1	4	3	
PPG		2	4	2		1	1	1	4			3	1			3	4	3	
Eyebrow			1			1	1		1							1	1	1	
Pupillometry			1			1	1		1							1	3	3	
Face muscle			1			1	1		1							1	1	1	
ECG		1	2	2		1			2		1	1				1	4	3	
TEMP		1	1	4				1	2			2	1		1	1	2	1	
GSR		2	1	6		1		1	3		1	4	1		1	2	9	8	
BVP				1													4	3	
RESP				2													2	2	
EMG						1			1		1	1				1	1	1	1
HR				3					1			1			1		2	1	1
Saccadic																	1	1	
RFID			1						1	1						1			
IMU			1			1	1		1							1	1	1	
SP2O				1															
Glucose				2															
BP				2															
FTT																	1	1	
FEA		1	1		1	1	1		1		1					1	2	1	
fNIRS																	1	1	
Typing																	1		
Acoustic		1	1				1										1		
	1	8	17	26	1	10	8	3	21	1	5	13	3	1	3	15	45	35	2

Appendix F

Graduation project reflection report
“Towards a taxonomy of wearable emotion recognition for user experience research”

Marc Fuentes Bongenaar S2258722
Creative Technology
Reflection II
Alexandria Poole

Table of Contents

Project description and vision

User experience (UX) research is an important part of the Creative Technology design process; the experience of the user while using a website, product, or even observing a data physicalization installation is something that is considered during the design process. The user experience can be measured in several quantitative and qualitative ways. In qualitative research, we might ask a user if they felt the new interface on their computer felt more efficient, and if made then feel happy. A general focus of the qualitative research in UX is towards the affect of the user, how did they feel during the use of the product. On the quantitative side are the measurements of the objective truths of the new design; the user might be faster using the new design, generate better output, or need fewer overall interactions with the interface. The problem of these measurements is that they do not consider the affect of the user during the user test. This can lead to poor design choices, for example, if you are creating an interface that is to be used daily in a workplace you are most likely looking for a design that optimizes the efficiency of the worker. A design might come forward that does exactly that, the efficiency is increased 10% over the last design, a great accomplishment. A problem with that design might be that during the user tests the user felt incredibly bored or anxious because of the interface. If this design was to be implemented in the workplace, the employer might be seeing more turnover or less efficiency over time because of the design that while being very efficient, did not consider the user experience.

While the previous anecdote demonstrates the importance of UX design well, it also already includes a solution: combine qualitative and quantitative measurements. While this is a solution that can work well in a lot of cases, it is not applicable to all and might not be the ideal solution overall. A problem with the qualitative measures is that they rely heavily on self-reporting and observations. Observations can be wrong and very limited to the extend of the actual affect of the user. Self-reporting is not always a wise choice because a participant of a user test might not remember their emotional state during the whole test and report them wrong. Further, self-reporting is not always possible for all target groups, for example children or people with autism spectrum disorder (ASD) [1][2]. A quantitative measurement of affect can offer a solution to these problems.

This is the focus of my graduation project, the creation of a taxonomy of wearable emotion recognition for UX research. The goal of my research is to read all the published work of the last few years relating to wearable emotion recognition and to determine what sensors, emotions, and algorithms can be used in what context, and what can be the best solution for UX research. This project lays the foundation of a future CreaTe graduation project that will use my research to create a wearable emotion recognition device for UX research. I hope that in the future my work, or at least the topic it pertains, can help those that can not voice or express their emotions in conventional ways, and can help guide design processes to create more user friendly and accessible products.

Ethical analysis

Ethical dilemmas

Four ethical dilemmas have been identified and are described below. Then in the next section a code of ethics is presented and finally, the ethical dilemmas are analyzed using the code of ethics and tools such as the Fleddermann diagram.

