

Emotionserkennung mit Hilfe von Bio-Sensoren

Andreas Kamenz¹

HAW Hamburg, Berliner Tor 7, 20099 Hamburg, Germany
andreas.kamenz@haw-hamburg.de

Abstract. Um Intelligente Systeme mit der Fähigkeit auszustatten, menschliche Emotionen zuverlässig zu erkennen, gibt es in der Wissenschaft zahlreiche Lösungsansätze basierend auf video-, audio- und/oder physiologische Daten des Menschen. In diesem Artikel wird die Emotionserkennung anhand der körperlichen Daten behandelt. In diesem relativ neuen Forschungsbereich fehlen momentan noch robuste, ausgereifte und im täglichen Leben einsetzbare Sensoren und Technik. Solche Verfahren haben allerdings durch die schwierige Beeinflussbarkeit der physischen Werte ggf. die besten Zukunftsperspektiven.

Keywords: emotion recognition, bio-sensors, physiological signals, eda, ecg

1 Einführung

Ein digitaler Haushaltshelfer, der die passende Musik und Beleuchtung anschaltet, wenn sein Besitzer spontan Besuch hat, oder eine Smartphone-App, die den Nutzer zu seinen Lieblingsorten navigiert - solche Technologien sind heute schon Realität. Diese intelligenten Systeme, auch Smart Systems genannt, werden in Zukunft eine immer größere Rolle spielen und immer mehr Menschen bei der Bewältigung ihres Alltags unterstützen.

Sollten sie in Zukunft im Stande sein, sich an die individuellen Fähigkeiten sowie aktuellen Bedürfnisse des jeweiligen Benutzers anzupassen und sich nach seiner emotionalen Lage zu richten, werden sie zu wahren Companion Systemen, die als vertrauenswürdige Partner an der Seite des Benutzers zur Verfügung stehen.

Für solch ein System ist es wichtig, den emotionalen Zustand einer Person zu ermitteln, damit eine möglichst passende Reaktion und dadurch eine gute Unterstützung gewährleistet werden kann. Dennoch ist das Erkennen von Emotionen für Companion-Systeme eine große Herausforderung. In der Wissenschaft gibt es verschiedene Lösungsansätze zur Emotionserkennung, die unter anderem auf Audio, Video oder zunehmend auch körperlichen Daten basieren. Außerdem gibt es sogenannte multimodale Verfahren, die die genannten Datenquellen kombinieren. Diese Verfahren zeigen zur Zeit unterschiedliche Forschungsfortschritte auf und sind dementsprechend unterschiedlich leistungsfähig in der Erkennung.

Die Verwendung von körperlichen Daten zur Emotionserkennung gehört zu den neuesten Forschungsrichtungen und zeigt sich aktuell als eine vielversprechende und sinnvolle Ergänzung der anderen Datenquellen. Ihr größter Vorteil liegt daran, dass solche physiologischen Werte wie Herzfrequenz, Blutdruck, elektrische Leitfähigkeit der Haut oder Körpertemperatur nicht einfach vorgetäuscht werden können und unabhängig von kulturellen und äußerlichen Unterschieden der Menschen sind.

Beim Vergleich der Lösungsansätze spielt jedoch nicht nur die Datenquelle, sondern auch das verwendete Dataset eine wichtige Rolle. Ohne ein einheitliches Dataset kann keine Aussage getroffen werden, welches Verfahren bessere Trefferquoten bei der Emotionserkennung liefert. Einen Versuch zur Gewährleistung einer besseren Vergleichbarkeit stellt das AFEW Dataset [14] dar, bestehend aus Audio- und Videodaten. Ab 2016 sollten auch physiologische Daten hinzugefügt werden, was aber bisher noch nicht erfolgte.

Mögliche Einsatzgebiete von automatisierter Emotionserkennung sind Spiele (Schwierigkeitsgrad anpassen [31]), Smart Home (je nach Stimmung passende Musik spielen oder Licht anpassen) [7], Health Care, Fitness, Bildungseinrichtungen/ eEducation/eLearning (wie Schule, Kindergarten, Uni) [36] und Psychologie (automatisierte Depressionserkennung) [2].

2 Grundlagen

2.1 Emotionsmodelle

Im Bereich der Emotionserkennung ist es wichtig, im Vorfeld zu klären, was unter dem Begriff *Emotion* verstanden wird. Folgende Fragen sind daher von zentraler Bedeutung: Wie viele Emotionen gibt es, welche Merkmale haben sie und wie sind sie voneinander abzugrenzen?

Im Allgemeinen sind Emotionen kurzzeitige, aber intensive Empfindungen und sind von den langfristiger wirkenden Stimmungen zu unterscheiden. Darüber hinaus haben Emotionen einen eindeutigen erkennbaren Auslöser und können positiv oder negativ wirken.

Eine konkretere Definition ist nach Scherer [35] folgende:

"Emotion ist [...] eine Antwort auf die Bewertung eines externalen oder internalen Reizereignisses [das] als bedeutsam für die zentralen Bedürfnisse und Ziele des Organismus [gewertet wird]."

