

ER-RIA Leitlinien

Leitlinien für die Emotionserkennung in roboter- unterstützten Interventionen für Kinder mit Autismus-Spektrum-Störung

EU Erasmus Plus EMBOA Projekt

<http://emboa.eu/>

Version: 1.2, Deutsch, August 2022

Dieser Bericht wird kostenlos unter der Creative Commons-Lizenz CC BY
vertrieben.

URHEBERRECHT UND VERVIELFÄLTIGUNG

Dieses Dokument geht aus dem internationalen Projekt EMBOA hervor, das von der Europäischen Union im Rahmen der Erasmus Plus-Förderschiene finanziert wird. Dieses Dokument wird kostenlos unter der Creative Commons-Lizenz CC BY vertrieben. Das Dokument ist erhältlich auf Englisch, Polnisch, Mazedonisch, Deutsch und Türkisch. Das Dokument darf frei distribuiert werden.

ERSTELLUNG DER LEITLINIEN

Duygun Erol Barkana
Katrín D. Bartl-Pokorny
Hatice Kose
Agnieszka Landowska
Michał R. Wróbel
Ben Robins
Tatjana Zorcec

DEUTSCHE VERSION

Katrín D. Bartl-Pokorny
Manuel Milling
Björn W. Schuller

FÖRDERUNG

Diese Publikation wurde zum Teil finanziert durch das von der Europäischen Kommission im Rahmen der Erasmus Plus-Förderschiene geförderte Projekt EMBOA, Affektive Schleife in sozial unterstützenden Robotern als Interventionswerkzeug für Kinder mit Autismus-Spektrum-Störung, Projektnummer 2019-1-PL01- KA203-065096. Die Förderorganisation nahm keinen Einfluss auf die Inhalte dieser Leitlinien.

HAFTUNGS AUSSCHLUSS

Diese Publikation spiegelt ausschließlich die Ansichten der Autor*innen wider und die Europäische Kommission ist nicht verantwortlich für jegliche Verwendung der in diesem Dokument enthaltenen Informationen.

VORGESCHLAGENE ZITIERUNG DIESER PUBLIKATION

<wird nach der Publikation ergänzt>

FÜR WEITERE INFORMATIONEN ÜBER DIE ER-RIA LEITLINIEN

Webseite des EMBOA-Projekts: <http://emboa.eu/>

Korrespondierende Autorin: Agnieszka Landowska, nailie@pg.edu.pl

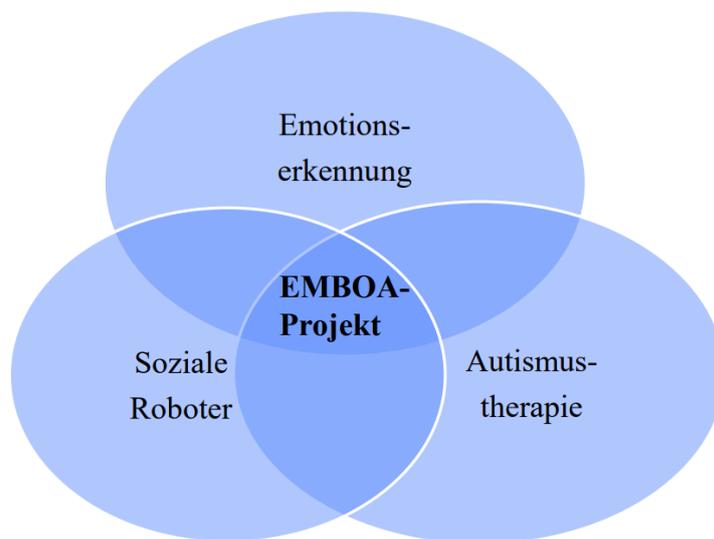
Inhaltsverzeichnis

Zusammenfassung	5
1. Hintergrund und Motivation	6
2. Zweck und Zielgruppe der Leitlinien	9
3. Development method	10
4. Guidelines	15
4.1. The choice of observational channels (CH)	18
4.2. Reliable acquisition of emotional symptoms (SYM)	21
4.3. Technologies and devices (TECH)	25
4.4. Activities in social robot interaction (INT)	28
4.5. Symptoms data processing (PROC)	31
4.6. Emotion recognition in children with autism (EMO)	36
4.7. Design of research studies (RES)	39
4.8. Reporting studies on children with ASD (REP)	41
5. Guidelines evaluation	39
5.1. Questionnaire	39
5.2. Focus group	40
5.3. AGREE expert evaluation	46
5.4. Changes - guidelines 1.0 and 1.2	46
6. Applicability	47
7. Future works	43
Literature	44