1. To achieve wearable emotion recognition different types of sensors are used. Most of these are biosensors, measuring anything from heartrate to galvanic skin response. The user provides a lot of personal biometric data for the device to work. This is sensitive information that the user might not want to be public. Furthermore, the user might be suffering from certain conditions (known or unknown) that influence the working of the system. Through a session, a user might be confronted with the fact that they have irregular heart rates; something they did not "sign up" for to know. These two aspects pose quite a conundrum for the use of emotion recognition in UX research because a lot of it is focused on children or people with disabilities that do not allow them to express emotions in surveys or interviews [3]. While it is great that this technology might help in the creation of interfaces, devices, and tools that will be more pleasant and accessible for that target group, it is also the target group that is generally not able to provide informed consent.

2. Because literature research is at the core of this graduation project the ethics relating to that are something to be considered too. This naturally relates to plagiarism but also extends further into the taxonomy. Everything that is recommended should be considered well and no claims that can not be substantiated should be made, as it is not ethical and can lead to problems in derivative work.

3. Another ethical dilemma is of a more philosophical nature. Is it possible? When discussing emotion recognition terms like valence and arousal are commonly mentioned. These are not thing we regularly discuss when talking about emotions. Emotions might be too abstract to fully grasp or measure with biometrics. This could lead to products designed for enticing specific emotions in the user, even though that is not the case. It could also become a very biased system, where gender, age, and race influence the biometrics in such a way that a system is not inclusive at all. This topic is touched upon in a large survey conducted by Law and his colleagues in "Attitudes towards user experience UX measurement" [4] where they surveyed many UX researchers about UX measurement. Emotions was a measurement that was included, and a lot of researchers indicated that they did not trust it to be able to correctly estimate human affect. While this survey has been conducted 8 years ago already and the availability of methods and amount of research has increased, they still post valid concern. Biases in the accuracy of emotion recognition systems are already visible in the current research. Generally, systems have a low subject-independent accuracy in research where there was a diverse group of participants in the experiments. Other systems that have a homogenous test group report higher accuracies up to 97% for detection of certain emotions [5].

4. Lastly is a dilemma I considered while writing the anecdote at the beginning of this report. It goes a bit beyond the scope of my graduation project, but I thought it would be a good thing to mention. When an accurate emotion detection system is developed and adopted into

general design processes, can we now “require” good user experience design. This is especially pertaining to workplaces where the situation described in the first chapter might occur at some point. At such a moment where we can quantifiably say that a new design for work software will actively make an employee more sad, anxious, tired, bored, etc. do we now limit the amount of “bad experience” an employer is allowed to incorporate into their work software for the sake of efficiency? Is it the employer’s freedom to create whatever software they desire? Or is it the employer’s responsibility to not actively worsen the mental state of their employees?

Code of ethics

This code of ethics was made for this specific graduation project. It takes into consideration the limitations of the graduation project (e.g. the taxonomy can merely inform and advise eventual experimental design). Inspiration was taken from the IEEE and NVIDIA code of ethics [6][7].

Privacy
- The privacy of test subject and their data will be always respected. This means that the data should not be stored in an unsafe place and should be destroyed when it is no longer needed.

Honesty
- The research will be conducted in an honest way. Work not created by me will be properly sourced. Limitations to the research will be clearly stated and will form the basis of recommendations of future research.

Inclusivity
- The resulting taxonomy must be clear in the applicability of the proposed methods regarding to age, gender, and race.

Clarity
- The results proposed methods produce will be properly assessed if they actually relate well to natural human emotion. If there is no substantiated claim to make, the results will be presented in a more abstract way that presents itself as merely an approximation of a complex human experience.

Accountability
- The creator of the project will be held accountable for all statements in the project report.

Analysis of dilemmas

The four ethical dilemmas presented earlier in the chapter will be reviewed according to the code of ethics for this graduation project. Where necessary tools will be utilized to consider the dilemmas.

During this graduation project no experiment involving the gathering of data will occur. Datasets do form the basis of the taxonomy, so the privacy of the participants of the studies that my work is derived from should also be considered. It is important that those studies handle their user data in an ethical way. If this is not the case, the study shall not be considered in the taxonomy. This is also to promote the importance of this topic in this (relatively) fresh field. By selecting papers on good practice, I can still try to help in promoting good handling of data even though no unique data is gathered during my graduation project. Because some of the papers are specifically about systems targeting children or people with disabilities the way informed consent is handled will also be looked at in those cases. This is because the proper gathering of data is of paramount importance in this emerging field. These considerations can be found in the taxonomy addition flowchart which can be found in appendix A [8].

