Modality face off

Detecting affect by analyzing electroencephalography and facial expressions

Svevo de Weert
ANR 778404

Master's thesis
Communication and Information Sciences
Specialization Human Aspects of Information Technology
Faculty of Humanities
Tilburg University, Tilburg

Supervisor: prof. dr. M.M. Louwerse
Second reader: M. Alimardani, Ph.D.

December 2017

Abstract

In this thesis, emotion recognition performance through encephalogram (EEG) analysis is compared to the more established performance of facial expression analysis. An experiment was conducted to assess recognition performance of five discrete affective states (fear, amusement, disgust, sadness, and neutral). Participants were shown emotion-eliciting movie fragments while their EEG and facial expressions were recorded. The EEG modality was able to detect the emotions only to a limited extent and significantly underperformed compared to the face modality. The combination of modalities, however, significantly outperformed both the individual modalities. These findings (1) contribute to our knowledge of the added value of EEG to facial expression recognition and (2) encourage future research to take a multimodal approach.

Table of contents

Introduction

Affective computing is computing that relates to, arises from, or influences emotion (Picard, 1997). Emotions affect many aspects of our lives such as decision-making, reasoning and well-being (Matiko, Beeby, & Tudor, 2014). Moreover, they play an important role in human–human communication and interaction (Wang, Nie, & Lu, 2014). It is theorized that computation that understands the affective context of users, and are able to express affect, could assist humans in areas such as communication, entertainment, health, and learning (Picard, 1997). These include digital learning systems that detect a student's frustration and adapt the study material; games that can recognize boredom and change the difficulty level to retain the player's engagement; or conversational customer service bots that change their tone of voice to accommodate customer needs. In these three scenarios, affective computing has been able to recognize a person's state of affect.

Assessment of emotions has attracted the attention of researchers from different fields such as psychology, education, gaming, and healthcare, who have addressed detection of emotion using individuals' heart rate, skin conductance, pupil dilation, tone of voice, facial expression and electroencephalogram (EEG) with varying accuracy (Matiko et al., 2014). Valstar, Mehu, Bihan Jiang, Pantic, and Scherer (2012) argued that inferring six basic discrete emotions from facial movements of known subjects can be considered largely solved, with seven out of ten participating systems in a facial expression recognition challenge achieving classification accuracies of over 80%, of which three systems achieved over 90% accuracy rates. However, despite the impressive results, these findings are limited to specific settings and populations, with limited opportunities for generalization, such as generalization to non-frontal head poses, and providing good performance across a broader range of ethnicities (Littlewort, Whitehill, Wu, Fasel, Frank, Movellan, & Bartlett, 2011). It may therefore be worthwhile to compare, or perhaps augment, existing techniques of measuring affect using facial expressions with non-facial expression measures, such as the use of EEG.

EEG is used to recognize emotions by analyzing the electrophysiological signals that reflect central nervous system (CNS) activity. Measurements recorded over various parts of the brain including the amygdala potentially enable observation of the emotions experienced (Gunes & Schuller, 2013). Despite the amygdala's role and importance in the brain activities due to emotions, the amygdala's deep location in the brain poses a challenge for obtaining valuable information via the EEG measurement (Gunes & Schuller, 2013). Nonetheless, some studies have suggested the possibility to use EEG data and machine learning to classify individual's emotions (Bos, 2006; Hosseini & Naghibi-Sistani, 2011; Sourina & Liu, 2011; Wang et al., 2014). Using various computational methods and algorithms, classification accuracies ranging from 62.3% to 94.4% were achieved (Kim, Kim, Oh, & Kim, 2013).

However, muscle activation and eye movement might influence EEG data recorded from the scalp (Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). Especially at frontal, temporal, and occipital locations, these non-CNS artifacts can affect the EEG measurements (Wolpaw et al., 2002). Since emotions-related features are mostly found in the frontal EEG signal (Kim et al., 2013), these artifacts might impede the emotion measurements. Thus, EEG analysis might add value as opposed to solely using facial expression recognition systems for identifying emotions.

This study answers three research questions. First, it intends to determine how well EEG analysis performs in discrete emotion recognition. Secondly, the performance of EEG will be compared to facial expression analysis to assess how capable EEG analysis is of recognizing affect. The third question is to what extent the combination of modalities (EEG and facial expression analysis) increases the performance of emotion classification.

To answer these questions, an experiment was conducted in which participants were presented with visual stimuli to evoke emotions. These emotions were captured using both an EEG measurement device and video recordings of facial expressions. Then, statistical analyses were used to classify the emotions captured with the EEG device. Furthermore, affective computing processing software CERT was used to classify the emotions captured with the video recordings of facial expressions. Finally, the classification performances were compared in order to assess to what extent unique information is provided by the different methods.

The first section of this paper will review psychological and neuroscientific perspectives on emotion. Moreover, several representations of emotion are discussed. In the next section the field of affective computing is examined, specifically the various methods that aim to detect affective states. The third section is concerned with the methodology used for this study. The final sections present the findings of the research and the discussion of these findings.

Emotion theories

Before continuing to explore the field of affective computing, let us first define the concept of emotion or affect (these terms will be used interchangeably throughout this thesis). This section provides an overview of emotion theories and the definitions of affect that will be used in the following sections.

Psychological and neuroscientific perspectives

Emotions as expressions

Academic emotion psychology became a scientific research topic around the last third of the 19th century (Calvo, D'Mello, Gratch, Kappas, & Reisenzein, 2015, Section 1). In his book *The expression of the emotions in man and animals*, Darwin (1872) noted the similarity in facial and body expressions between species. He theorized that emotional expressions emerged from 'serviceable associated habits' (Calvo & D'Mello, 2010). An example is baring teeth when angry. This could be a vestige of biting, an action that served species in the past and is still associated to a state of 'anger', even when there is no object or person to bite.

Emotions as bodily reactions

Instead of viewing emotions as expressions, William James, a contemporary of Darwin, saw emotions as the sensation of bodily reactions (James, 1884). Around the same time, independently of James, Lange published his ideas on emotions. Given the similarity of their theories, they are often referred to as the James-Lange theory (Lange & James, 1922; Van den Broek, 2011). This 'feeling theory of emotion' states that emotions are the perception of physiological changes (similar to taste or touch) due to emotion-evoking situations. James uses the example of encountering a bear during a hike. This event makes your heart race and your stomach twitch, which is picked up by your senses and relayed to the brain, where these reactions are merged into an experience: fear. The theory's notion of emotions as perception of physiological changes implies that affect can be studied by analyzing the pattern of these changes associated with each emotion, which is an aim of the field of affective computing (Calvo & D'Mello, 2010).

The James-Lange theory was challenged on several aspects. We highlight three challenges. Firstly, the same physiological changes are experienced in different emotional and non-emotional states, which arguably makes them too uniform as a means of distinguishing emotions (Cannon, 1927). For instance, an acceleration of the heart in combination with sweating and erection of hairs can be interpreted as signals of fear. However, the same signals are known to be induced by fever or a state of severe excitement. As such, the physiological changes are not sufficiently capable of differentiating emotional and non-emotional experiences. Moreover, emotions are experienced both with and without bodily feelings (Cannon, 1927). Secondly, emotions do not seem to have a motivational function in the James-Lange theory. Emotions are caused by bodily changes but do not influence action, apart from instinctive action such as fleeing in the case of fear. This was contradicted by motivational theories where emotions are regarded as modes of behavior that are purposive and serve as means for realizing ends (Dewey, 1895). Lastly, the feeling theory does not account for the (representation of) objects that emotions appear to be about. We are not just angry, but angry *about* something. The cognitive processing of this object influences the emotional state (Worcester, 1893). Going back to the example, if a hunter would encounter the bear he might not experience fear because he is convinced that the bear cannot harm him.

Emotions as cognitive appraisals

Arnold (1960) elaborated on the assumption that emotions are object-directed. According to Arnold, emotions presuppose a cognitive representation of an object. Her theory is that objects must be appraised as directly affecting oneself in order to experience emotion (Calvo & D'Mello, 2010). This appraisal consists of three dimensions: an evaluation of the object as being good or bad, absent or present, and feasibility of attaining or avoiding it. These dimensions can be used to describe emotions. For instance, joy can be seen as a positive and easy to maintain event that is present. Fear can be described as a bad and hard to avoid event, absent right now but possible in the future. The three dimensions proved not to be sufficient for distinguishing emotions as several additional and alternative appraisal criteria have been proposed over the years. Roseman (1984) added the appraisal dimensions of responsibility and probability. And Scherer (2001) even suggested four groups of appraisal objectives, named Stimulus Evaluation Checks, containing 16 appraisal criteria. According to Moors (2014), appraisal theories can be divided in two main "flavors". The first flavor theories are aimed at establishing the set of dimensions to differentiate specific emotions. The second flavor theories assume that the appraisal process can have numerous combinations of dimensions, and do not attempt to define patterns that indicate discrete emotions.

In summary, appraisal theories have emphasized the cognitive appraisal process as an important part of emotional experiences. Nevertheless, some questions remain unsolved such as the minimal number of criteria necessary to distinguish affect. Or why some emotional experiences do not require conscious registration (Calvo & D'Mello, 2010).

Neuroscientific views on emotions

Some contributions to the understanding of affect that might solve these questions are coming from advancements in neuroscience. Scholars have used neurophysiological concepts to reinforce their theories throughout the years. First it was posed that the perceiving of bodily changes through the cortex dictates emotions (James, 1884). This was challenged by studies of patients who were able to experience some form of emotions even though their biosignal perception were affected, suggesting the importance of processes in subcortical structures such as the thalamus and hypothalamus (Cannon, 1927). Therefore, both regions of the brain are viewed as involved in the experience of emotion by subsequent theories (Van den Broek, 2011). One example is the Limbic System Theory, which advocates a center of emotions consisting of closely connected subcortical and cortical areas (including the thalamic region and the amygdala) that are relatively isolated from other parts of the brain (MacLean, 1952). The idea of the limbic system as a separated processing system has been criticized on the grounds of not being clearly neuroanatomically distinct from other brain regions; in other words, there may not be an "emotional motor system" in the brain (Ledoux, 1998).

Over the last three decades technological developments, such as brain imaging and electrophysiology, have enabled more research into the functioning of the brain and its role in affect. A number of brain regions are put forward as being involved in emotional experiences:

- The *prefrontal cortex* influences the emotional motivation and regulation (Lindquist, Wager, Kober, Bliss-Moreau & Barrett, 2012). Moreover, hemispheric asymmetry in prefrontal cortical activity is associated with positive and negative valence (Begley & Davidson, 2012).
- Secondly, the *amygdala* region is argued mainly to be concerned with negative emotions such as fear and anxiety (Ledoux, 1998). However, activation of the amygdala during experiences of positive affect has been examined in specific cases, hinting at usage of weak elicitive stimuli in neuroimaging experiments (Hamann, Herman, Nolan, & Wallen, 2004).
- Thirdly, the *anterior cingulate cortex* region of the brain is connected to conscious emotion experience, calmness and interest in particular (Calvo et al., 2015, Section 1).
- Fourth, the *insula* area seems to regulate bodily sensations of emotional experiences and is most consistently activated with feelings of disgust (Murphy, Nimmo-Smith, & Lawrence, 2003).

Overall, affective neuroscience contributed to the field of emotion theory by disproving the division of emotion and cognition through providing evidence that the neural foundations of cognition and emotion overlap substantially (Calvo & D'Mello, 2010). Furthermore, emotions are bidirectional: the brain triggers bodily reactions and these physiological changes are fed back to the brain (Thayer & Lane, 2009; Damasio, 1994). Lastly, affective experiences or behavior may occur both with and without conscious cognitive processing (LeDoux, 1998).

Ill-defined concepts of emotion

In conclusion, numerous theories have been conceived in order to unravel the concept of emotion. Although the frameworks designed and studies conducted from multiple disciplines have contributed to our understanding, the complexity of the topic leaves open questions on the causes, processes and effects of emotional experiences. One of the reasons for its complexity is that the concept of emotion remains ill-defined (Van den Broek, 2011). Based on a qualitative study by Izard (2010) on definitions of emotion that surveyed 34 scientists concerned with emotion theory and research, there is no consensus on the word "emotion". It cannot be defined as a unitary concept, given that the definitions of emotion given by the researchers contained relatively distinct structures and functions. This interchangeable usage of emotional concepts and tackling of the same questions from different theoretical standpoints is arguably leading to an intellectual stalemate (LeDoux, 2012). However, Izard also found that compared to an earlier survey on emotion (Kleinginna & Kleinginna, 1981), researchers showed stronger agreement on the multi-aspect nature of emotion, signaling an advance in our understanding of emotions. Outlined in the next section are the various representations of emotions that are used in the fields of emotion theory and affective computing.

Representation of emotions

Previously we discussed several emotion theories, mentioning labels of affective states such as anger or happiness. Several of these labels of emotional experiences have been suggested, which can be mainly divided in *discrete* and *dimensional* representations of emotion.

Discrete representations of affect

Darwin (1872), viewing emotions as expressions, described tens of discrete concepts such as joy, grief, and meditation using combinations of bodily and facial movements he observed in humans and other species. These concepts closely align with the *emotion mechanisms* that James (1884) thought to be triggered by the perception of bodily reactions. McDougall, (1908/2003) build on these ideas and proposed several *instinct modules*. The term "instincts" emphasizes that he, just as Darwin and James, viewed discrete emotional states as having developed through evolution as solutions to adaptive problems. The emotional states defined by Darwin, James, and McDougall are listed in Table 1. These are just a few examples of the various discrete representations of emotions - ranging from a handful discrete states (e.g. James, 1884) to over a hundred (e.g. Parrot, 2000) - that have been proposed by researchers throughout the years. In this thesis we focus on the discrete emotions proposed by Ekman (1992), as the usage of these concepts is widespread in the field of computational emotion recognition.