Nach Plutchik [33] existieren über 90 weitere wissenschaftliche Definitionen zum Begriff *Emotion*, und zwar allein aus dem 20. Jahrhundert. Diese sind ein kontroverses Thema, das seit jeher intensiv diskutiert wird. Allein diese Diskussion würde den Rahmen der Ausarbeitung sprengen, deswegen wird auf die Publikationen von Kleinginna [24] und Cowie [10] verwiesen.

In Fig. 1 wird verdeutlicht, wie eine Emotion abläuft. Laut Plutchik sind Emotionen menschliche Feedbackschleifen, die anfangs durch ein Reizereignis (z.B. Bedrohung) ausgelöst werden. Daraufhin erfolgt die Wahrnehmung des Reizereignisses, die auf persönliche Erfahrung oder Instinkt basiert (in Gefahr). Dies löst einen Gefühlszustand (Angst) aus und sorgt für eine körperliche Erregung (Muskelanspannung). Nachfolgend findet ein Handlungsimpuls (Wille wegzulaufen) statt. Dieser Impuls wird dann mit einem sichtbaren Verhalten (Weglaufen) umgesetzt und mündet schließlich in einer konkreten Auswirkung der Handlung (sich verringernde Bedrohung). Dieser Prozess wird solange in einer Schleife durchlaufen, bis das Reizereignis nicht mehr wahrnehmbar ist, und es erfolgt eine Entspannung (in Sicherheit).

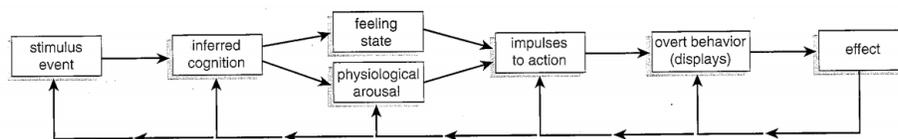


Fig. 1. Menschliche Feedbackschleife (Emotion) [33]

Für die Klassifizierung von Emotionen wurden in der Psychologie verschiedenste Modelle entwickelt. Unter den Modellen, die sich mit dem Inhalt der Emotionen beschäftigen, haben sich zwei konträre Gruppen herauskristallisiert.

Dazu zählen auf der einen Seite diskrete Modelle, die von einer diskreten Menge an Basisemotionen ausgehen. Hierbei wird einer Person eine oder mehrere dieser Basisemotion zugeordnet. Ein weit verbreitetes diskretes Modell ist beispielsweise das Emotionsmodell nach Ekman [11]. Es besteht aus 6 Basisemotionen (vgl. Tab. 1). In späteren Publikationen [12], [13] wurde die *Verachtung* als zusätzliche Kategorie hinzugefügt.

Table 1. Menschliche Grundemotionen nach Ekman [11]

| | | | |
|--------|-------------|--------------|--------------|
| Wut | Ekel | Furcht | (Verachtung) |
| Freude | Traurigkeit | Überraschung | |

Ein System, welches dieses Emotionsmodell einsetzt, ist das Facial Action Coding System (FACS) [12], ebenfalls von Ekman entwickelt. Dabei werden Gesichtsausdrücke in Action Units (AU) zerlegt und kategorisiert. In vielen Algorithmen wird zudem noch der "emotionslose" Zustand hinzugefügt, um einen Grundzustand einzuführen.

Neben den diskreten Modellen existieren auf der anderen Seite die kontinuierlichen Modelle. Diese klassifizieren die Emotionen nicht in Kategorien, sondern bilden sie auf Punkte in einem mehrdimensionalen Raum ab. Zu den Vertretern dieser Theorie gehört u.a. Wundt [47]. Nach seinen Ansichten ist die Gefühlswelt nicht die Summe im Wesentlichen unverändert bleibender Elementargefühle. Vielmehr sei der qualitative Gefühlsreichtum unerschöpflich, und somit ist keine Kategorisierung in Basisemotionen möglich.

Wundt stellt 3 Dimensionen als qualitative Gefühlsfärbungen auf, und zwar mit folgenden Achsen: *Lust - Unlust*, *Erregung - Beruhigung* und *Spannung - Lösung*. Dabei ist "Lust - Unlust" die Qualitätskomponente, "Erregung - Beruhigung" die Intensitätskomponente und "Spannung - Lösung" die Zeitkomponente einer Emotion. Genauer gesagt, entsteht im zeitlichen Verlauf Spannung, bevor eine Emotion eintritt, und Lösung, nachdem die Emotion eingetreten ist.

Zur Vereinfachung des Modells von Wundt wird nach Russel [34] die Zeitkomponente weggelassen, und die verbleibenden Achsen sind Erregung (Arousal) und Wohlfühlen (Valence) (vgl. Fig. 2). Darin ist ersichtlich, dass dem Koordinatensystem an bestimmten Punkten Emotionen zugeordnet wurden.

Der Vorteil von kontinuierlichen Modellen mit 2 Dimensionen ist die einfache Darstellung und das Aufzeigen von Ähnlichkeit oder Diversität durch die räumliche Entfernung der Emotionen im Koordinatensystem.