The second dilemma is regarding ethical research. Many guidelines exist for this already and proper training in this has been provided by the university. Still, honesty is an important part of the code of ethics for this graduation project and should still be considered even when following these guidelines correctly. Not everything can be ethically assessed in its entirety with these guidelines so upholding this code is still important. All claims made should be substantiated by research with accompanying limitations or be noted as subjective thought. This is especially important for the sections relating to the specific translation from measurement to actual emotion. In these sections the nuance and division in the field regarding this topic should be made clear. This is included in the taxonomy addition flowchart in appendix A.

As stated in the previous paragraph, the limitations of emotion measurements should be made clear, this is included in the clarity section in the code of ethics. In addition, the taxonomy needs to address the diversity or homogeneity of the participants of the experiments conducted in the research supporting the taxonomy. Homogeneity drastically improves the performance of emotion measurements systems and this needs to be considered in the way results are put into words in the taxonomy [5]. The limitations of the research used should be clear to the reader and this is included in the taxonomy addition flowchart in appendix A.

The last dilemma is though because the taxonomy has little influence on it. At this moment the technology is not there yet, and the usability of emotion recognition is still debated. Still, it is an interesting dilemma. When we have to ability to measure the experience of the worker, should we set expectations as to how the employers acts upon those measurements? If we do require employers to act on this, the employer is most likely to forego such a UX test to be able to focus purely on improving efficiency. Is it then better to require those tests to employers that want to make a new program to be used in a workspace? In that case, we take away a large part of the autonomy of the employer. If we do not require action to be taken, it can lead to cases where the employers actively choose for a system that has an objectively bad user experience for the sake of efficiency. Is it ethical for those employers to make that decision? Do they have a larger responsibility to their employees or to their investors? To put some possible situations into perspective I put them in a line drawing (figure 1). This line drawing is from the perspective that optimal employee experience is most ethical (positive paradigm) and that bad employee experience is the least ethical (negative paradigm) [8].

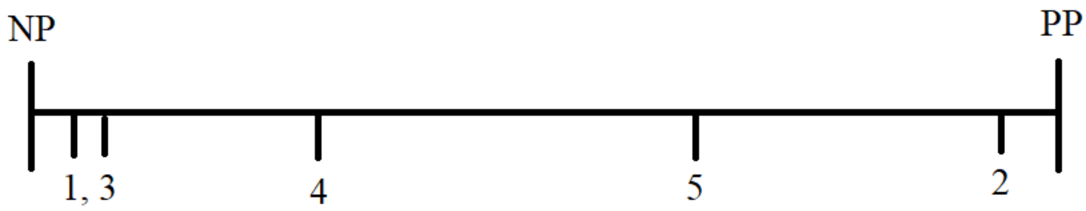


Figure 1: line drawing

Situations:
1. Employer does not want to measure the affect of employees
2. Employer measures affect of the employees and acts to improve it
3. Employer measures affect of the employees and does not act to improve it
4. Employer wants to measure affect of the employees but doesn't (e.g., lack of resources)
5. Employer actively tries to improve employee affect but does not measure it quantitively

An interesting idea voiced by a good friend of mine that I discussed this dilemma with was the idea of implementing a position of affect supervisors into company structures that keep an eye on and maintain a balance between efficiency and positive affect in the employees. Of course, this whole topic is quite nuanced because there is a case to be made for positive affect being more efficient in the long term and there are a lot more other things that have not been considered yet, it is just an interesting dilemma I encountered during the writing of this reflection.