Ekman adopted a subset of the labels that were used in previous theories and referred to them as *basic emotions* (see Table 1). Thereby underlining his view of the emotional mechanisms denoted by these labels as "biologically basic" (Barrett et al., 2016, Chapter 1). By designating some emotions as basic it is also postulated that other non-basic emotions are combinations of the basic emotions (Ekman, 1992). There are six basic emotions: anger, fear, sadness, enjoyment, disgust, and surprise. In his research on affective states, Ekman predominantly focused on facial expressions. Several expressions that share certain characteristics are viewed as one emotion family, or a basic emotion. For instance, the basic emotion of anger is a family of states comprising over 60 expressions that share commonalities. Three characteristics identify the separate discrete emotions according to Ekman. Firstly, there is evidence of distinctive, universal facial expressions for these emotions, based on high agreement in labeling across cultures, and studies of deliberate and spontaneous expressions. Secondly, there seem to be distinctive patterns in autonomic nervous system activity across the basic emotions as well. Thirdly, triggering these emotions are prototypical antecedent events.

Facial parameterization

The validity of studies using human observations of emotional expressions of the face was questioned by Ekman and his colleague Friesen (1976). They raised doubt on the reliability and accuracy of human observations of the face used in research, which were mostly subjective analyses of facial expressions that could be influenced by context (Bettadapura, 2012). For instance, a observer could subconsciously give more prominence to voice rather than face. Moreover, observer interpretations of facial expressions could vary across cultures.

To study the expression of basic emotions they developed the Facial Action Coding System (FACS). This system consists of a set of parameters to represent facial expressions. The parameters are muscles and muscle movements that, individually or in groups, cause changes in expressions. The limitations of human observers are theorized to be overcome by representing the face with a fixed set of parameters. The sets of parameters are named Action Units (AUs). To illustrate, AU4 is lowering and pulling the brows together, involving the Corrugator Procerus muscle, whereas AU12 is pulling the lip corners (smiling), involving the Zygomatic major muscle. Combinations of AUs can also be used to represent expressions, such as AU12 followed by AU24 and then by AU51+54, associated with embarrassment.

Coding schemes similar to FACS have been developed, most notably the Facial Animation Parameters (FAPs; Pandzic & Forchheimer, 2003). Although originally intended to facilitate computer animations, the FAPs have been used for facial expression recognition (Bettadapura, 2012). Each FAP is a set of movements of predefined Feature Points on the face, its value indicating the magnitude of the action. FAPs are closely related, and can be mapped to, the AUs of the FACS (Malatesta, Raouzaiou, Karpouzis, & Kollias, 2007).

Dimensional representations of affect

The assumption that basic emotions are associated with particular physiological changes is criticized by psychological constructionists (Barrett et al., 2016, Chapter 1). The existence of basic emotions is supported by cross-cultural studies of observers recognizing affective states in facial expressions. However, observer agreement on the expressed emotions is inflated, and decreases significantly with increasing distance to Western culture according to Russell (1994). Furthermore, no clear neurological basis has been found for grouping processes together as emotion (Ledoux, 1998). These findings have led to the proposal that emotions occur in a continuous space, instead of being discrete states (Russell, 2003). Russell names this space the Core Affect: a conscious experience (raw feeling) evident in moods and emotions. This feeling is an assessment of one's current condition and can be made tangible as a coordinate on a spatial model. Russell proposed a circumplex model of affect with two dimensions: Valence and Arousal (Russell, 1980). In his framework, the Valence dimension can be represented by a horizontal axis ranging from displeasure, through neutral, to pleasure. The arousal dimension can be seen as a vertical axe ranging from deactivation, through neutral, to activation. This results in a space with four quadrants (see Table 1). Wielding this framework, core affect can be plotted on the two-dimensional space. Core affect can be situated around the neutral center of the space, or in the more extreme periphery. Core affect is related to the emotional states as defined by James and categorized as discrete basic emotions by Ekman or accounts by Russell. These accounts of emotion are called prototypical emotional episodes by Russell. A core affect of pleasure may qualify as the emotion of joy, and the combination of displeasure and arousal as fear. However, emotions should be seen as the continuous fluctuations in core affect (Russell, 2003).

The temporal dynamics of affect

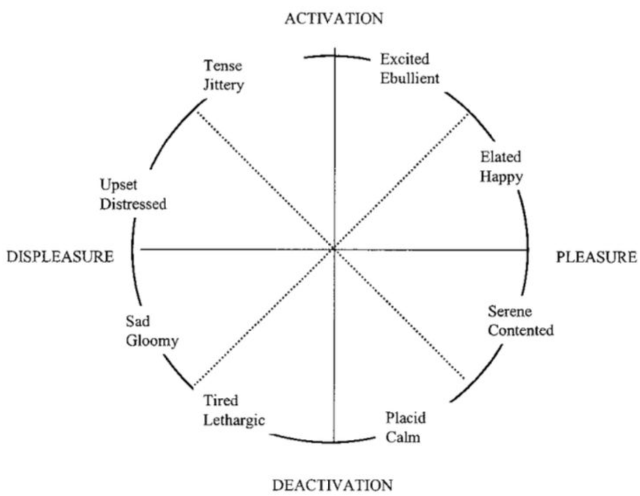
In this section and the previous one, several theories and studies were reviewed that each provide a different perspective of the experience and expression of emotions. So far, the time aspect of emotions has not been explicitly discussed. Not all theories and research on affect provide clear definitions of duration of emotional states. The dimensional approach to affect views emotions as occurring in a continuous space. However, the speed at which one's state of affect moves along the dimensions, or the duration of a prototypical emotional episode, is not declared. This is similar for the discrete approach to emotion representation. Both the concepts of emotions as expressions (Darwin) and as reflex-like bodily reactions (the James-Lange theory) suggest that emotions are relatively short-term phenomena that last for seconds to minutes but not multiple hours. This is in line with Ekman's view of basic emotions as mobilising quickly to be responsive to one's environment, usually lasting only for seconds (Ekman, 1992). Several factors have been mentioned as moderating the duration of emotions. One factor is the type of affective states, some emotions are typically shorter than others (Ekman, 1992; Ekman & Friesen, 1976). For instance the affective state of surprise tends to last shorter than feelings of anger. Another factor moderating the duration of emotions is the situation that evokes the emotion, related to the discussed cognitive emotion theories (Arnold, 1960). The object-directedness of emotions might influence the duration. For instance, going back to the frightening bear encounter example mentioned in the previous section, the context of encountering a bear in real-life or reading about it in a book might influence the duration of the fear emotion.

Furthermore, Ekman theorizes that a single emotion lasting for minutes or hours is likely the same emotion being repeatedly elicited (Ekman, 1992). Moreover, he proposes such a short duration of emotions to distinguish them from moods, that are highly saturated with a specific emotion but can last for hours or days. This view is supported by Scherer, who distinguishes emotions from other psychological constructs using several concepts including duration and rapidity of change (Borod, 2000, Chapter 6). According to Scherer, emotions are both of shorter duration and change faster than moods, attitudes or personality traits.

To conclude, there is little agreement on how long emotions last in the literature on emotion theories. Overall, most researchers agree that emotions are brief in duration, spanning several seconds to minutes rather than hours or days, in contrast to other psychological constructs such as moods or behaviors. Factors that moderate duration of affect include the specific emotional state and context in which the emotional state is evoked.

Table 1. *Examples of representations of affect in literature*

Author	**Representations of affect**						
Darwin (1927) *Expressions of emotion*	Low spirits, anxiety, grief, dejection, despair	Joy, high spirits, love, tender feelings, devotion	Reflection, meditation, ill-temper, sulkiness, determination	Hatred, anger	Disdain, contempt, disgust, guilt, pride, helplessness, patience, affirmation, negation	Surprise, astonishment, fear, horror	Self-attention, shame, shyness, modesty
James (1884) *Basic emotion mechanisms*	Joy	Grief	Fear	Anger	Love	Hate	Pride
McDougall (1908) *Instinct modules*	Instinct of flight (fear)	Instinct of repulsion (disgust)	Instinct of curiosity (wonder)	Instinct of pugnacity (anger)	Instinct of self-abasement (subjection)	Instinct of self-assertion (elation)	Parental instinct (tender)
Ekman & Friesen (1976) *Basic emotions*	Anger	Fear	Sadness	Enjoyment	Disgust	Surprise	
Russell (2003) *Valence and arousal models*							



Detection of affect

Numerous approaches have been developed to measure affect. Self-reporting is one way to assess what emotions people are experiencing. However, it is a rather subjective method. Researchers have been studying people's physiological and behavioral changes as more objective proxies to determine affective states. Several modalities and combinations of modalities have been researched. A meta-analysis of 90 peer-reviewed multimodal affect detection systems shows that facial expressions or vocal behavior have been used in over 75% of these systems (D'Mello & Kory, 2015). Posture, physiology, and text were part of the modalities in at least 10% of the cases. Furthermore, eye gaze and context models were used. Some modalities are briefly discussed in this section. Since facial expressions and physiology are the focus modalities of this thesis, they will be reviewed more elaborately.

Vocal behavior

Information transmitted via speech can be divided in the linguistic message, and vocal non-verbal behavior (paralinguistic features) that influences the message (Calvo & D'Mello, 2010). Major components of non-verbal behavior are speech production and vocalizations (Vinciarelli, Pantic, & Bourlard, 2009).

Speech production system

The speech production system consists of two components, one generating and one modulating the airflow (Calvo et al., 2015, Section 2). Airflow is modulated by the vocal cords (located in the glottis, which is part of the larynx) resulting in variation of pitch and voice quality. Vocal folds open and close with different levels of tension while letting through air, thus creating variations in speech waveforms. Studies show that the vel*ocity at which the vocal folds open a*nd close is different between discrete emotional states. Voice qualities (harsh, breathy, or whispery, to name a few) produced by vocal fold control are associated with affective states. Another paralinguistic feature, pitch, appears to be an index to arousal (Calvo & D'Mello, 2010). Furthermore, after airflow passes the vocal folds, it can be modulated by the articulatory controls such as the tongue and lips, which causes spectral changes in the speech signal (Calvo et al., 2015). More advanced positioning of the tongue tip, jaw and lip has been reported for emotional speech compared to neutral speech. And high articulatory speed modulations are reported for angry speech.

Linguistic and non-linguistic vocalizations

Vocalizations are non-words used in speech such as "uhm" and "ah", or vocal outbursts (non-linguistic) such as laughing and sobbing (Vinciarelli et al., 2009). The linguistic vocalizations are often referred to as disfluencies. They can signal difficulties in finding the proper words to express when speaking, or be used as a way to relay feedback (for instance attention or disapproval) when listening to someone else speaking. The non-linguistic vocalizations provide additional information in an interaction. Some affective states were accurately derived from non-linguistic vocalizations (Russell & Fernández-Dols, 1997).

Effect of language on emotion recognition

The importance of paralinguistic features in vocal emotion expression and recognition appears to be largely unaffected by language. Pell, Paulmann, Dara, Alasseri, and Kotz (2009) compared the acoustic-perceptual underpinnings of vocal emotion expressions in multiple language corpora. In their study, native speakers (in Arabic, English, German, and Hindi) produced pseudo-utterances that were meant to express six basic emotions. Native listeners judged the recordings for their perceived emotional meaning. The research showed similarities in the vocal expressions of emotion across languages, despite the differences in language and linguistic-cultural backgrounds. Recognition rates varied by language, but all emotions could be recognized from vocal expressions at levels exceeding chance showing that recognition performance does not depend on language.

Achievements and challenges of affect detection using speech

A recent review of research on emotion recognition through voice can be found in Zeng, Pantic, Roisman, and Huang (2009). It shows that most efforts were aimed at recognizing (a subset of) the basic emotions. Some studies focused on application-dependent affective states such as stress, confidence, frustration, and empathy. Researchers investigating recognition of spontaneous emotions in more naturalistic interactions often utilized more coarse affective states such as valence or arousal. Zeng et al. also raise current challenges of emotion recognition through audio expression. Although several different paralinguistic features have been used for affect detection, no optimal set has been found. Secondly, reliable extraction of paralinguistic features (such as glottal airflow information) from audio recordings is difficult.

Body expressions

Aspects of physical appearance such as natural characteristics (height, body shape) and artificial characteristics (clothes, cosmetics) convey affective signals and may influence perception of affective states. For example, ectomorphic body types (tall, thin) are associated with pessimistic personality traits (Vinciarelli et al., 2009). In research, increasing attention is devoted to the development of affectively aware technologies that use body expressions, especially posture and gender (Kleinsmith & Bianchi-Berthouze, 2013).

There are benefits to using body expressions for recognition of emotion expressions (Calvo & D'Mello, 2010). The body is relatively large and can assume a large variety of positions and movements. Posture and body movements are considered to be assumed subconsciously for the most part, thus less affected by social editing (Vinciarelli et al., 2009). Nevertheless, body expressions have been given less attention in research on emotions and affect detection than vocal and facial expressions (Calvo & D'Mello, 2010).

Coding of body expressions

There is no predominant coding system, like FACS for facial expressions, developed for describing body expressions (Kleinsmith & Bianchi-Berthouze, 2013). However, a method by De Meijer (1989) showed that movement-dimensions can be used to represent specific emotions. He listed seven dimensions to describe direction (e.g. opening versus closing, and upward versus backward), force, and velocity of body expressions. Combinations of the dimensions indicate distinct emotions. To give an example, happiness can be described by a strong force, fast velocity and an opening arm movement. Other methods take more high-level approach using a reduced set of features, for example upper-body only (Glowinski et al., 2011) .