Emotion wie Freude aus. Dazu sollte der Fragende jedoch Hintergrundwissen zum Probanden haben, um entsprechende emotionsauslösende Fragen stellen zu können.

Ein weiterer häufig verwendeter Ansatz ist es, dem Probanden (multi-)mediale Inhalte wie Fotos, Musik oder Videos vorzuspielen und dadurch Emotionen auszulösen [38], [21].

Darüber hinaus besteht eine mögliche Vorgehensweise darin, dem Probanden selbst agieren zu lassen, und zwar entweder in der realen Welt mit Objekten und Personen oder in einer virtuellen Welt [42].

Im Rahmen des Projektes EmotionBike fiel die Entscheidung auf den Ansatz mit Interaktion in der virtuellen Welt. Der Grund dafür ist, dass die Probanden einerseits an das Geschehen gefesselt werden, aber andererseits nicht zu großen Bewegungsfreiraum lässt.

2.3 Emotionserkennung

Trotz allem Fortschritt von Intelligenten Systemen bzw. Companion Systemen in den letzten Jahren bleibt die Emotionserkennung nach wie vor das größte Problem in diesem Bereich. Aufgrund der Komplexität des menschlichen Verhaltens ist es für diese Systeme schwierig, richtige Schlussfolgerungen daraus zu ziehen, und es kommt oft zu Missverständnissen.

Ein Mensch kann auf viele verschiedene Arten seine Emotionen ausdrücken. Die Ausdrucksweisen sind jedoch laut Holodynski [19] sehr stark von seinem Charakter, der Sozialisation und vom aktuellen Kontext abhängig, was wesentlich zur Komplexität der Aufgabe beiträgt.

Untersuchungen von Mehrabian [29] ergaben, dass Emotionen bei der Kommunikation über verschiedene Kanäle verbreitet werden. Und zwar zu 55% über Mimik, 38% über die Stimme und 7% über Sprache (textlichen Inhalt).

Dabei haben Mimik, Gestik und Körperhaltung zwar einen sehr großen Anteil, allerdings ist es für videobasierte Emotionserkennungsalgorithmen schwierig, immer zuverlässige Aussagen zu treffen. Ein Grund dafür ist, dass das menschliche Aussehen stark variieren kann und dass es in verschiedenen Kulturen verschiedene Ausdrucksweisen gibt [33]. Ein weiteres Problem ist, dass die Forschung der menschlichen Ausdrucksweise stark auf den europäischen und nordamerikanischen Kulturkreis ausgerichtet war, meist auch nur auf eine Sozialschicht. Aber auch aufgrund des zunehmenden Interesses im asiatischen Raum in den letzten Jahrzehnten wurden die Algorithmen bereits verbessert bzw. haben durch die neuesten Erkenntnisse im Bereich Deep Learning [27] noch viel Verbesserungspotenzial.

Der zweitwichtigste Kanal, über den Emotionen ausgedrückt werden, ist die Stimme. Als Mittel dienen unter anderem Tonhöhe, Geschwindigkeit oder Har-

monie in der Aussprache. Jedoch muss die Person ihre Emotionen auch mündlich kommunizieren, damit diese Algorithmen Eingangsdaten bekommen. Wenn eine Person aus verschiedenen Gründen (z.B. Krankheit, Depression, Angst, Scham) nicht sprechen möchte oder kann, dann ist auch keine Emotionserkennung mit diesen Verfahren möglich [3].

Die am schwierigsten zu manipulierende Art, Emotionen zu ermitteln, ist über die physiologischen Daten eines Menschen. Denn diese sind am wenigsten durch die Person selber beeinflussbar und damit sehr aussagekräftig. Das Problem stellt jedoch die genaue Erfassung der Daten mittels Bio-Sensoren dar, und zwar ohne Rauschen und Störungen. Außerdem dürfen die Sensoren nicht zu stark einschränken, da es sonst zu Einflüssen auf die Emotion als solches kommen kann. Beispielsweise darf es nicht passieren, dass nicht mehr die Emotion der Freude über ein Ereignis im Vordergrund steht, sondern der Ärger über unbequeme Sensoren.

Im Folgenden werden die jeweiligen "State-of-the-Art"-Lösungsansätze vorgestellt.

Videobasierte Verfahren. Videobasierte Emotionserkennungsverfahren verwenden als Trainingsdaten entsprechende Videoaufnahmen, die einzelne oder mehrere Personen beinhalten, von denen bestimmte Emotionen erlebt werden. Nach dem Training kann die Erkennungsrate während einer Testphase ermittelt werden.

Als Beispiel einer solchen Trainingdatenbank dient das CK+ Dataset, das von idealen Bedingungen ausgeht. So zeigen die dazugehörigen Videos Personen, die Emotionen vorspielen, und zwar einmal mit frontaler Ansicht und einmal um 30° versetzt. Auf diesem Dataset wurden Erkennungsraten von über 90% dokumentiert [8].