Conclusion

My research does not propose new methods of emotion recognition but merely summarizes what is there already and tries to propose the most optimal way for people to model their designs in an informed manner. I cannot control the way the user tests of these designs happen because I am not the one conducting them. I can however try to voice the dilemmas I think prospective affect recognition researchers should consider and select the research I base my taxonomy on in an ethical way. I can not see the future of this technology yet, emotion recognition is mostly a gimmick implemented in some smart watches, but it has the potential to grow into something larger over time. My influence on the field is very limited and I am aware of that. I hope that the little influence I do have is a positive one.

References

[1] R. W. Robins, R Chris Fraley, and R. F. Krueger, *Handbook of research methods in personality psychology*. New York ; London: Guilford, 2010.

[2] G. T. Fujiura, “Self-Reported Health of People with Intellectual Disability,” *Intellectual and Developmental Disabilities*, vol. 50, no. 4, pp. 352–369, Aug. 2012, doi: 10.1352/1934-9556-50.4.352.

[3] A. Bagirathan, J. Selvaraj, A. Gurusamy, and H. Das, “Recognition of positive and negative valence states in children with autism spectrum disorder (ASD) using discrete wavelet transform (DWT) analysis of electrocardiogram signals (ECG),” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 405–416, May 2020, doi: 10.1007/s12652-020-01985-1.

[4] E. L.-C. Law, P. van Schaik, and V. Roto, “Attitudes towards user experience (UX) measurement,” *International Journal of Human-Computer Studies*, vol. 72, no. 6, pp. 526–541, Jun. 2014, doi: 10.1016/j.ijhcs.2013.09.006.

[5] J. A. Miranda, M. F. Canabal, L. Gutiérrez-Martín, J. M. Lanza-Gutierrez, M. Portela-García, and C. López-Ongil, “Fear Recognition for Women Using a Reduced Set of Physiological Signals,” *Sensors*, vol. 21, no. 5, p. 1587, Feb. 2021, doi: 10.3390/s21051587.

[6] “IEEE Code of Ethics,” *@IEEEorg*, 2022. https://www-ieee-org.ezproxy2.utwente.nl/about/corporate/governance/p7-8.html (accessed Jan. 09, 2022)

[7] J. Huang, “NVIDIA CORPORATION WORLDWIDE CODE OF ETHICS.” [Online]. Available: http://media.corporate-ir.net/media_files/nsd/nvda/corpgov/worldwide_codeofethics.pdf.

[8] C. B. Fleddermann, *Engineering ethics*. Upper Saddle River, N.J.: Prentice Hall, 2012.

Appendix A

Taxonomy addition flowchart (chart 1). Created using LucidChart.

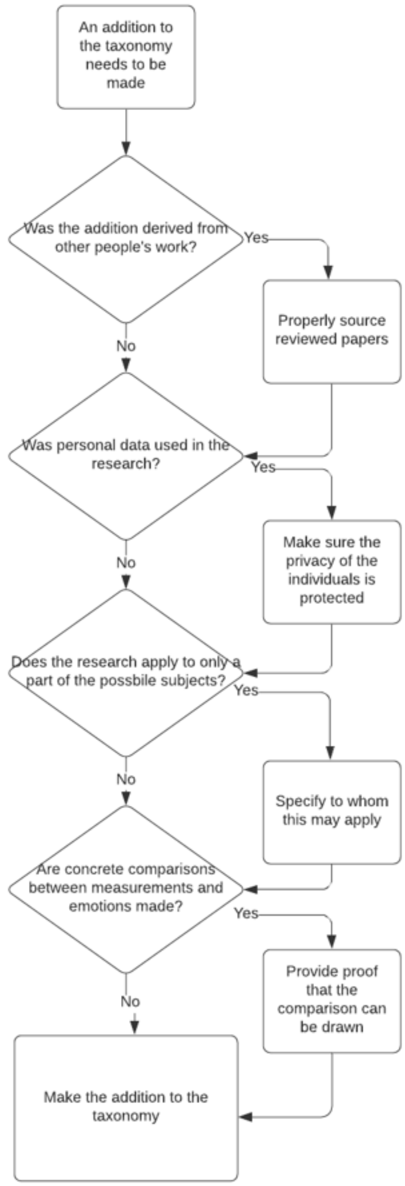


Chart 1: taxonomy addition flowchart