Achievements and challenges of affect detection using body expressions

A recent survey of 18 studies of body expressions and affect was carried out by Kleinsmith and Bianchi-Berthouze (2013). Most of the studies analyzed used discrete representations of emotion instead of dimensions. Body features conveyed important signals for recognizing affect independently of performed actions (e.g., walking, dancing, gesturing, etc.). Overall, automatic recognition of affect using body expressions achieved well above chance level. However, the researchers note that comparing the system performances properly is difficult considering the many differences in methodology. Two main challenges in recognition of body expressions are mentioned: one is the the need for studies using natural, non-acted data. To illustrate, 11 out of the 18 studies surveyed analyzed the body expressions of dance sequences, which entail exaggerated movements. Affect detection systems that are able to process natural body expressions regardless of actions are desired. Secondly, the importance of taking into account individual differences in body expressions is stipulated, since high variation was found.

Facial expressions

The human face contains the eyes, nose, and ears, enabling us to see, smell, and hear. It is believed to be a central part of the human system of communication and understanding of affect and intention (Pantic, Nijholt, Pentland, & Huanag, 2008). Therefore, it is not surprising that a large part of the research on emotion recognition has been aimed at detecting affect from facial expressions. There are four routes in which the face conveys information (Pantic & Stewart, 2007). Firstly, the static features of the face that are permanent like overall proportions and skin type. Secondly, the facial changes that occur over time such as the developments of wrinkles that signal age. Thirdly, artificial additions to the face, for instance glasses and makeup, which may obscure or enhance facial features. Lastly, temporal changes in neuromuscular activity that lead to changes in facial appearance, like smiling or blushing. These signals underlie facial expressions and have been studied to detect affect, both manually (observing) and automatically (computational).

Studies are carried out to research which aspects of the face are eminent for displaying and recognizing facial expressions. In such studies, parts of the face are occluded to test the effect on recognition rates of facial expressions. Bettadapura (2012) lists findings on the role of the various facial features. The eyebrows and mouth are considered to convey the most information on the face expression that is displayed. Movements of these parts of the face are associated with the discrete emotions of happiness, surprise, and disgust. Sadness is mostly signalled by configurations of the mouth. Furthermore, it has been found that the occlusion of one vertical half of the face does not impact recognition rates to a great extent, since facial expressions are mainly symmetrical.

Steps to recognize facial expressions

Detecting the face

Robust systems for recognizing facial expressions have been developed since the early 1990s (Bettadapura, 2012). The systems address one or more tasks underlying the challenge of facial expression recognition (Vinciarelli et al., 2009). The first step to (computationally) analyzing facial expressions is face detection. The shape of the face has to be localized in an image or video. In the case of video, the face has to be tracked across different frames. The presence of multiple faces, occlusions of the face, varying positions of the face, and changing lighting conditions can make face detection more difficult. Despite these challenges, methods have been developed to carry out the task (for an overview see Bettadapura, 2012). Most of the recent approaches focus on probabilistic and statistical learning techniques (Vinciarelli, 2009). Rowley, Baluja, and Kanade (1998) attempted to infer the presence of a face from the pixel values. They used a neural network to classify specific portions of video frames (e.g. 20 by 20 pixel regions) as face or non-face, from the spatial relationships and intensities of pixels in face and non-face images. This technique has limitations as it is difficult to train a classifier on the non-face class, since images without a face can contain all kinds of information. An example of a more recent, commonly used face detector was developed by Viola and Jones (2004). It consists of a cascade of classifiers trained by the AdaBoost learning algorithm. The cascading of successively more complex classifiers that focus on "promising" image regions (that are likely to contain the face) allows for faster separation of the background from the face. The classifiers employ integral image filters, or box filters, which are representations of the image allowing for computation of features at scale.

Extraction of facial features

After the face is detected, facial features need to be extracted. Various approaches are described in Pantic and Stewart (2007). One method to accomplish feature extraction is by tracking geometric features like shapes (mouth) or points (lip corners). Another method, appearance-based, is analyzing skin texture changes due to muscle contraction (wrinkles and furrows). Geometric-based extraction has been outperformed by methods based on appearance features (Pantic & Stewart, 2007). However, since some methods are more prone to specific conditions than others, e.g. appearance-based features may be negatively affected by lighting variations, combinations of geometric and appearance feature extraction might be most suited for facial expression recognition.

Affective state classification

Eventually, the information of extracted facial features is translated into an affective state. In some cases this is done directly, in other cases the facial features are first converted to Action Units from which emotions are determined. Several methods are utilized, among which rule-based models, Naïve Bayes, Hidden Markov Models, and Support Vector Machines. Usually, translating data into affective states is approached by applying classification algorithms using machine learning. For an overview of methods, see Pantic and Stewart (2007), and Bettadapura (2012).

Example of an emotion recognition system using facial expression analysis

An example of an appearance-based approach to affect detection in images and videos is the Computer Expression Recognition Toolbox (CERT; Littlewort et al., 2011). First, it detects the facial regions using the approach proposed by Viola and Jones (2004). Then, a set of 10 facial feature positions is located by using feature-specific detectors. After Gabor image filters are applied to the image (or video frame), a feature vector is generated. This feature vector is used as input for separate linear support vector machines (SVM) for 19 AUs. The SVMs were trained on a collection of face expression databases coded with AUs. The output is a continuous value for each AU per image, providing information on the intensity of the facial actions. To recognize the basic emotions (happiness, sadness, anger, disgust, surprise, and fear), AU values are put into a multivariate logistic regression classifier. The classifier was trained on a dataset of facial expressions and their emotion labels. As a result, the probability for each basic emotion is provided. In short, CERT is able to recognize intensity of 19 AUs, and the probabilities of 6 basic emotions achieving accuracy rate of 87% (Littlewort et al., 2011).

Achievements of affect detection using facial expressions

Reviews of the state-of-the-art of computational recognition of affect from facial expression have been published (Bettadapura, 2012; Valstar et al., 2012; Zeng et al., 2009). From these reviews it can be derived that the majority of studies are aimed at detecting AUs and / or a discrete set of prototypical emotions (mostly the six basic emotions). As for AU detection, since manual coding images or segments of videos is a time-intensive task, automation is deemed an answer. A challenging endeavor, yet progress has been made as was exemplified by the description of CERT. Studies have been able to computationally detect facial behavior (Pantic & Stewart, 2007).

Despite that, the reliability of automatic face action recognition in videos is lower than that of humans (Calvo & D'Mello, 2010). This can be partly ascribed to humans' propensity to rely on context of the interaction to correctly infer a person's affective state (Hoque, Kaliouby, & Picard, 2009). A face that is made of AUs that corresponds to a basic emotion can be labeled as a different emotion when the context is modified. However, automated classifiers often blindly associate a set of AUs with specific affective states, regardless the of context (Hoque et al., 2009). Therefore, it is important to take into account the sequence in which the AUs appear and interact with each other when classifying videos.

Using the detected AUs, or via other approaches, researchers have been trying to recognize affect from facial expressions. Not many studies have analyzed recognition performance of the basic emotions across publications. Finding a generic method for such a comparison is said to be highly challenging and controversial, given the different datasets and methodologies applied (Poria, Cambria, Bajpai, & Hussain, 2017). Valstar et al. (2012) reported on classification accuracies in a meta-analysis of studies submitted for a facial expression recognition challenge. With achieved classification accuracies of over 80% (and up to 100%), it is argued that inferring discrete affective states from (acted) face expression videos on subject-dependent test sets can be regarded as a solved problem. As for subject-independent methods, accuracy rates of between 80% and 90% have been achieved. Out of the six basic emotions, happiness and surprise are most accurately detected (Bettadapura, 2012). Automatic face recognition systems most often confuse anger with disgust, fear with happiness, fear with anger, and sadness with anger (Bettadapura, 2012).

Limitations of using the face modality for emotion recognition

Although promising recognition performance has been achieved, several limitations have to be noted. One problem is that the facial expressions used in research are often staged by actors. Approximately half of the 29 studies reviewed by Zeng et al. (2009) utilized data sets with posed expressions. Acted expressions of emotions are likely to be exaggerated in comparison to natural displays of emotion. Since overstated facial expressions can be picked up with relatively more ease, this inflates detection accuracy and results in models that are not suited to real-world data. Aspects of the movements of facial features, such as duration and occurrence of brow actions, have been reported to differ between acted and spontaneous expressions as well (Valstar, Pantic, Ambadar, & Cohn, 2006). Another limitation in automatic affect detection from facial expressions is that present systems mostly rely on the input of unobstructed, close-up, frontal views of the face (Littlewort et al., 2011). Some studies have showed the possibility of profile-view affect detection (Li & Jain, 2011, Chapter 19; Pantic & Patras, 2006). However, these systems do rely on full-profile-views with little occlusion of facial features.

Physiology

Several measuring methods are used for affect detection (Calvo & D'Mello, 2010). Examples are functional Magnetic Resonance Imaging (fMRI) and lesion studies. Other techniques are relatively less invasive, such as measuring electrical signals produced by the body: muscle activity (Electromyogram), skin conductivity (Electrodermal Activity), heart activity (Electrocardiogram), eye movement (Electrooculogram), and brain activity (Electroencephalography).

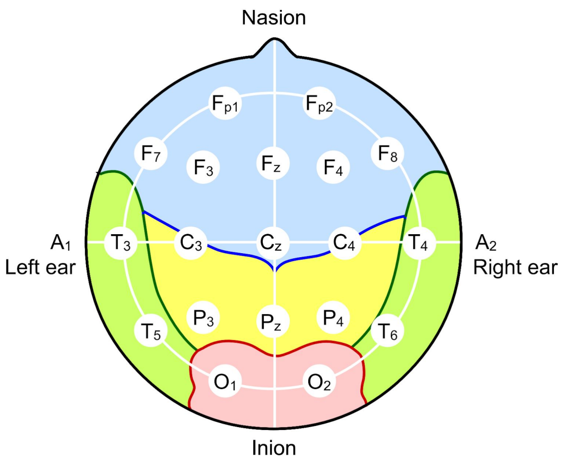
Electrical activity in the brain

Cells in the human brain are commonly referred to as neurons. Neurons receive signals from, and transmit signals to, other neurons in the form of a wave of electrical discharge (Bos, 2006). The electrical activity of a single neuron is too small to be detected. However, when numerous neurons fire simultaneously, the waves of electrical discharge can be measured through electrodes on the head (EEG analysis). This is a relatively non-invasive method of measuring electrical activity in the brain. However, measurement is complicated by the skull and tissue, weakening the signal. A disadvantage of EEG is the relatively low spatial resolution (compared to for instance fMRI), prohibiting detailed anatomical studies. Nevertheless, it gives insight into the electrical activity of the cortex.

Measurement of brain activity through EEG

Standardized locations for EEG electrodes on the scalp enable replication of setups (Bos, 2006). The International 10-20 system is a standardized set of electrode positioning (Jasper, 1958; see Figure 1) . Since the size of the head is variable, the 10-20 system defines electrode distances in percentages, spanning the top of the head in sections of 10 and 20 percent. Electrode placement starts from 10% above the nasion (F_p) and inion (O) points, followed by electrodes on 20% distances in between the starting points. The electrode positions are labeled by combinations of letters and numbers. The F, T, C, P and O letters refer to the lobes, such as F for frontal and T for temporal, with the exception of C (central) which is present for positioning purposes. The numbers indicate electrodes at either the right (even) or the left (odd) hemisphere.

Figure 1. *The International 10-20 system of electrode positioning (Bos, 2006)*



Expression of EEG as event-related potentials

The electrophysiological potentials recorded via electrodes can be expressed as event-related potentials (ERPs). ERPs are patterns of EEG measurements that occur as a response to stimuli. The varying perceived emotional values, and conscious evaluation, of stimuli result in different ERPs. Some potentials occur relatively fast after exposure to stimuli, for example 100 ms. Others ERPs, detected at later times like 300 ms, are regarded late potentials. Two ERPs that have been found to respond to salient emotional stimuli are the P300 and late positive potential (Mühl, Allison, Nijholt, & Chanel, 2014). Furthermore, signals captured by the electrodes can be expressed in frequency ranges that describe oscillatory characteristics or brain rhythms. Five conventional frequency bands are delta (0.5 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 13 Hz), beta (13 to 30 Hz), and gamma (above 30 Hz).

Expression of EEG as frequency bands

Recognition of dimensional states of affect

Frequency bands are used to detect affective states via affective brain-computer interfaces (BCI), mostly in the valence-arousal space (Mühl et al., 2014). An indication of valence is hemispheric alpha asymmetry. Negative emotions are associated with left frontal inactivation, and positive affect with right frontal inactivation (Bos, 2006). Since alpha bands are more dominant in a more relaxed state state of mind, an increase in alpha activity has been linked to brain inactivation. An observed increased alpha power in the left frontal region of the brain together with decreased beta activity indicates a negative affective state.

Since beta waves are associated with an alert state of mind, the beta/alpha ratio could be an indication of the state of one's arousal (Bos, 2006). Taking this perspective into account, an Asymmetry Index (AsI) was introduced to compute the information flow from one hemisphere to another (Petrantonakis & Hadjileontiadis, 2010). Using the AsI, more intense asymmetry (a lower amount of shared information between the hemispheres) was associated with high-arousal stimuli. Examples of other signals of arousal are the power spectral density (estimated by averaging the electrode periodograms) of alpha bands in the occipital area (Soleymani, Pantic, & Pun, 2012) and higher beta power and coherence in the parietal lobe, plus lower alpha activity (Choppin, 2000).