Seit 2013 findet ein jährlicher Wettbewerb um die besten Erkennungsraten in realitätsnahen Situationen statt, das sogenannte EmotiW (Emotion Recognition in the Wild [14]) im Rahmen der ACM International Conference on Multimodal Interaction [1]. Dabei besteht das Dataset "Acted Facial Expressions in the Wild" (AFEW) aus Videoclips, die aus Kinofilmen geschnitten wurden. Für die 2016er Challenge wurde das Dataset um Reality TV Clips erweitert. Dies soll das Dataset realitätsnäher machen, da keine Schauspieler, sondern Laien die Emotionen verkörpern. Insgesamt bietet das Dataset eine stark heterogene Datenbasis, d.h. es gibt Videosequenzen mit oder ohne Bewegung der Kamera, verschiedenste Lichtverhältnisse, Aufnahmen in Gebäuden oder im Freien, freie Körperbewegungen der Person, andere Personen und Gegenstände im Bild usw. Es stellt damit einen Kontrast zu Datensets dar, die in Laborumgebungen entstanden sind. Klassifiziert werden soll dabei in die 7 Emotionen nach Ekman: *Wut, Ekel, Furcht, Freude, Traurigkeit, Überraschung* und *Neutral*.

Die besten Erkennungsraten lagen in den letzten Jahren im Bereich von 50%. Zur Erkennung auf Grundlage der Videodaten werden hauptsächlich Convolutional

Neural Networks (CNN) eingesetzt. Ausführlichere Details zu dem Wettbewerb und den Siegereverfahren werden in den entsprechenden Veröffentlichungen behandelt (siehe [8], [46], [16], [48]). Weiterführende Artikel zu videobasierten Verfahren sind Bernin [4] und Corneanu [9].

Audiobasierte Verfahren. Audiobasierte Verfahren nutzen die Stimme als Grundlage für die Emotionserkennung. Sie haben allerdings den Nachteil, dass ein großer Teil des emotionalen Ausdrucks einer Person verloren geht (Mimik/Gestik). Außerdem ist die Ausdrucksweise auch stark von der jeweiligen Sprache bzw. Sprachfamilie abhängig. Es sind bisher zwei Drittel aller Sprachen auf ihre spezifischen Eigenschaften untersucht. Somit gibt es keine leistungsfähigen Verfahren, die generalisiert für alle Sprachen der Welt eine gute Erkennungsrate aufweisen.

Auch für audiobasierte Verfahren existieren diverse Datasets, die immer nur eine bestimmte Sprache beinhalten. In der Publikation von Feraru [17] werden die 8 verbreitetsten Datasets gegeneinander verglichen. Als Verfahren wurde eine Featureextraktion basierend auf 31 Low Level Descriptors (Lautstärke, Harmonie, etc.) und 42 statistische Funktionen sowie eine Support Vector Machine (SVM) als Klassifikator eingesetzt. Klassifiziert wurde jeweils nach Arousal (positiv/negativ) und Valence (positiv/negativ). Dabei wurde es mit einer Datenbank trainiert und mit einer anderen getestet.

Das Ergebnis zeigt auf, dass die Erkennungsraten bei Training und Test mithilfe derselben Sprache sich an 100% nähern, bei zwei unterschiedlichen Sprachen jedoch stark abfallen.

Verfahren für physiologische Sensordaten. Die rasante Verbreitung von mobilen Endgeräten wie Smartphones und Wearables und deren Einsatz in sich ständig ändernden Umgebungen haben auch in der Forschung zu einer stärkeren Beachtung von anderen Datenquellen als Video- und Audiosignalen geführt. Die Mensch-Maschine-Interaktion findet nicht mehr nur fest mit einem Terminal statt, sondern es erfolgt in einem dynamischen Umfeld. Physiologische Verfahren haben den Vorteil, dass auch weitere körperliche Zustände (z.B. Ermüdung oder Überanstrengung) erkannt und in die Bewertung mit einfließen können.

Für die Messung von physiologische Daten zur Emotionserkennung wurden bisher häufig EEG oder ein multimodaler Ansatz mit EEG und weitere periphere Sensoren verwendet, beispielsweise das DEAP Dataset [26]. EEG hat jedoch das Problem, dass es sehr genau positioniert und kalibriert werden muss und nur im Laborumfeld nutzbar ist, keinesfalls im täglichen Leben.

Für die EmotiW 2016 Challenge war geplant [41], physiologische Daten mit zu integrieren, was aber nicht umgesetzt wurde. Es sollte als weitere Eingabequelle dienen, und die Erkennungsraten der video- und audiobasierten Verfahren verbessern. Da es im Allgemeinen schwierig ist, ein großes Dataset mit physiologischen Daten

zu generieren, liegt die Vermutung nahe, dass es für diese Challenge bisher zu komplex war, ein geeignetes Dataset zu entwickeln.

Bekannte Verfahren wie von Haag et. al [18] oder Kim et. al [23] sind mit einem multimodalem Sensorverbund und beide mit rudimentärem Emotionsmodell, welches positive und negative Valence und Arousal klassifizieren können, ausgestattet. Dabei werden von beiden Forschungsteams Erkennungsraten von 90% erreicht. Hierbei ist jedoch ein weiterer Schritt im Verfahren nötig, um Valence und Arousal auf den einzelnen Emotionen abzubilden. Denn in jedem Quadranten des Koordinatensystems von kontinuierlichen Emotionsmodellen sind mehrere Emotionen vorhanden und mit diesen beiden Verfahren, ist eine genaue Zuordnung nicht möglich.