Recognition of discrete states of affect

There have been efforts to recognize discrete emotional states via EEG as well. In an overview of research on emotion detection via EEG for brain-computer interfaces, 6 out of 18 reviewed studies were aimed at detecting emotions other than the valence-arousal dimensions (Mühl et al., 2014). In an overview of computational methods for emotion detection via EEG, this was the case for 6 out of the 15 reviewed publications (Kim et al., 2013). Of these studies, not all report on the unimodal EEG results, and the ones that do mainly use the frequency bands to detect affective states. Balconi and Mazza (2009) found right-frontal cortical asymmetry as a response to negative potentially threatening emotions (fear, anger, and surprise). Li and Lu (2009) found EEG gamma band activities to be suitable for detection of happiness and sadness. Park et al. (2011) showed participant videos to induce the emotions of happy, sad, peaceful, fearful. Comparing to a neutral state, they found a decrease in alpha in the left temporal lobe for negative emotions, an increase in beta in the left temporal lobe when exposed for fearful emotions, a decrease in alpha at the C4 location for happy emotions, and an increase in gamma at the T5 location for the peaceful emotions.Petrantonakis and Hadjileontiadis (2010) elicited the six emotions of happiness, surprise, anger, fear, disgust, and sadness by showing participants images of facial expressions and were able to differentiate participants' emotions with a 83% mean classification accuracy rate using SVM classifier.

Challenges of affect detection using EEG

Most studies mentioned in the paragraphs above are affected by a number of limitations and challenges in the fields of emotion recognition via EEG. Certain facial expressions like frowning and smiling involve movement of the scalp muscles. Together with other activities such as blinking, this can contribute to the electrical activity measured from the scalp (Wolpaw et al., 2002). This means that such cases of facial activity can trigger power increases in EEG frequencies, especially of the high-frequency bands of beta and gamma (Mühl et al., 2014; Wolpaw et al., 2002). These disruptions of the EEG signal are referred to as artifacts (Bos, 2006). Usually, researchers act on artifact influence to prevent potentially misleading EEG results (Wolpaw et al., 2002). Artifacts can be partly prevented by asking people whose EEG signals are measured not to blink and move. A different approach is to filter artifacts from the data after recording.

A second challenge is the participant-independent affect detection using EEG signals. Participant-independent, generalizable affect detection allows a system to maintain recognition accuracy when applied to new individuals (D'Mello & Kory, 2015). This is desirable since training procedures tend to be long and intrusive (Mühl et al., 2014).

Finally, obtrusiveness of EEG measurement devices is a challenge to efforts of emotion detection via EEG. EEG measurement is less physically constraining than fMRI scans, yet requires participants to wear saline-soaked electrodes on the scalp that take time to be correctly placed. Therefore, the technique used to be confined to hospitals and laboratories (Aspinall, Mavros, Coyne, & Roe, 2015). However, in recent years efforts have been made to make the EEG devices smaller, lighter, and thus less obtrusive. These devices tend to have a reduced number of electrodes, but several studies have shown the ability to detect affect to some extent using only 3 to 4 electrodes (Petrantonakis & Hadjileontiadis, 2010; Frantzidis et al., 2010). Moreover, a smaller number of sensors might entail some advantages (Mühl et al., 2014). Generally, wearing a device with more sensors means less comfort for the user. Increasing numbers of sensors also lead to high-dimensional features spaces that can make classification analyses more challenging.

Examples of commercially available portable EEG devices are Brainquiry PET 4(2017), NeuroSky MindWave (2017), Interaxon MUSE (2017), Emotiv Insight (2017), and Emotiv EPOC (2017). All these devices have in common that they connect to a computer over bluetooth. Furthermore, aside from the software installation, they are all relatively easy to set up within a couple of minutes, in contrast to medical-grade EEG devices. Most mentioned devices can be used with dry sensors, expect for the EPOC which needs a saline solution to be applied to the sensors before usage. The portable EEG devices differ on more aspects, such as price and the number of channels accessible through the EEG, ranging from 4 (the PET 4 device) to 14 (the EPOC device). Furthermore, the frequency bands measured by the devices range from 0 to 400 Hz. Lastly, the sample rates per sensor across these devices vary between 128 Hz (EPOC) and 512 Hz (MindWave) per sensor. In this study, the Emotiv EPOC device is used for EEG measurement. The main reason for this is ease of access to the device.

EEG affect detection using the Emotiv EPOC device

Duvinage et al. (2013) compared the performance of the Emotiv EPOC to a medical-grade system. They found that the EEG classification performance of the EPOC is above random and not due to muscular or ocular artefacts. However, compared to the medical-grade system the EPOC showed a medium to large underperformance. Although the device performance does not seem to be on par with medical-grade systems, several studies have been able to use the EPOC for affect detection. We have listed an overview of these studies in Table 2, adopting a format D'Mello and Kory (2015) used to compare multimodal affect detection systems. *Data Type* refers to the sort of affective expressions that were used. All studies made use of expressions that were induced via experimental methods. Representation model (*Rep. Model*) pertains to whether dimensional or discrete representations of emotion were used. Only one study used discrete representations of emotion. *Class Model* describes that type of analysis that was used. Apart from one study using a regression analysis, all research used classification analyses. The number of affective states that the researchers aimed to recognize is noted by *k*. For multimodal affect detectors, *Fusion Type* refers to the level at which the modalities were combined. Only one study employed multiple modalities and fused the data at the feature level, meaning features were independently computed from each modality and fused before classification. The *Validation Method* of the studies was reported person-dependent, since they cross-validated within an individual or did not report on person independence across training and testing sets. *Calculation* refers to whether studies manually calculated features from the raw EEG sensor data or used the calculated affect variables provided by the Emotiv software. The utilized EEG sensor locations are summarized under the *Locations* header.

Lastly, *Measure* and *Accuracy* are the reported performance metric (correlation coefficient or accuracy) and corresponding value. Inventado, Legaspi, Suarez, and Numao (2011) researched recognition of students' emotional state (frustration and excitement) as a result of feedback from an intelligent tutoring system. Using the calculated affect variables from the Emotiv output as features, they achieved a performance of 0.627 (R^2) for frustration and 0.484 (R^2) for excitement. Ramirez and Vamvakousis (2012) studied detection of valence and arousal elicited by sound fragments. The EEG data of locations AF3, AF4, F3, and F4 were used to obtain average classification accuracies of 77.82%

Table 2. *Comparison of studies using the Emotiv EPOC for affect detection*

	Inventado et al.	**Ramirez & Vamvakousis**	**Liu & Sourina**	**López-Gil et al.**	**Pham & Tran**
Year	2011	2012	2012	2016	2012
Participants	10	6	14	44	Undisclosed
Data Type	induced	induced	induced	induced	induced
Rep. Model	dim	dim	dim	dim	disc
Class Model	reg	class	class	class	class
k		4	2	3	3
Affect States	frustration, excitement	valence, arousal	dominance (high & low)	pleasant, unpleasant, neutral	amusement, fear, neutral
Modalities	EEG	EEG	EEG	EEG + EDA	EEG
Fusion Type				feat	
Validation Method	dep	dep	dep	dep	dep
Calculation	emotiv	raw	raw	raw	raw
Locations	(all)	AF3,AF4,F3,F4	FC6, F8	(all)	(all)
Measure	cc	acc	acc	acc	acc
Accuracy	0.555	0.790	0.742	0.425	0.928

for arousal and 80.11% for valence. Again using sound fragments, Liu and Sourina (2012) investigated detection of dominance (low versus high) from EEG. Employing the locations FC6 and F8, resulting in a classification accuracy of 74.15% (average over three experiments). Pham and Tran (2012) employ videos in the form of movie fragments for eliciting amusement, fear and neutral states of emotion. Using the EEG data from all electrode locations, a classification accuracy of 92.8% is achieved. López-Gil, Virgili-Gomá, Gil, and García (2016) showed participants both emotion-eliciting images and videos to detect valence and arousal. In the multimodal research, EEG data are used only to detect valence but classification accuracies are not reported for the separate modalities.

In conclusion, the comparison of emotion recognition studies using the Emotiv EPOC device for EEG measurement shows that the device is to some extent capable of detecting affect. Furthermore, to our knowledge, there has been only one study aimed at recognizing (three) discrete emotional states using the EPOC device. Moreover, the EPOC device has only been used in one multimodal study of affect detection.

Multimodality

In the previous paragraphs several modalities were discussed that are being used for emotion recognition. These modalities differ in the data that they gather (speech, body posture, brain activity), which gives certain modalities advantages and disadvantages. For example, over a long distance, it can be more difficult to recognize a face than it is to recognize body posture (Calvo & D'Mello, 2010). Differences in accuracy between modalities have been found as well, as lower accuracy rates have been reported for (basic emotions) affect detection from speech compared to facial expressions (Calvo & D'Mello, 2010). Given the variety between modalities, it has been theorized that combining them might yield increased performance of affect detection (Calvo & D'Mello, 2010; D'Mello & Kory, 2015; Mühl et al., 2014).

Methods of fusing modalities

In order to apply a multimodal setup for affect detection, the various modalities have to be fused. This can be done on several levels, depending on when information from the modalities is combined (Calvo & D'Mello, 2010). At the data level, the raw data of each source is merged. This is only possible when sources have the same temporal resolution. Mostly, data-level fusion is used for signals coming from the same recording devices (video data from multiple cameras). At the feature level, the processed output of modalities prior to classification is combined. For instance, the mean AU values from facial expressions with the mean values of the EEG channels. Decision-level fusion is performing classification on the individual modalities and merging the outputs.

Achievements of multimodal affect detection

In a meta-analysis of 90 peer-reviewed studies, D'Mello and Kory (2015) described the state of the art of multimodal affect detection systems. As mentioned at the start of this section, the majority of multimodal affect detection employs face and speech modalities (over 75%). As for the source data, multimodal detectors were mostly trained on actor-portrayed displays of emotions (52%) as opposed to induced (28%) or induced (20%) affective displays. The reviewed studies mainly represented emotion in discrete categories (64%) rather than dimensional (36%). Feature-level (39%) and decision-level (36%) fusion were applied most frequently.

Moreover, D'Mello and Kory assessed whether multimodal approaches led to enhanced classification scores compared to unimodal approaches. From their comparison of effects and accuracy rates, it was concluded that multimodal approaches outperform their best unimodal counterparts. The improvements were modest, especially for affect detection of naturalistic displays. The main reason for this is argued to be the redundancy among various modalities. The best performing unimodal accuracies explained 76% of the variance in multimodal accuracies. Furthermore, this idea is supported by the finding that multimodal accuracies increased when the unimodal accuracies proved more diverse, as multimodal systems in that case had more non-redundant data.

Experiment

The research reviewed in the previous sections shows that emotion is a broad concept involving multiple disciplines and various theoretical standpoints. There is little consensus over the definition of emotions, but considerable agreement over their activation, functions, and regulation. Although there are various proposed representations of emotions, progress has been made in emotion recognition over the past decades; especially through the rise of computational affect detection using facial expression and voice modalities. More recently, new modalities such as physiology (e.g. EEG) and the combination of modalities have contributed to the field of affect detection.

Most research on affect detection using EEG analysis has been aimed at recognizing dimensional representations of emotion, specifically the valence-arousal dimensions. Nevertheless, previous studies suggest that EEG activity can be used to detect discrete emotional states as well. A challenge of using EEG as a modality for emotion recognition is the obtrusiveness of the measurement devices. The emergence of portable, less-obtrusive EEG devices has tried to solve for this issue. Although the number of sensors on these less-obtrusive devices tends to be lower, research has shown their suitability for valence-arousal recognition to some extent. However, discrete affective state detection performance when using less-obtrusive EEG devices is still largely unknown.

Another challenge of EEG analysis highlighted in the reviewed literature is the influence of disruptions of the EEG signal due to electrical activity on the scalp caused by facial activities such as blinking, frowning, or smiling. Influence of these artifacts on EEG analysis can be partly removed by filtering the data, but the EEG signal can still be affected by facial activity. This raises questions on the previously reported performance of EEG as modality for affect detection. To what extent was the EEG data of these studies influenced by facial activity? Previous research established facial expression analysis as a robust modality for emotion recognition. So, facial activity influence on the EEG measurement might contribute to, or even explain, an automated classifier's ability to distinguish emotional states. To date, however, the effects of the interaction of facial activity and EEG activity on affect detection performance have not been closely examined.

An experiment was conducted in order to evaluate the extent to which discrete emotional states can be recognized using EEG, compare the EEG analysis performance to that of facial expression analysis, and assess the added value of EEG data to facial expression data.

Method

Participants

A total of 27 participants were recruited for this study. The participants consisted of students at Tilburg University. Participants were between 18 and 27 years old. 16 men and 11 women participated in the experiment. All participants in this study were volunteers.

Design

The study had a within-subjects design. Each participant was shown a series of 5 movie scenes. All participants were shown the same stimuli in counterbalanced order. Each stimulus was expected to elicit an affective response. This affective response was captured using two methods: a video camera recording facial expressions and an EEG device recording brain activity. These modalities were then compared using classification and mixed-effect analyses. Before and after watching a movie fragment, participants performed a filler task of solving a maze. The filler task goals were to conceal the actual aim of the experiment and to stimulate a neutral state of mind in between the videos.

Materials

To measure brain activity, an Emotiv EPOC device was used (see Figure 2). A Qconferencing Ex'ovision device was used to record participants’ facial expressions (see Figure 3). Movie fragments were chosen in order to elicit amusement, sadness, disgust, fear, or a neutral state. To determine which movie fragments to use, several studies of emotion elicitation using films were reviewed (Bartolini, 2011; Carvalho, Leite, Galdo-Álvarez, & Gonçalves, 2012; Coan & Allen, 2007, Chapter 1; Gabert-Quillen, Bartolini, Abravanel, & Sanislow, 2015; Gross & Levenson, 1995; Hagemann et al., 1999; Hewig et al., 2005; Philippot, 1993; Schaefer, Nils, Sanchez, & Philippot, 2010; Tomarken, Davidson, & Henriques, 1990). Attributes of the studies are listed in Table 3.

Figure 2. *Emotiv EPOC*

Figure 3. *Example of participant recording*

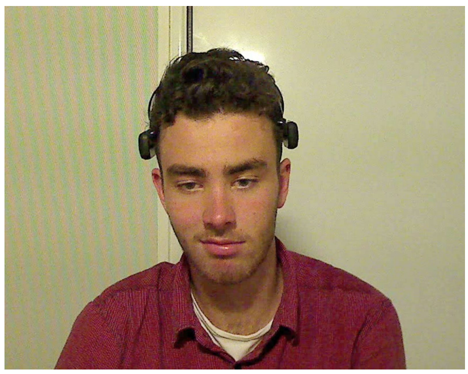


Table 3. *Film stimuli used for eliciting emotions*

	Philippot	**Gross & Levenson**	**Coan & Allen**	**Schaefer et al.**	**Bartolini**	**Gabert-Quillen et al.**
	1993	1995	2007	2010	2011	2015
Amusement	Le magnifique (1973)	When Harry Met Sally, Robin Williams Live	When Harry Met Sally, Robin Williams Live, Cosby, Whose Line	The Three Brothers, When Harry Met Sally, Something About Mary	The Hangover, Modern Times, Monthy Python and the Holy Grail	Modern Times, The Hangover
Disgust	Faces of death (1986)	Pink Flamingos, Amputation	Pink Flamingos, Amputation, Foot Surgery	Trainspotting, The Silence of the Lambs	The Fly, National Lampoon's Van Wilder,	The Fly, National Lampoon's Van Wilder,
Sadness	Kramer vs. Kramer (1979)	The Champ, Bambi	The Champ, Lion King, Return to Me	City of Angel, Life is Beautiful, Dangerous Mind	My Girl, Saving Private Ryan, The Shawshank Redemption	My Girl, Shawshank Redemption
Neutral	Belgian tv documentary	Abstract Shapes, Colour Bars	Sticks screen saver, Denali	Blue, The Lover	Planet Earth: Shallow Seas, Pride and Prejudice, Searching for Bobby Fisher	Pride and Prejudice, Searching for Bobby
Fear	Psycho (1953)	The Shining, Silence of the Lambs		The Blair Witch Project, The Shining	Psycho (1953)	Psycho, The Ring
Clip length	3-6 min	0-9 min	0-9 min	1-7 min	0-8 min	1-7 min
Language	French (dubbed), subtitles	English	English	French (dubbed)	English (dubbed), subtitles	English
Groups	Alone	Max. 30		Max.5	Max. 10	Max. 17
Measure-ment	Modified Differential Emotional Scales (Izard, 1974) Modified Semantic Differential Scales, IAPS (Osgood, 1964)	Modified Differential Emotional Scales (Izard, 1974)	Developed Post Film Questionnaire	Modified Differential Emotional Scales (Izard, 1974) PANAS (Coan & Allen, 2007; Watson, Clark, & Tellegen, 1988)	Post Film Questionnaire, (Coan & Allen, 2007) Modified Semantic Differential Scales, IAPS (Osgood, 1964)	Post Film Questionnaire, (Coan & Allen, 2007) Modified Semantic Differential Scales, IAPS (Osgood, 1964)

Four of the reviewed studies were omitted from Table 3 for not having included sound in the movie fragments (Carvalho et al., 2012; Hagemann et al., 1999; Hewig et al., 2005; Tomarken et al., 1990). Two of the remaining studies reported in Table 3 were dismissed for using French (or French dubbed) movie fragments. One of the listed studies was excluded since there were no stimuli included to elicit fear (Coan & Allen, 2007). Of the remaining studies, film stimuli of Gabert-Quillen et al. (2015) were chosen because of their recency compared to the other stimuli. Out of the film clips included in their study, five fragments with the highest average intensity scores for the to be elicited affective states were chosen (see Table 4). The intensity scores were derived from indications of emotional experience obtained through postfilm questionnaires that were filled out by participants who watched the movie fragments (Gabert-Quillen et al., 2015).

Table 4. *Selection of movie clips taken from Gabert-Quillen et al. (2015)*

Affective state to be elicited	**Origin movie of film clip**	**Length in minutes**	**Average intensity (SD) ***
Amusement	The Hangover (2009)	5:47	6.78 (1.4)
Disgust	Van Wilder (2002)	3:07	6.98 (1.6)
Sadness	The Shawshank Redemption (1994)	4:14	6.88 (1.5)
Neutral	Pride & Prejudice (2005)	1:37	6.25 (1.9)
Fear	The Ring (2002)	2:45	5.30 (2.4)

* On a scale of 0 (not at all/ none) to 8 (extremely/a great deal)

In order to guide the participants through the experiment procedure a website was developed specifically for this study (see screenshots of the interface in Table 6). Since the website was displayed in a browser set to fullscreen mode, participants were able to view only the experiment content on the computer screen. Mouse and keyboard input were tightly controlled, for instance participants could not launch the context-menu by right-clicking anywhere on the webpage. Or when the video started playing they could not change the playback (pause, forward, replay, etcetera.) in any way. Participant input and timestamps (such as the time the video started and finished) were recorded through text areas and radio buttons. All submitted data were automatically sent to and collected in Google Spreadsheets.

As a filler task, the solving of randomly generated mazes was chosen (Boström, 2010; see Appendix A). Solving a maze made up of black lines is an abstract task suited as a filler task that does not influence the affective state of participants. The generated mazes were equally difficult and solvable within a minute, thus not evoking feelings of frustration.

Measures

During the viewing of each clip, the data of 16 EEG channels (names according to the international 10-20 electrode location system) are recorded: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, with CMS/DRL references in the P3/P4 locations (Emotiv, 2017). Emotiv calculates several variables based on the raw data. The EmoState Logger application in the software development kit was used to extract these variables. Furthermore, facial expressions in the video recordings of participants were measured automatically using CERT (Littlewort et al., 2011), a computational method for analyzing component movements that correspond roughly to individual facial muscles, named Action Units (AU), to describe facial expressions (Ekman & Friesen, 1976). For an overview of all the measures that were recorded by CERT and Emotiv, please refer to Appendix B. For an overview of the variables that were used in the analyses, see Table 5. All these variables were continuous values between 0 and 1.

Table 5. *CERT and Emotiv variables used in the experiment*

Modality	Variable	Definition
Emotiv *	Long-Term Excitement	*"Long-Term Excitement is [...] designed and tuned to be more accurate when measuring changes in excitement over longer time periods, typically measured in minutes."*
	Short-Term Excitement	*"[Short-Term Excitement] is experienced as an awareness or feeling of physiological arousal with a positive value. Excitement is characterized by activation in the sympathetic nervous system which results in a range of physiological responses including pupil dilation, eye widening, sweat gland stimulation, heart rate and muscle tension increases, blood diversion, and digestive inhibition. [Short-Term Excitement] detection is tuned to provide output scores that more accurately reflect short-term changes in excitement over time periods as short as several seconds."*
	Engagement / Boredom	*"Engagement is experienced as alertness and the conscious direction of attention towards task-relevant stimuli. It is characterized by increased physiological arousal and beta waves (a well-known type of EEG waveform) along with attenuated alpha waves (another type of EEG waveform). The opposite pole of this detection is referred to as 'Boredom' [...]; however, please note that this does not always correspond to a subjective emotional experience that all users describe as boredom."*
CERT **	AU (1,2,4,5,6,7,9,10,12,14,15,17,18,20,23,24,25,26,28,45)	Action Units from the Facial Action Coding System.
	Anger, Contempt, Joy, Sad, Joy, Disgust, Fear, Neutral	*"CERT implements a set of [...] basic emotion detectors, plus neutral expression, by feeding the final AU estimates into a multivariate logistic regression (MLR) classifier. The classifier was trained on the AU intensities, as estimated by CERT, on the Cohn-Kanade dataset and its corresponding ground-truth emotion labels. MLR outputs the posterior probability of each emotion given the AU intensities as inputs."*

* Emotiv (2014). ** Littlewort et al. (2011)

Procedure

First, participants read the experiment introduction, instructions, and consent form (see Appendix C). The instructions stated that the experiment was about the effect of visual impressions on analytical capacities. After participants read and agreed with the instructions, they were asked to enter a soundproof booth and take place in front of a computer screen. Next to the screen a stack of five sheets was placed, with a randomly generated maze printed on each sheet. Then, the EEG measurement device was placed on the participant's head. Once the device was set up, the experimenter started the video recording.

Participants started the experiment by clicking the start button on the screen. After solving a maze, the participants watched the videos to elicit affective states in five sets (see Table 6). In each set, the participant watched the movie clip followed by solving a maze. The movie fragments were shown in counterbalanced order. At the end of the experiment the participants were shown screenshots of the clips, and asked to write a short summary of the video they just watched. After solving the final maze, participants were asked to fill out their gender, age, and indicated their command of the English language. Finally, the participants were debriefed.

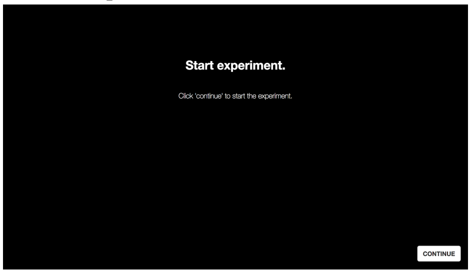
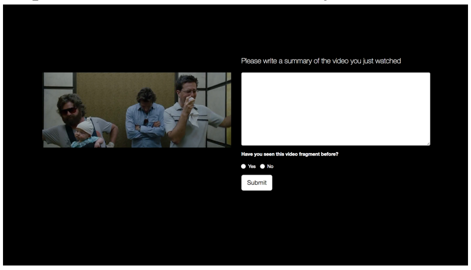
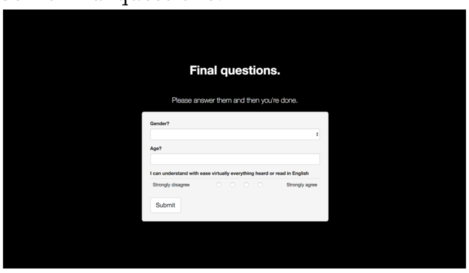
Data pre-processing

Before and after the videos were played, a bright white screen was displayed. These flashes of light were visible in the video recording and were used to pinpoint the timestamps during which participants were watching video fragments. The codes identifying the video that were shown were then manually added to the rows of video and EEG data.

Face recordings were missing for three participants, due to errors during the data collection. For the same reason, EEG data were not measured for one participant. These cases were excluded. CERT was not able to recognize facial expressions in all video frames. For instance, in some frames the face could not be recognized because participants looked sideways, or held their hands against their face. This was the case in 3.85% of all the recorded video frames. These instances were coded as missing values.

The next step was to fuse the video and EEG data. The number of data points per participant differed between the two modalities. CERT recorded a row of measures for every frame in the video corresponding to 15 frames per second. The Emotiv EPOC device recorded 9 rows of EEG data per second. In order to fuse the video and EEG data, the data of each participant were aggregated in 100 subgroups per video for both modalities. This resulted in 100 rows of data per video, per participant.

Table 6. *Experiment procedure*

Step	Event	Experiment tool screenshots
Start	1. Participant reads experiment instructions. 2. Consent to use EEG and video recordings is requested. 3. EEG is placed on participants head. 4. Video recording is started. 5. Participant solves first maze.	Start of experiment: 
Set 1 to 5	1. Participant watches a movie clip. 2. Participant states whether or not he or she had seen the movie clip before. 3. Participant writes summary of movie clip. 4. Participant solves another maze.	Video is playing:  Request to write video summary:  Instruction to solve maze: 
End	1. Participant states gender, age, and command of the English language. 2. Participant is debriefed.	Some final questions: 

Data processing

Following the fusing of the two modalities, correlations between the variables were studied. Then, a mixed effects regression analysis was carried out in order to assess whether the experiment conditions (namely the watching of emotion-eliciting videos) were able to predict the affective states (as measured through the two modalities). The CERT and Emotiv variables reported in Table 5 were used as dependent variables. The videos that participants watched were used as predictors.

Furthermore, classification analyses were performed to gain a better understanding of how well the measured variables facilitate the classification of videos that participants were exposed to. An overview of the variables used as feature in the classification analysis is given in Table 5. Classification analyses were done for each video, for each participant, and for each of the modalities (only the features of CERT or Emotiv) plus the modalities combined (both the features of CERT and Emotiv). Moreover, since it was theorized that the apex of the affective state elicited through the videos that were shown would take place at the end of the videos, the process was repeated for both the full video data, and the last 30% of the video data. Weka software was used for the classification analyses; an SVM classifier (SMO) was selected, with 10-fold cross-validation.

Finally, in order to compare the classification performance between CERT, Emotiv, and the combination of both modalities, the differences in F-measures were studied. Classification performance was compared between the full video data and the last 30% of video data as well. This was done with a factorial repeated-measures ANOVA.

Results

First, it was tested whether the measurement distributions were normal. All the distributions of the measured variables were significantly non-normal (see Table 7). Furthermore, the data were tested for homogeneity of variance. The variances were significantly different between the five videos for Engagement / Boredom, $F(4, 10480) = 21.35$, $p < .001$, Joy, $F(4, 10480) = 102.23$, $p < .001$, Sad, $F(1, 4256) = 43.88$, $p < .001$, Neutral, $F(4, 10480) = 101.13$, $p < .001$.