Letztendlich sind alle Verfahren stark vom jeweiligen verwendeten Trainingsset und dem Emotionsmodell abhängig. Dies erschwert den direkten Vergleich zwischen audio- und videobasierten Verfahren, aber auch zwischen den Verfahren innerhalb einer Kategorie. Außerdem sind viele Datasets nicht repräsentativ genug oder bilden nur einen kleinen Teil der Realität ab.

Die Forschung ist also nicht abgeschlossen und es besteht weiterhin der Bedarf nach aussagekräftigen Datasets, die möglichst viele Datenkanäle (Audio-/Video- oder physiologische Daten) gleichzeitig beinhalten. Deshalb wird viel Wert darauf gelegt, im Rahmen des EmotionBike Projektes ein eigenes Dataset zu kreieren.

3 EmotionBike

3.1 Systemdesign

Um die Emotionserkennung zu gewährleisten, sollen im Rahmen eines Feldversuchs Daten von Probanden erfasst und ausgewertet werden. Dazu werden die Probanden in eine virtuelle Welt versetzt, in der sie verschiedene Levels mit einem Fahrrad durchqueren und an bestimmten Punkten bestimmte Herausforderungen bewältigen müssen. Über den gesamten Zeitraum werden Daten vom Probanden erfasst, und anschließend wird ermittelt, wie sie sich während eines solchen Ereignisses verändern, um damit Rückschlüsse auf die empfundene Emotion zu bekommen.

Um die Ergebnisse vergleichbar zu machen, werden die Probanden in den verschiedenen Levels an festgelegten Stellen durch bestimmte Ereignisse beeinflusst und in einen erwarteten emotionalen Zustand versetzt. Dies erfolgt durch plötzliches Auftreten von Schockmomenten oder schwierigen Herausforderungen, wie das Überspringen eines Abgrunds. Nach jedem Level wird zusätzlich ein Fragebogen vom Probanden ausgefüllt, um für die spätere Auswertung auch den subjektiven Eindruck über den emotionalen Zustand zu haben.

Der Versuchsaufbau besteht aus einem modifizierten Ergometer als Eingabegerät, einem Flachbildschirm zur Anzeige und aus verschiedenen Sensoren, um den aktuellen Zustand eines Probanden zu erfassen.

Das Ergometer wurde neben den standardmäßigen Pedalen zur Beschleunigung mit einem Lenker zur Richtungsänderung [20] und mit Bremsgriffen [39] ausgestattet. Dies ermöglicht eine realitätsnahe Bedienung des Spielgeschehens in der virtuellen Welt und erreicht somit eine hohe Immersion des Probanden.

Genauere Ausführungen zum Versuchsaufbau sind den Veröffentlichungen [22] und [30] zu entnehmen.

3.2 Hypothese

Es wird die Hypothese aufgestellt, dass allein mit physiologischen Sensordaten eine Emotionserkennung nach dem Modell mit 7 Grundemotionen von Ekman möglich ist. Dazu wird ein Versuchsaufbau realisiert, der die benötigten Daten liefern soll. Es werden von 20 Probanden die Daten aufgezeichnet, um ein ausreichend großes Dataset zu erhalten. Die gewonnenen Daten werden entstört, gefiltert und normalisiert. Anschließend sollen sie zum Training und zum Testen eines noch genauer zu spezifizierenden Verfahrens eingesetzt werden.

Zusätzlich soll im Gesamtprojekt EmotionBike die mögliche Genauigkeit der Emotionserkennung erhöht werden, indem die physiologischen Daten als weitere Datenquelle zur Entscheidungsfindung hinzugezogen wird. Wenn die Videodaten keine eindeutige Zuordnung zulassen, dann können die physiologischen Daten bei der Klassifizierung den Ausschlag geben.

Nach Abschluss des Versuchs soll eine Auswertung der Daten erfolgen, indem die vom jeweiligen Klassifikator (Video, physiologische Daten) ermittelten Emotionen mit den erwarteten Emotionen verglichen werden.

Die Auswahl der Sensoren basiert auf den Erkenntnissen von Levenson [28]. Demzufolge lösen Emotionen im autonomen Nervensystem unterbewusst bestimmte Prozesse aus. So werden beispielsweise durch ein angst- oder wutauslösendes Reizereignisse die Gliedmaßen stärker durchblutet und dadurch steigt die Temperatur an. Außerdem wird die Atmung beeinflusst, Blutdruck und Puls steigen an, es wird Schweiß abgesondert, wodurch der Hautleitwiderstand sinkt und es kann zu nervösen Bewegungen kommen.

Im Folgenden werden die eingesetzten Sensoren und ihre zu erwartende Relevanz beschrieben.

Temperatur In der Studie von Nummenmaa et. al [40] zur Emotionserkennung mittels Ganzkörperwärmebildern wurde nachgewiesen, dass Erkennungsraten von 38% erzielt werden können. Beispielhafte Wärmebilder sind in Fig. 3 dargestellt.