Table 7. *Test of normality*

Modality	Variable		*D*	*p*
Emotiv	Short Term Excitement	(10485)	0.08	< .001
	Long Term Excitement	(10485)	0.04	< .001
	Engagement / Boredom	(10485)	0.48	< .001
CERT	Digust	(10485)	0.38	< .001
	Fear	(10485)	0.44	< .001
	Joy	(10485)	0.39	< .001
	Sad	(10485)	0.26	< .001
	Neutral	(10485)	0.06	< .001

The mean values of the measured variables of both modalities per condition are listed in Table 8. It can be seen that for Emotiv, the mean Short Term Excitement was highest when participants watched the Neutral video, and lowest when watching the Fear video. This contradicts the given explanation of this variable as “physiological arousal” (see Table 5). Mean Long Term Excitement is highest for Sadness, and lowest for Fear, which fits the concept of sadness being an affective state that develops over time, while the emotion of fear can be experienced more sudden. Engagement / Boredom varied less than the other Emotiv variables and was highest for the Sadness condition. The least amount of Engagement was measured during the Fear and Neutral conditions.

For CERT, the mean value of Disgust was highest when participants watched the Disgust video. This is also the case for the Fear emotion when the Fear video was shown. The mean value of Joy was high in the Amusement condition, but higher when participants watched the Disgust video. During the Neutral condition the mean of Sad was highest. Conversely, the highest mean of Neutral affective states was when participants watched the Sadness video.

A number of observations emerged from the mean values. The measurements of CERT suggest that the Disgust and Fear emotions were experienced least across the experimental conditions. Moreover, a neutral state of affect was measured the most when participants watched videos. For Emotiv, Engagement / Boredom varied only slightly between the conditions. In the next part of this section, the data are explored further, starting with correlations.

Table 8. *Mean values of modality variables, SD in brackets*

Video	Emotiv			CERT				
	Short Term Excitement	Long Term Excitement	Engagement / Boredom	Disgust	Fear	Joy	Sad	Neutral
Amu-sement	0.419 (0.289)	0.419 (0.228)	0.556 (0.031)	0.016 (0.044)	0.004 (0.015)	0.061 (0.163)	0.057 (0.086)	0.325 (0.211)
Fear	0.403 (0.255)	0.401 (0.183)	0.553 (0.023)	0.014 (0.044)	0.014 (0.077)	0.020 (0.073)	0.057 (0.084)	0.341 (0.213)
Disgust	0.413 (0.251)	0.429 (0.202)	0.557 (0.038)	0.022 (0.058)	0.003 (0.013)	0.066 (0.190)	0.061 (0.099)	0.319 (0.232)
Sadness	0.431 (0.266)	0.439 (0.195)	0.559 (0.043)	0.011 (0.035)	0.003 (0.012)	0.016 (0.053)	0.051 (0.078)	0.401 (0.230)
Neutral	0.445 (0.291)	0.425 (0.202)	0.553 (0.022)	0.013 (0.057)	0.003 (0.010)	0.011 (0.037)	0.085 (0.131)	0.378 (0.212)

Correlations

Several significant relations were found between the measured variables of both modalities, see Table 9 for an overview. For Emotive, the strongest positive relationship was found between Short Term Excitement and Long Term Excitement (r = .69). It can be expected that people who experience excitement over a longer time period do also show signs of short time excitement, and vice-versa. For CERT, the Joy emotion was negatively correlated to a Neutral affective state (r = -.25). This is in line with the measurements of Joy, which were lowest when participants watched the Neutral video.

Furthermore, a number of significant but weak relationships were found ($r \leq .10$). Short Term Excitement was positively related to Fear, Sad, and Neutral. Furthermore, Long term excitement was positively related to Fear, Joy, Sad, and Neutral. A negative correlation was found between Long term excitement and both Disgust and Engagement. Conversely, there was a positive relation between Engagement and Disgust, Joy, and Sad. A negative correlation was found between Engagement and Neutral. Disgust was only correlated (negatively) to Neutral. Fear was positively correlated to Sad, and negatively related to Neutral. Finally, a negative relation was found between Joy and both Sad and Neutral.

Table 9. *Correlations between Emotiv and CERT variables, 95% CI in square brackets*

	Emotiv			CERT				
	Short Term Excitement	Long Term Excitement	Engagement / Boredom	Disgust	Fear	Joy	Sad	Neutral
Short term excitement	1	.69 [.680, .702]	-.003 [-.022, .017]	-.02 [-.037, .004]	.07 [.053, .089]	.004 [-.016, .023]	.05 [.026, .068]	.10 [.030, .071]
Long term excitement	*p* < .01	1	-.03 [-.050, -.009]	-.05 [-.043, -.014]	.03 [.016, .043]	.04 [.019, .057]	.05 [.034, .073]	.05 [.026, .066]
Engagement / boredom	*p* = .73	*p* < .01	1	.05 [.032, .065]	-.001 [-.014, .017]	.05 [.020, .085]	.05 [.024, .072]	-.04 [-.061, -.026]
Disgust	*p* = .08	*p* < .01	*p* < .01	1	-.01 [-.015, .012]	.001 [-.008, .012]	.003 [-.012, .018]	-.05 [-.067, -.038]
Fear	*p* < .01	*p* < .01	*p* = .94	*p* = .64	1	.003 [-.004, .011]	.04 [.022, .065]	-.05 [-.066, -.031]
Joy	*p* = .70	*p* < .01	*p* < .01	*p* = .94	*p* = .76	1	-.11 [-.111, -.098]	-.25 [-.260, -.238]
Sad	*p* < .01	*p* < .01	*p* < .01	*p* = .76	*p* < .01	*p* < .01	1	-.002 [-.021, .015]
Neutral	*p* < .01	*p* < .01	*p* < .01	*p* < .01	*p* < .01	*p* < .01	*p* = .80	1

Mixed effects regressions analysis

Table 10 lists the results of the mixed effects regression analysis. Several predictors were found to be significant. For Emotiv, Short Term Excitement was negatively predicted by the conditions of Fear, Amusement, and Disgust. This does not fit the description of Short Term Excitement as "physical arousal", which is expected to be elicited through a condition such as Fear. Long Term Excitement was negatively predicted by Fear and Sadness. This reinforces the concept of sadness as an affective state that develops over time. Amusement and Disgust videos positively predicted Engagement. Moreover, Sadness negatively predicted Engagement (or positively predicted Boredom).

For CERT, the Fear emotion was solely predicted by the Fear video. Joy was positively predicted by Amusement, but also by Fear and Disgust. Conversely, the affective state of Disgust was positively predicted by the Disgust video, and the Amusement video as well. Furthermore, the Sadness video positively predicted the emotion of Sad. The videos of Fear, Amusement, and Disgust negatively predicted Sad. Lastly, the Neutral state of affect was predicted negatively by the Fear, Amusement, Disgust, and Sadness videos.

Table 10. *Mixed effects regression analysis of experienced emotions based on watched videos*

Modality	Emotion	Video	b	*SE* b	df	*t*	*p*	95% CI
Emotiv	Short Term Excitement	Fear	-.042	.008	11358	-5.58	< .01	-.057, -.027
		Amusement	-.025	.008	11358	-3.39	< .01	-.040, -.011
		Disgust	-.031	.008	11358	-4.16	< .01	-.046, -.017
		Sadness	-.014	.008	11358	1.82	.07	-.001, .028
		Neutral[a]						
	Long Term Excitement	Fear	-.025	.005	11358	-4.69	< .01	-.035, -.014
		Amusement	-.006	.005	11358	-1.15	.25	-.016, .004
		Disgust	.003	.005	11358	0.62	.53	-.007, .014
		Sadness	-.013	.005	11358	-2.53	< .05	-.024, -.003
		Neutral[a]						
	Engagement / Boredom	Fear	-.000	.001	11358	-0.01	.99	-.001, .001
		Amusement	.003	.001	11358	3.85	< .01	.001, .004
		Disgust	.004	.001	11358	4.90	< .01	.002, .005
		Sadness	-.005	.001	11358	-7.07	< .01	-.066, -.004
		Neutral[a]						
CERT	Fear	Fear	.011	.001	10465	10.29	< .01	.009, .014
		Amusement	.002	.001	10462	1.64	.10	-.000, .004
		Disgust	.001	.001	10461	1.02	.31	-.001, .003
		Sadness	-.001	.001	10463	-0.47	.64	-.003, .002
		Neutral[a]						
	Joy	Fear	.011	.004	10466	2.93	< .01	.004, .018
		Amusement	.052	.004	10462	14.06	< .01	.044, .059
		Disgust	.056	.004	10462	15.06	< .01	.049, .063
		Sadness	-.006	.004	10464	-1.77	.08	-.014, .001
		Neutral[a]						
	Disgust	Fear	.001	.001	10461	0.98	.33	-.001, .004
		Amusement	.003	.001	10459	2.48	< .05	.001, .006
		Disgust	.011	.001	10459	7.67	< .01	.008, .013
		Sadness	.002	.001	10460	1.60	.11	-.000, .005
		Neutral[a]						
	Sad	Fear	-.029	.003	10459	-11.37	< .01	-.034, -.024
		Amusement	-.027	.003	10459	-10.87	< .01	-.032, -.022
		Disgust	-.023	.003	10459	-8.89	< .01	-.028, -.018
		Sadness	.034	.002	10459	13.68	< .01	.029, .039
		Neutral[a]						
	Neutral	Fear	-.031	.005	10459	-6.27	< .01	-.041, -.021
		Amusement	-.052	.005	10458	-10.65	< .01	-.061, -.042
		Disgust	-.052	.005	10458	-10.64	< .01	-.062, -.043
		Sadness	-.023	.005	10459	-4.70	< .01	-.032, -.013
		Neutral[a]						

[a] neutral is set to zero because it is redundant

Classification analysis

The F-measures of the classification analyses are reported in Table 11. We can make a number of observations based on this data. Firstly, the means across all modalities suggest that when using only the last 30% of video data, classification performance is higher. Secondly, the means of the combination of Emotiv and CERT are higher than CERT-only in all conditions. Thirdly, the means of CERT are higher than Emotiv in all conditions. Furthermore, there seems to be no apparent pattern in classification performance between the experimental conditions. The differences in means were tested for significance which is reported in the following paragraphs.

Table 11. *Mean classification F-measures (and standard deviation) per modality and condition*

	Full video			**Last 30% of video**		
	Emotiv	CERT	Emotiv & CERT	Emotiv	CERT	Emotiv & CERT
Fear	.055 (.181)	.294 (.357)	.383 (.390)	.186 (.357)	.431 (.444)	.592 (.423)
Amusement	.194 (.347)	.304 (.403)	.503 (.418)	.228 (.396)	.412 (.458)	.550 (.478)
Disgust	.056 (.185)	.272 (.385)	.396 (.396)	.174 (.345)	.473 (.437)	.661 (.407)
Sadness	.127 (.267)	.319 (.413)	.445 (.437)	.316 (.419)	.458 (.482)	.572 (.482)
Neutral	.140 (.288)	.441 (.392)	.487 (.399)	.380 (.421)	.569 (.425)	.656 (.405)

Effect of the amount of data on classification performance

First, the classification performance was compared between using the full video data and only the last 30% of video data. This comparison was done for each of the three modalities. For Emotiv, there was a significant main effect of the amount of data on the classification results, $F(1, 22) = 32.40$, $p < .001$. Contrasts revealed that classification performance when using the last 30% of video data was significantly higher than when using the complete video data, $F(1, 22) = 32.40$, $p < .001$, $r = .77$. Furthermore, there was a non-significant main effect of the experiment condition on the classification performance, $F(4, 88) = 1.05$, $p = .386$. Lastly, there was a non-significant interaction effect between video and modality, $F(4, 88) = 1.06$, $p = .382$.

For CERT, there was a significant main effect of the amount of data on the classification performance, $F(1, 22) = 7.26$, $p = .013$. Contrasts revealed that the classification performance when using the last 30% of video data was significantly higher than when using the full video data, $F(1, 22) =$

7.26, p = .013, r = .50. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of experiment condition, $\chi^2(9) = 24.65$, p = .003, and the interaction effect of condition and modality $\chi^2(9) = 27.35$, p = .001. Therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity (ε = .69 for the main effect of condition and ε = .69 for the interaction effect of condition and modality). There was a significant main effect of the video on the classification results, $F(2.76, 60.68) = 3.17$, p = .034. Contrasts revealed that classification performance of fear videos was significantly lower than of the neutral videos, $F(1, 22) = 6.14$, p = .021, r = .47. Conversely, classification performance of amusement videos was significantly lower than of the neutral videos, $F(1, 22) = 9.97$, p = .005, r = .56. The classification performance of sadness videos was significantly lower than of the neutral videos, $F(1, 22) = 8.11$, p = .009, r = .52. Lastly, there was a non-significant interaction effect between experiment condition and modality, $F(2.74, 60.27) = 0.90$, p = .441.

Thirdly, for the combination of Emotiv and CERT, there was a significant main effect of the amount of data that was used on the classification results, $F(1, 22) = 11.51$, p = .003. Contrasts revealed that the classification performance when using the last 30% of video data was significantly higher than when using the full video data, $F(1, 22) = 11.51$, p = .003, r = .59. Furthermore, there was a non-significant main effect of video, $F(4, 88) = 0.34$, p = .850. The interaction effect between video and modality was non-significant as well, $F(4, 88) = 2.62$, p = .055.

In conclusion, there was a significant effect of the amount of data on the classification performance. The classification performance was higher when only the last 30% of the recorded data of the video fragment was used, as opposed to using data of the whole recording. This finding is in line with the idea that participants' emotional experience was stronger towards the end of the watched videos.

Comparison of classification performance between modalities

The classification performance was compared between the three modalities. This was repeated for the full video data, and for the last 30% of video data. For the full video data, Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of modality, $\chi^2(2) = 35.25$, p < .001, and the interaction effect of experiment condition and modality $\chi^2(35) = 81.16$, p < .001. Therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity (ε = .55 for the main effect of modality and ε = .53 for the interaction effect of condition and modality). There was a significant main effect of the modality on the classification results, $F(1.10, 24.26) = 15.63$, p < .001. Contrasts revealed that classification performance of Emotiv-only was significantly lower than performance of the combination of Emotiv and CERT, $F(1, 22) = 21.70$, p < .001, r = .70. Conversely, classification performance of CERT-only was significantly lower than of the combination of Emotiv and CERT, $F(1, 22) = 38.26$, $p < .001$, $r = .80$. Furthermore, there was a non-significant main effect of the experiment condition on the classification performance, $F(4, 88) = 2.11$, $p = .087$. Lastly, there was a non-significant interaction effect between condition and modality, $F(4.24, 93.38) = 0.79$, $p = .544$.