Im Projekt wird der Temperatursensor am linken Unterarm angebracht, da der Oberkörper bereits per Thermalkamera aufgenommen wird und somit diese Stelle als am besten geeignet angesehen wird, um einen weiteren Temperaturwert zu erhalten. Zu untersuchen bleibt, ob die Daten von einem Temperatursensor ausreichend sind, um verlässliche Rückschlüsse auf die Emotion zu erhalten oder ob weitere Temperatursensoren an anderen Körperstellen notwendig sind (vgl. Kletz und Kleimann [25]).

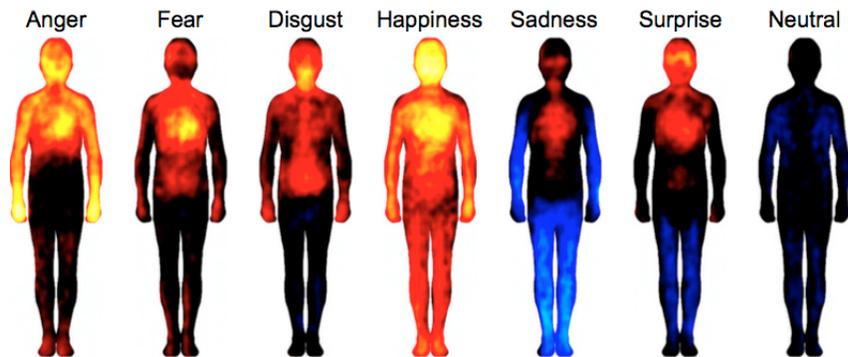


Fig. 3. Ganzkörperwärmebilder für Basisemotionen [40]

ECG und BVP Das Elektrokardiogramm zeichnet ähnlich zum BVP-Sensor den Herzschlag auf. Dies erfolgt anhand der Herzmuskelaktivität. Cai et. al [5] zeigen, dass bei einem Emotionsmodell mit zwei Emotionen mit Hilfe der ECG-Daten eine Erkennungsrate bei der Klassifizierung von Freude (zu 80%) und Traurigkeit (zu 90%) möglich ist.

Bei jedem Pulsschlag des Herzens wird Blut durch die Venen gepresst. Mit dem Blood Volume Pulse Sensor kann somit auf Grundlage des Blutflusses der Puls ermittelt werden. Diese Werte werden als Zweitwert zum ECG verwendet. Es dient also nur als Backup, falls die ECG-Werte fehlerhaft sind.

EDA Die elektrodermale Aktivität wird maßgeblich durch Stress und Erregung beeinflusst und es soll sich nach Haag et. al [45] gut zur Unterscheidung von stressigen- bzw. nichtstressigen Situationen und Wut bzw. Angst eignen. Physiologisch erfolgt dies dadurch, dass die Schweißbildung angeregt wird. Nach Taylor [44] sind EDA-Werte bei Emotionen sehr charakteristisch ausgeprägt. So steigt der Wert sehr schnell an und fällt anschließend gleichermaßen schnell wieder ab, wodurch zeitversetzt zum Reizereignis Peaks entstehen (vgl. Fig. 4).

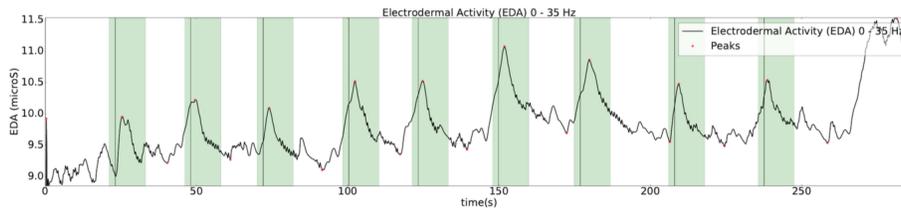


Fig. 4. EDA-Peaks bei bestimmten Reizereignissen [30]

Atmung Mittels des Atmungssensors kann ermittelt werden, wie schnell bzw. tief geatmet wird. Es liefert laut Wagner [45] Rückschlüsse auf eine eingetretene Emotion.

Bewegung Da bedingt durch den Versuchsaufbau keine freie Bewegung des Probanden im Raum stattfinden soll, werden die Bewegungsdaten auch nur dazu eingesetzt, zu erkennen, wann bei den anderen Sensoren Störungen durch Bewegungen auftreten könnten.

EMG Bei einem rein auf Biosensoren bestehenden Versuchsaufbau wäre es außerdem interessant Muskelbewegungen im Gesicht mittels EMG-Sensoren zu erfassen. Da es im Projekt jedoch die Videoaufnahme und die Thermalkamera beeinträchtigen würde, wurde darauf verzichtet.

Alle physiologischen Daten sind einzeln betrachtet nicht ausreichend, um Emotionen erkennen zu können. Jedoch wird die ausgewogene Kombination aller Daten das entscheidende Kriterium sein, um eine verlässliche Erkennung zu erhalten. Es stellt sich dennoch die Frage, ob allein mit physiologischen Daten eine zuverlässige Emotionserkennung möglich ist oder ob diese Daten lediglich in Kombination mit anderen Verfahren die Erkennungsrate verbessern können. Zu klären wäre im letzteren Fall, wie die Kopplung mit anderen Verfahren zu gestalten ist.