For the last 30% of video data, Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of modality, $\chi^2(2) = 35.31$, $p < .001$. Therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\varepsilon = .55$). There was a significant main effect of the modality on the classification results, $F(1.10, 24.26) = 16.30$, p < .001. Contrasts revealed that classification performance of Emotiv-only was significantly lower than of the combination of Emotiv and CERT, $F(1, 22) = 29.91$, $p < .001$, $r = .76$. Conversely, classification performance of CERT-only was significantly lower than of the combination of Emotiv and CERT, $F(1, 22) = 26.88$, $p < .001$, $r = .74$. Furthermore, there was a non-significant main effect of the video on the classification performance, $F(4, 88) = 1.21$, $p = .313$. Lastly, there was a non-significant interaction effect between experiment condition and modality, $F(8, 176) = 0.83$, $p = .576$.

In summary, there was a significant effect of the modality on the classification performance. Only using Emotiv as a modality yielded the lowest performance. Compared to Emotiv, the performance of CERT as the only modality for affect detection was higher. However, the most striking result to emerge from the data is that combining the Emotiv and CERT data yielded significantly higher classification performance. This finding suggests that EEG data does have an added value for computational affect detection based on facial activity. Lastly, the combination of CERT and Emotiv resulted in significantly higher classification performance both when using the data of the whole video fragment and when using only the last 30% of video data, suggesting the multimodal approach outperformed the unimodal approaches throughout the video.

Discussion

This thesis intended to contribute to the field of affective computing, specifically to the computational recognition of emotions. As has been shown by the literature review, this is a developing area of research. The ability to computationally recognize affective states is improving, partly thanks to developments in technology - both hardware (low cost, less-invasive measurement) and software (machine learning) - and partly thanks to an increasing understanding of affect and affect detection. As for affect detection, most studies have focused on face and speech modalities, while much less is known about EEG as a modality for emotion recognition (D'Mello & Kory, 2015). The growing number of relatively low-cost, less-invasive EEG measurement devices facilitates research on affect recognition performance of EEG. Therefore, this thesis was designed to study EEG as a modality for affect recognition, comparing it to the more common modality of facial expressions. We addressed three questions.

The first question focused on how well EEG analysis performs in emotion recognition. Prior research has shown that EEG can be used to detect affect. Most of this research has been aimed at recognizing emotions on a dimensional scale such as valence and arousal. Nevertheless, there have been studies that showed promising ability recognizing discrete emotional states (Petrantonakis & Hadjileontiadis, 2010). However, the observed classification performance in this study (F-measure between .055 and .380) do not support these findings.

A possible explanation for these results may be the difference in number of affective states that were being detected. The reviewed studies utilizing the same EEG measurement device as in the present study aimed at recognizing no more than three discrete affective states. Recognizing additional discrete emotional states increases the complexity of the task, possibly to the extent that classification performance suffers. However, it has been shown that up to six emotions can be differentiated using EEG (Petrantonakis & Hadjileontiadis, 2010). This raises the possibility that the EEG measurement device used for this study did not have the capability to provide data to accurately recognize five discrete affective states.

Furthermore, it may be that the affect variables that were used in the analyses are not adequate for distinguishing between the emotions. The three variables that were used as features were calculated estimates of affect based on the raw EEG (sensor) data. They represented short-term and long-term excitement and engagement or boredom. Potentially, these variables are more aligned to dimensional affective values (like valence and arousal) which could make the recognition of discrete emotions (such as fear or disgust) more challenging. However, in previous research various methods of calculating affect variables from the EEG data have been applied and shown to be effective in detection of emotions.

Another possible explanation is that the videos do not elicit the emotions strongly enough. If the videos did not - or only weakly - elicit specific affective states, it would not be possible to detect emotions using EEG data (or any other modality for that matter). However, this is disputed by the classification performance of the other modality in this study, which is discussed in the following paragraphs.

The second question sought to determine how EEG compares to facial expression analysis. The experiment results indicate that face recognition is more equipped as a modality for the task of recognizing affective states, as the classification performance of the face modality was significantly higher compared to the EEG modality. This finding is in line with earlier research listed by Valstar et al. (2012), which has shown the possibility to recognize discrete affective states from face expression videos.

The third and final research question was to establish to what extent the combination of modalities (EEG and facial expression recognition analysis) increases the performance of emotion classification. The classification results when data of both modalities were combined were significantly higher than the classification results of either the EEG modality, or the facial expression analysis. Two conclusions can be drawn from this result. Firstly, it supports the finding of D'Mello and Kory (2015) that multimodal affect detection systems outperform unimodal counterparts at recognizing emotions. Secondly, it shows that EEG can be of added value when recognizing emotions using facial expression analysis, even though EEG frequency power is influenced by facial activity (Mühl et al., 2014; Wolpaw et al., 2002).

Furthermore, there was an additional finding pertaining the effect of data selection on emotion recognition performance. At first, all data that were recorded via the two modalities while participants watched the videos were used for the classification analysis. However, it was found that classification performance was higher when only using a part of the recorded data, namely the final 30% (of each video). These results are likely to be related to the temporal segments of emotion. The videos shown to participants were chosen to elicit specific emotional states, but it cannot be expected emotions were elicited from the first second the videos were displayed. Instead, the experienced emotions developed while watching the video. Therefore, the results suggest that emotions are more strongly experienced towards the end of the video, thereby highlighting the importance of data selection for emotion classification.

The findings in this study are subject to a number of limitations. Firstly, affective states were induced via emotion-eliciting video fragments. Although this method more closely approaches a natural display of emotion than letting participants act out specific emotions, the emotional experience may be limited or exaggerated due to the experimental setup. This in turn may have influenced the classification performance. Secondly, the validation method of this study was person dependent. In the analysis, cross-validation was applied within an individual, but not between individuals (assigning participants to either the training group or the test group but not both). An implication of this is the possibility that the test results do not generalize to new individuals. Thirdly, this study attempted to recognize more discrete emotional states than was done before by reviewed studies using the same EEG measurement device. The increased number of emotional states to classify might have had an impact on the relative poor classification performance compared to the face modality, which potentially could have been addressed by collecting data from more individuals. Fourthly, a note of caution is due for the video modality. In this study, facial expression recognition was limited to the frontal view. Profile views of the face and occlusions of facial areas were not taken into account. This is a common limitation of affect detectors using the face as modality.

To conclude, this study has shown the limited extent to which (relatively low cost and less-invasive) EEG measurement can be used for recognition of discrete affective states. To current knowledge, it is one of the few existing studies on recognition of five discrete states of affect using a this type of EEG measurement devices. Moreover, this study confirmed the possibility to detect discrete emotions using facial expressions. Furthermore, it has shown that the combination of EEG and facial expression recognition leads to better classification performance, thereby adding to the growing body of research on multimodal affect detection. To develop a better understanding of affect and affect detection, additional studies will be needed.

Firstly, further work is required to establish the viability of said EEG measurement devices to recognize affective states. Being able to accurately detect emotions with such devices enables to broaden the field of research which is currently dominated by face and voice modalities. Secondly, further research should be undertaken to investigate the amount of EEG variance that can be explained by face muscle movements. This study has shown that EEG data contains additional information that improves emotion recognition performance. There is room for further progress in determining to which extent EEG data enhances affect detection. More knowledge on the incremental value of an EEG modality in emotion detection will also help future research choose whether or not to include this modality in their research setup, as it is a more time-consuming form of measurement compared to recording facial expressions only. Thirdly, this research focused on the recognition of five discrete emotional states. Future studies can further investigate additional discrete emotional states (such as surprise or anger) and dimensional concepts of emotion (such as valence and arousal). Fourthly, the findings suggest that emotions are more strongly experienced towards the end of an video that elicits emotions. Further research on the effect of temporal dynamics of emotions on detection of affective states is needed to confirm this. Fifth and last, the definition of a standardized method of reporting on the performance of affect detection is important. Given the variety of (combinations of) modalities and classification methods, comparing results to those reported in earlier studies is already difficult. The various reporting methods that are used in research on emotion recognition only adds to this complexity. A standardized approach like the one being used here is a first step to solve the issue.

Acknowledgements

First I would like to thank prof. dr. M.M. Louwerse for his advice, enthusiasm and patience throughout the whole process. I am grateful for the insightful and inspiring discussions we had. Even though the lab often required immediate attention you always managed to find time to help. In the Dante building, or via a Skype connection between Tilburg and Dublin.

Then a big thank you to my parents who are and always have been there for me. You have helped me to get to this point from the moment I set foot to De Aanloop, with an incredible dose of love and attention. I hope a MSc degree will help me to answer the phone more often.

A shout out to my friends (thanks Pim for the proofreading and critical notes) and family for your support and for taking my mind of this project more than enough. Now you finally won't be able to make fun of me anymore for not having a *diploma*.

Lastly, a note of appreciation to my current employer for providing me with snacks in the evening hours when I was writing this thesis, a search engine to look up previous studies on affect detection, and a salary so I could pay my tuition fees.

Svevo

Literature

Allard, E. S., & Isaacowitz, D. M. (2008). Are preferences in emotional processing affected by distraction? Examining the age-related positivity effect in visual fixation within a dual-task paradigm. *Aging, Neuropsychology, and Cognition, 15*(6), 725-743.

Arnold, M. B. (1960). *Emotion and Personality*. New York: Columbia University Press.

Aspinall, P., Mavros, P., Coyne, R., & Roe, J. (2015). The urban brain: analysing outdoor physical activity with mobile EEG. *British Journal of Sports Medicine*, *49*(4), 272–276.

Balconi, M., & Mazza, G. (2009). Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues. ERS/ERD and coherence measures of alpha band. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 74(2), 158–165.

Barrett, L. F., Lewis, M., & Haviland-Jones, J. M. (2016). *Handbook of Emotions, Fourth Edition*. New York, NY: Guilford Publications.

Bartolini, E.E. (2011). *Eliciting emotion with film: development of a stimulus set* (bachelor’s thesis). Wesleyan University, Middletown, United States of America.

Bettadapura, V. (2012). Face expression recognition and analysis: The state of the art. *arXiv Preprint arXiv:1203.6722*.

Borod, J. C. (Ed.). (2000). *The neuropsychology of emotion*. New York: Oxford University Press.

Bos, D. O. (2006). EEG-based emotion recognition: The Influence of Visual and Auditory Stimuli. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.226.8188&rep=rep1&type=pdf.

Boström, J. (2017). *Maze Generator*. Retrieved from http://www.mazegenerator.net/

BraInquiry (2017). *brainquiry - Wireless, high quality signal EEG, EMG, ECG*. Retrieved from http://www.brainquiry.com/

Calvo, R., & D’Mello, S. (2010). Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37.

Calvo, R., D’Mello, S., Gratch, J., Kappas, A., & Reisenzein, R. (2015). A Short History of Psychological Perspectives on Emotion. In *The Oxford Handbook of Affective Computing*. New York, NY: Oxford University Press.

Cannon, W. B. (1927). The James-Lange theory of emotions: a critical examination and an alternative theory. *The American Journal of Psychology*, *39*(1), 106-124.

Carvalho, S., Leite, J., Galdo-Álvarez, S., & Gonçalves, O. F. (2012). The Emotional Movie Database (EMDB): a self-report and psychophysiological study. *Applied Psychophysiology and Biofeedback*, *37*(4), 279–294.

Choppin, A. (2000). *Eeg-based human interface for disabled individuals: Emotion expression with neural networks* (master’s thesis). Tokyo Institute of Technology, Tokyo, Japan.

Coan, J. A., & Allen, J. J. B. (2007). *Handbook of Emotion Elicitation and Assessment*. New York, NY: Oxford University Press.

Damasio, A. R., Everitt, B. J., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences, 351*(1346), 1413-1420.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Murray.

Davidson, R. J., & Begley, S. (2012). *The emotional life of your brain*. London, UK: Penguin.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, *13*(4), 247–268.

Dewey, J. (1894). The theory of emotion. *Psychological Review, 1*(6), 553-569.

Dhami, T. S. (2015). *The neural detection of emotion in naturalistic settings* (doctoral dissertation). Queen Mary University of London, London, United Kingdom.

D'Mello, S., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, 47(3), 1–36.

Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., & Dutoit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications. *Biomedical Engineering Online*, *12*, 56.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3), 169–200.

Ekman, P., & Friesen, W. V. (1976). *Measuring facial movement. Environmental psychology and nonverbal behavior, 1*(1), 56-75.

Elkins, A. C., Derrick, D. C., & Gariup, M. (2012). The voice and eye gaze behavior of an imposter: Automated interviewing and detection for rapid screening at the border. *Proceedings of the Workshop on Computational Approaches to Deception Detection*, 49–54.

Emotiv. (2017). *Homepage*. Retrieved from https://www.emotiv.com/

Emotiv. (2014). *EmotivEPOCUserManual2014*. Retrieved from https://emotiv.zendesk.com/hc/en-us/article_attachments/200343895/EPOCUserManual2014.pdf

Frantzidis, C. A., Bratsas, C., Papadelis, C. L., Konstantinidis, E., Pappas, C., & Bamidis, P. D. (2010). Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE Transactions on Information Technology in Biomedicine, 14*(3), 589-597.

Gabert-Quillen, C. A., Bartolini, E. E., Abravanel, B. T., & Sanislow, C. A. (2015). Ratings for emotion film clips. *Behavior Research Methods*, 47(3), 773–787.