Ein Risiko könnte sein, dass die Sensoren bei manchen Probanden aufgrund der individuellen körperlichen Situation zu schlechte Daten liefern oder falsch eingesetzt wurden.

Es kann passieren, dass die erwarteten Emotionen bei bestimmten Ereignissen in den Levels gar nicht bei allen Probanden auftreten, sodass bei jedem Proband unterschiedliche oder auch gar keine Emotionen vorkommen. Um das Problem jedoch zu erkennen und in den weiteren Auswertungen zu berücksichtigen, wurden spezielle Fragebögen nach jedem Level von den Probanden ausgefüllt.

4 Zusammenfassung

Es wurde aufgezeigt, dass physiologische Sensoren immer besser in der Praxis eingesetzt werden und eine vielversprechende Ergänzung bzw. Ersatz zu den bestehenden Lösungsverfahren sein können. Dies wird erreichbar durch die fortschreitende Miniaturisierung und Entwicklung von entsprechender Sensorik und Übertragungstechnik, aber auch von immer besseren Algorithmen zur Erkennung von Emotionen, im speziellen durch die Möglichkeiten des Deep-Learnings.

Für die Akzeptanz von Companionsystem wird es entscheidend sein, wie gut sie menschliche Emotionen erkennen können. Somit ist es wichtig, dass es gute, für den jeweiligen Einsatzzweck angepasste Algorithmen gibt. Mit dem Hauptprojekt und der folgenden Masterarbeit soll dazu ein Beitrag geleistet werden.

References

1. ACM International Conference on Multimodal Interfaces (ACM ICMI), URL: <http://icmi.acm.org/2016/>
2. Alghowinem, S., Goecke, R., Wagner, M., Parker, G., Breakspear, M., Head pose and movement analysis as an indicator of depression. In: Active Computing and Intelligent Interaction (ACII), Humaine Association Conference on IEEE, pp. 283-288, (2013)
3. Anagnostopoulos, C. N., Iliou, T., & Giannoukos, I., Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011, *Artificial Intelligence Review*, 43(2), pp. 155-177, (2015)
4. Bernin, A., et. al, Towards More Robust Automatic Facial Expression Recognition in Smart Environments, PETRA '17, ACM. ISBN123-4567-24-567/08/06. (2017)
5. Cai, J., Liu, G., & Hao, M., The research on emotion recognition from ECG signal. In *Information Technology and Computer Science*, 2009. ITCS 2009. International Conference on (Vol. 1, pp. 497-500). IEEE, (2009)
6. Chanel, G., Ansari-Asl, K., and Pun, T. Valence-arousal evaluation using physiological signals in an emotion recall paradigm. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Montreal, Quebec, Canada (pp. 2662-2667). Washington, DC: IEEE Computer Society. (2007)
7. Chaudhuri, S., Demiris, G., Thompson, H.J., Translating Smart Home Research into Practice, In book: *Handbook of Smart Homes, Health Care and Well-Being*, pp.555-565 (2016)
8. Chen, Jinhui; Takiguchi, Tetsuya; Arika, Yasuo, Facial Expression Recognition with Multithreaded Cascade of Rotation-invariant HOG, In: *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 155-168, (2015)
9. Corneanu, C. A., Simón, M. O., Cohn, J. F., & Guerrero, S. E., Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE transactions on pattern analysis and machine intelligence*, 38(8), pp. 1548-1568, (2016)
10. Cowie, R., Sussman, N., and Ben-Ze'ev, A., Emotion: Concepts and definitions. In *Emotion-oriented systems* (pp. 9-30). Springer Berlin Heidelberg, (2011)
11. Ekman, P., Friesen, W. V., Ellsworth, P., *Emotion in the human face*. Pergamon Press, New York (1972)

12. Ekman, P., Friesen, W. V., Facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto (1978)
13. Ekman, P., Expression and the nature of emotion, In K. Scherer and P. Ekman (eds), *Approaches to Emotion*, Hillsdale, NJ: Erlbaum, (1984)
14. Emotion Recognition in the Wild (EmotiW), URL: <https://sites.google.com/site/emotiw2016/>
15. Euler, H., Mandl, H., *Emotionspsychologie. Ein Handbuch in Schlüsselbegriffen*. Weinheim: Beltz, (1983)
16. Fan, Y., et. al, Video-based emotion recognition using CNN-RNN and C3D hybrid networks, *ICMI 2016 Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 445-450 (2016)
17. Feraru, Silvia Monica; Schuller, Dagmar; Schuller, Björn, Cross-language acoustic emotion recognition: An overview and some tendencies, In: *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 125–131, (2015)
18. Haag, A., Goronzy, S., Schaich, P., & Williams, J., Emotion recognition using biosensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems* (pp. 36-48), Springer Berlin Heidelberg, (2004)
19. Holodynski, M.: Milestones and Mechanisms of Emotional Development, In: *Emotions as Bio-cultural Processes*, pp. 139-163, Springer US, New York (2009)
20. Hornschuh, J., *Weiterentwicklung eines Fahrradergometers als intuitive Steuerung für virtuelle Welten* (2015) URL: <http://users.informatik.haw-hamburg.de/~ubicomp/arbeiten/bachelor/hornschuh.pdf>
21. Juslin, P. N., & Sloboda, J. A., *Music and emotion: Theory and research*, Oxford University Press, pp. 487, (2001)
22. Kamenz, A., *Emotionserkennung mittels Bio-Sensoren*, URL: <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master2016-proj/kamenz.pdf>
23. Kim, J., & André, E., Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12), 2067-2083, (2008)
24. Kleinginna, P. R., and Kleinginna, A. M., A categorized list of emotion definitions, with suggestions for a consensual definition. In: *Motivation and Emotion*, Vol. 5, pp. 345-379, (1981)
25. Kletz, F., Kleimann, J., *Verwendung einer Thermografiekamera im Human-computer interaction Kontext*, URL: <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master2017-proj/kletz-kleimann.pdf>
26. Koelstra, S., et. al, Deap: A database for emotion analysis; using physiological signals, *IEEE Transactions on Affective Computing*, 3(1), 18-31, (2012)
27. LeCun, Y., Bengio, Y., & Hinton, G., Deep learning. *Nature*, 521(7553), pp. 436-444, (2015)
28. Levenson, R. W., Blood, sweat, and fears: The autonomic architecture of emotion, *Emotions Inside Out*, *Annals of the New York Academy of Sciences*, New York, Vol 1000, pp 348–366. (2003)
29. Mehrabian, A., *Communication without words*. *PsychologyToday*, Vol. 2, pp. 53–56, (1968)
30. Müller, L., *EmotionBike: A Study of Provoking Emotions in Cycling Exergames*, In: *Entertainment Computing - ICEC 2015*, Volume 9353 of the series *Lecture Notes in Computer Science* pp. 155-168 (2015)
31. Müller, L., et al.: *Physiological Data Analysis for an Emotional Provoking Exergame*. In: *Proceedings of Computational Intelligence, 2016 IEEE Symposium Series on* (2016)

32. Picard, R., *Affective computing*, The MIT Press, (1997)
33. Plutchik, *The Nature of Emotions*, In: *American Scientist* Vol. 89 (2001)
34. Russel, J., *A Circumplex Model of Affect*, In: *Journal of Personality and Social Psychology* 1980, Vol. 39, No. 6, pp. 1161-1178 (1980)
35. Scherer, K. R., *Neuroscience projections to current debates in emotion psychology*. In: *Cognition and Emotion*, 7, 1-41. (1993)
36. Shen, L. , Wang M., and Shen, R., *Affective e-Learning: Using Emotional Data to Improve Learning in Pervasive Learning Environment*, *Educational Technology & Society*, pp. 176–189, (2009)
37. Schwarz, N., *Cognitive aspects of survey methodology*, John Wiley & Sons, Ltd. (2007)
38. Simons, R. F., Detenber, B. H., Roedema, T. M., & Reiss, J. E., *Emotion processing in three systems: The medium and the message*. *Psychophysiology*, 36(05), pp. 619-627, (1999)
39. Matthiessen, E., *Entwicklung einer Gangschaltungs- und Bremsensteuerung für ein Fahrradergometer* (2015) URL: <http://users.informatik.haw-hamburg.de/~ubicomp/arbeiten/bachelor/matthiessen.pdf>
40. Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K., *Bodily maps of emotions*, *Proceedings of the National Academy of Sciences*, 111(2), 646-651, (2014)
41. Ringeval, Fabien; Schuller, Björn; Valstar, Michel; Jaiswal, Shashank; Marchi, Erik; Lalanne, Denis; Cowie, Roddy; Pantic, Maja, *AV+ EC 2015 – The First Affect Recognition Challenge Bridging Across Audio, Video, and Physiological Data*, In: *AVEC '15 Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, (2015)
42. Riva, G., et. al, *Affective interactions using virtual reality: the link between presence and emotions*, *CyberPsychology & Behavior*, 10(1), pp. 45-56, (2007)
43. Szwoch, Wioleta, *Emotion Recognition Using Physiological Signals*, In: *MIDI'15 Proceedings of the Multimedia, Interaction, Design and Innovation*, Article No. 15, pp. 1-8, (2015)
44. Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R., *Automatic identification of artifacts in electrodermal activity data*. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (pp. 1934-1937). IEEE, (2015)
45. Wagner, J., Kim, J., & André, E., *From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification*. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on* (pp. 940-943). IEEE, (2005)
46. Wen, Guihua; Li, Huihui; Li, Danyang, *An ensemble convolutional echo state networks for facial expression recognition*, In: *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 873–878, (2015)
47. Wundt, W., *Das Gemeingefühl und andere Totalgefühle*, In: *Vorlesungen über die Menschen- und Tierseele*. Verlag von Leopold Voss, p. 235, Leipzig (1906)
48. Yao, A., et. al, *HoloNet: towards robust emotion recognition in the wild*, *ICMI 2016 Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 472-478 (2016)