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., & Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, *2*(2), 106–118.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, *9*(1), 87–108.

Gunes, H., & Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, *31*(2), 120–136.

Hagemann, D., Naumann, E., Maier, S., Becker, G., Lürken, A., & Bartussek, D. (1999). The assessment of affective reactivity using films: Validity, reliability and sex differences. *Personality and Individual Differences*, *26*(4), 627–639.

Hamann, S., Herman, R. A., Nolan, C. L., & Wallen, K. (2004). Men and women differ in amygdala response to visual sexual stimuli. *Nature neuroscience, 7*(4), 411-416.

Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., & Bartussek, D. (2005). *Cognition & Emotion*, *19*(7), 1095–1109.

Hoque, M.E., El Kaliouby, R., & Picard, R.W. (2009). When human coders (and machines) disagree on the meaning of facial affect in spontaneous videos. In Ruttkay, Z., Kipp, M., Nijholt, A., & Vilhjálmsson, H.H. (Eds.), *Intelligent Virtual Agents*, pp.337–343. Amsterdam, NL: Springer.

Hosseini, S. A., & Naghibi-Sistani, M. B. (2011). Emotion recognition method using entropy analysis of EEG signals. *International Journal of Image, Graphics and Signal Processing*, *3*(5), 30–36.

Interaxon (2017). *MUSE ™ | Meditation Made Easy*. Retrieved from http://www.choosemuse.com/

Inventado, P. S., Legaspi, R., Suarez, M., & Numao, M. (2011). Predicting student emotions resulting from appraisal of ITS feedback. *Research & Practice in Technology Enhanced Learning*, *6*(2), 107-133.

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review,* 2(4), 363-370.

Izard, C. E., Dougherty, F. E., Bloxom, B. M. and Kotsch, N. E. (1974). *The Differential Emotions Scale: A method of measuring the meaning of subjective experience of discrete emotions*. Nashville: Vanderbilt University.

James, W. (1884). What is an Emotion? *Mind, 9*(34), 188-205.

Jasper, H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, *10*, 370–375.

Kim, M.-K., Kim, M., Oh, E., & Kim, S.-P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Computational and Mathematical Methods in Medicine*, 2013. Retrieved from http://www.hindawi.com/journals/cmmm/2013/573734/abs/

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion, 5*(4), 345-379.

Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective Body Expression Perception and Recognition: A Survey. *IEEE Transactions on Affective Computing*, 4(1), 15–33.

Lange, C. G., & James, W. (1922). *The emotions (Vol. 1)*. Philadelphia: Williams & Wilkins.

Ledoux, J. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.

Li, M., & Lu, B. L. (2009). Emotion classification based on gamma-band EEG. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE,* 1223-1226.

Li, S. Z., & Jain, A. (2011). *Handbook of Face Recognition*. New York, NY: Springer.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and brain sciences, 35*(3), 121-143.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). in *Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture*

Recognition, 298–305.

Liu, Y., & Sourina, O. (2012). EEG-based dominance level recognition for emotion-enabled interaction. In *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo*, 1039–1044.

López-Gil, J.-M., Virgili-Gomá, J., Gil, R., & García, R. (2016). Method for improving EEG based emotion recognition by combining it with synchronized biometric and eye tracking technologies in a non-invasive and low cost way. *Frontiers in Computational Neuroscience*, *10*, 85.

MacLean, P. D. (1952). Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Electroencephalography and Clinical Neurophysiology*, 4(4), 407–418.

Malatesta, L., Raouzaiou, A., Karpouzis, K., & Kollias, S. (2007). Towards modeling embodied conversational agent character profiles using appraisal theory predictions in expression synthesis. *Applied Intelligence*, *30*(1), 58–64.

Matiko, J. W., Beeby, S. P., & Tudor, J. (2014). Fuzzy logic based emotion classification. In Proceedings of the *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 4389–4393.

McDougall, W. (2003). *An Introduction to Social Psychology.* Mineola, NY: Dover Publications.

Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review, 6*(4), 303-307.

Mühl, C., Allison, B., Nijholt, A., & Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces*, *1*(2), 66–84.

Murphy, F. C., Nimmo-Smith, I. A. N., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3*(3), 207-233.

NeuroSky (2017). *EEG - ECG - Biosensors.* Retrieved from http://neurosky.com/

Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist*, *66*(3), 171–200.

Pandzic, I. S., & Forchheimer, R. (2003). *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. Hoboken, NJ: John Wiley & Sons.

Pantic, M., Nijholt, A., Pentland, A., & Huanag, T. S. (2008). Human-Centred Intelligent Human-Computer Interaction (HCI^2): how far are we from attaining it? *International Journal of Autonomous and Adaptive Communications Systems*, *1*(2), 168–187.

Pantic, M., & Patras, I. (2006). Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *36*(2), 433–449.

Pantic, M., & Stewart, M. (2007). Machine Analysis of Facial Expressions. In Delac, K. & Grgic, M., (Eds.) *Face Recognition*. Vienna, Austria: I-Tech Education and Publishing, 377-416.

Park, K. S., Choi, H., Lee, K. J., Lee, J. Y., An, K. O., & Kim, E. J. (2011). Emotion recognition based on the asymmetric left and right activation. *McGill Journal of Medicine: MJM: An International Forum for the Advancement of Medical Sciences by Students*, *3*(6), 201–209.

Parrott, W. G. (Ed.). (2001). *Emotions in social psychology: Essential readings*. Philadelphia, PA: Psychology Press.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417–435.

Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Adaptive extraction of emotion-related EEG segments using multidimensional directed information in time-frequency domain. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 1-4.

Pham, T.D., & Tran, D. (2012). Emotion Recognition Using the Emotiv EPOC Device. In Huang T., Zeng Z., Li C., Leung C.S. (Eds.), *Neural Information Processing. ICONIP 2012. Lecture Notes in Computer Science,* 7667, 394-399. Berlin: Springer.

Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition & Emotion*, 7(2), 171–193.

Picard, R. W. (1997). *Affective computing*. Cambridge: MIT press.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *An International Journal on Information Fusion*, *37*, 98-125.

Ramirez, R., & Vamvakousis, Z. (2012) Detecting Emotion from EEG Signals Using the Emotive Epoc Device. In Zanzotto F.M., Tsumoto S., Taatgen N., Yao Y. (Eds.), *Brain Informatics. BI 2012. Lecture Notes in Computer Science, 7670*, 175–184. Berlin: Springer.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of Personality & Social Psychology, 5*, 11-36.

Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence, 20*(1), 23-38.

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1),102-141.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Russell, J. A. & Fernández-Dols, J. M. (1997). *The Psychology of Facial Expression*. Cambridge, UK: Cambridge University Press.

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition & Emotion*, *24*(7), 1153–1172.

Scherer, K. R. (2001). Emotion, the psychological structure of. In Smelser, N. J. & Baltes, P. B. (Eds.), *International encyclopedia of the social and behavioral sciences*. Oxford, UK: Pergamon.

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

Sourina, O., & Liu, Y. (2011). A Fractal-based Algorithm of Emotion Recognition from EEG using Arousal-Valence Model. *BIOSIGNALS*, 209–214.

Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience & Biobehavioral Reviews, 33*(2), 81-88.

Tomarken, A. J., Davidson, R. J., & Henriques, J. B. (1990). Resting frontal brain asymmetry predicts affective responses to films. *Journal of Personality and Social Psychology*, *59*(4), 791–801.

Valstar, M. F., Mehu, M., Bihan Jiang, Pantic, M., & Scherer, K. (2012). Meta-Analysis of the First Facial Expression Recognition Challenge. *IEEE Transactions on Systems, Man, and Cybernetics, (Part B)*, *42*(4), 966–979.

Valstar, M. F., Pantic, M., Ambadar, Z., & Cohn, J. F. (2006). Spontaneous vs. posed facial behavior: automatic analysis of brow actions. In *Proceedings of the 8th international conference on Multimodal interfaces,* 162-170. New York, NY: ACM.

Van den Broek, E. L. (2011). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing*, *17*(1), 53–67.

Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12), 1743–1759.

Viola, P., & Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154.

Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, *129*, 94–106.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *113*(6), 767–791.

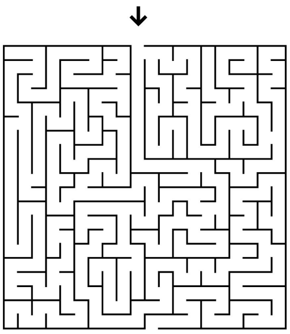
Worcester, W. L. (1893). Observations on Some Points of James's Psychology: II. Emotion. *The Monist*, 3(2), 285-298.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 39–58.

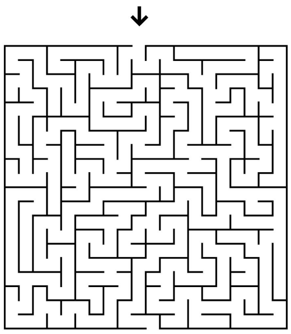
Appendix

Appendix A. Filler task mazes

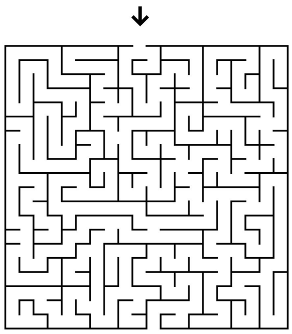
Maze 9LQ



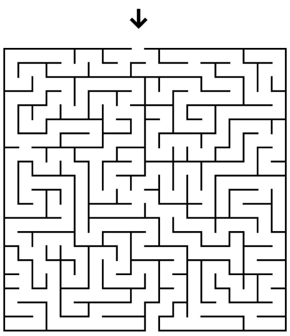
Maze D5P



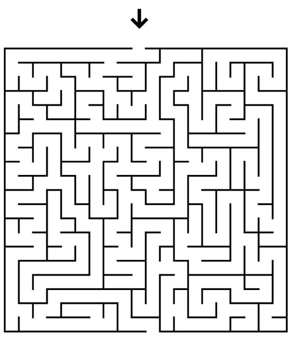
Maze PH7



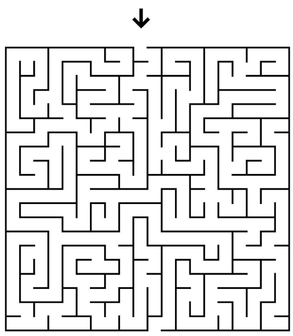
Maze S4F



Maze T1A



Maze XE1



Appendix B. Recorded variables

Table 11. *Variables measured by CERT and Emotiv*

Modality	Variables
CERT	Mouth Imp X, Mouth Imp Y, Left Eye Imp X, Left Eye Imp Y, Right Eye Imp X, Right Eye Imp Y, Mouth Left Corner Imp X, Mouth Left Corner Imp Y, Right Eye Nasal Imp X, Right Eye Nasal Imp Y, Mouth Right Corner Imp X, Mouth Right Corner Imp Y, Right Eye Temporal Imp X, Right Eye Temporal Imp Y, Left Eye Temporal Imp X, Left Eye Temporal Imp Y, Left Eye Nasal Imp X, Left Eye Nasal Imp Y, Nose Imp X, Nose Imp Y, Mouth Left Corner X, Mouth Left Corner Y, Mouth Right Corner X, Mouth Right Corner Y, (AU 1) Inner Brow Raise, (AU 2) Outer Brow Raise, (AU 4) Brow Lower, (AU 5) Eye Widen, (AU 9) Nose Wrinkle, (AU 10) Lip Raise, (AU 12) Lip Corner Pull, (AU 14) Dimpler, (AU 15) Lip Corner Depressor, (AU 17) Chin Raise, (AU 20) Lip stretch, (AU 6) Cheek (AU 7) Lids Tight, (AU 18) Lip Pucker, (AU 23) Lip Tightener, (AU 24) Lip Presser, (AU 25) Lips Part, (AU 26) Jaw Drop, (AU 28) Lips Suck, (AU 45) Blink/Eye Closure, Fear Brow (1+2+4), Distress Brow (1, 1+4), AU 10 Left, AU 12 Left, AU 14 Left, AU 10 Right, AU 12 Right, AU 14 Right, Gender, Glasses, Yaw, Pitch, Roll, Smile Detector, Anger (v3), Contempt (v3), Disgust (v3), Fear (v3), Joy (v3), Sad (v3), Surprise (v3), Neutral (v3)
Emotiv	Time, UserID, Wireless Signal Status, Blink, Wink Left, Wink Right, Look Left, Look Right, Eyebrow, Furrow, Smile, Clench, Smirk Left, Smirk Right, Laugh, Short Term Excitement, Long Term Excitement, Engagement/Boredom, Cognitiv Action, Cognitiv Power

Appendix C. Participant instruction and consent form

Informed Consent

Procedure
Thank you for participating in this study about the effect of visual impressions on analytical capacities. In this study, you are asked to solve a series of mazes. We are interested in what brain regions are active while solving a maze. To distract your mind, a movie fragment will be played after solving each maze.

Recording
Your brain activity will be recorded, and you will be video recorded. These recordings will be stored and analysed. Information about your identity will not be linked to the recordings.

Participation
Participation is voluntary. At any moment during the experiment you can decide to withdraw participation, without needing to specify a reason.

Time involvement
Your participation will take approximately a half hour.

Questions
If you have any questions, please ask them to the experimenter now.

I have read and understood the information stated above, and have been given the opportunity to ask questions. I confirm that …

- ◯ … the recorded data may be used
- ◯ … the recorded data may be used only for this study
- ◯ … the recorded data may not be used

NAME DATE SIGNATURE

____________________ _______________ ____________________

The researcher declares that the participant has received the information needed and that questions have been answered.

NAME DATE SIGNATURE

____________________ _______________ ____________________