Zusammenfassung

Dieses Dokument enthält in der Praxis evaluierte Leitlinien für den Einsatz von Technologien zur automatischen Emotionserkennung im Rahmen von roboter-unterstützter Intervention für Kinder mit Autismus-Spektrum-Störung. Die Leitlinien wurden im Rahmen des EU-Erasmus-Plus-Projektes EMBOA erstellt. Sie konzentrieren sich auf die Kombination von automatisierten Emotionserkennungsansätzen mit zu Therapiezwecken eingesetzten sozial unterstützenden Robotern. Das Hauptziel ist, eine affektive Feedback-Schleife in roboter-unterstützte Autismus-Therapie zu implementieren.

Visuelle Zusammenfassung



Schlagwörter

Sozial unterstützende Roboter; Autismus; Autismus-Spektrum-Störung (ASS); Emotionserkennung; Affekterkennung

1. Hintergrund und Motivation

1.1. Über dieses Dokument

Dieses Dokument enthält eine Reihe von Leitlinien und Empfehlungen für den Einsatz von Technologien der automatisierten Emotionserkennung basierend auf der Beobachtung der Interaktion zwischen einem Roboter und einem Kind mit Autismus-Spektrum-Störung (ASS). Das Dokument wurde im Rahmen des von der Europäischen Union geförderten Erasmus-Plus-Projektes EMBOA erstellt.

Das Dokument ist der wichtigste Output des Projekts und zielt darauf ab, die Erkenntnisse zu beschreiben und zu erklären, die im Rahmen von Literaturrecherchen und Beobachtungsstudien gewonnen wurden. Das Dokument ist erhältlich auf Englisch, Polnisch, Mazedonisch, Deutsch und Türkisch. Dieses Dokument wird kostenlos unter der Creative Commons-Lizenz CC BY vertrieben.

Das Dokument ist in den verschiedenen Sprachen auf der Webseite des EMBOA-Projektes verfügbar: <http://emboa.eu/>.

1.2. Motivation

Autismus-Spektrum-Störung (ASS) ist eine Entwicklungsstörung, die für Betroffene lebenslange Auswirkungen auf die Wahrnehmung ihrer Umwelt und ihre Interaktion mit anderen Personen hat. Per Definition umfasst ASS Defizite in der sozialen Kommunikation und Interaktion wie auch eingeschränkte/repetitive Verhaltensweisen/Interessen/Aktivitäten und betrifft ca. 1% der Bevölkerung (7,5 Millionen Bürger*innen Europas). Diese Entwicklungsstörung kann zu Schwierigkeiten im selbstständigen Leben führen, wie auch mit eingeschränkten Ausbildungs- und Beschäftigungsmöglichkeiten einhergehen. In etwa die Hälfte der Menschen mit ASS haben eine intellektuelle Beeinträchtigung und erwerben die Lautsprache ggf. nicht bzw. nur sehr eingeschränkt. Die ASS kann aktuell nicht „geheilt“ werden. Es gibt allerdings eine Reihe von Interventionsmöglichkeiten, um die Entwicklung von Kindern mit ASS positiv zu beeinflussen. Interventionen sollten sich in erster Linie darauf konzentrieren, die sozialen Fähigkeiten der Kinder zu fördern, da sich herausgestellt hat, dass soziale Kompetenz in der Kindheit in Zusammenhang mit der Erreichung langfristiger Ziele von Menschen mit ASS steht. Autismus-Spektrum-Störung ist eine sehr heterogene Störung, die eigens für jede*n Betroffene*n zugeschnittene Therapien erfordert. Ein Ansatz, der für ein spezielles Kind gut geeignet ist, funktioniert womöglich für ein anderes Kind überhaupt nicht. Ohne zielgerichtete Intervention kann ASS zu großen familiären Problemen und psychischen Erkrankungen führen und Familienmitglieder werden oftmals lebenslange Pflegende.

Kinder mit ASS können von vielseitigen Defiziten betroffen sein, wie unter anderem eingeschränkte soziale Fähigkeiten und Schwierigkeiten bei der Imitation, die ihre Kommunikation und Interaktion mit anderen beeinflussen. Schwierigkeiten in der Mensch-Mensch-Interaktion können ihre Beziehungen zu Familienmitgliedern,

Gleichaltrigen und Therapeut*innen beeinflussen. Kinder mit ASS haben vielfach auch eingeschränkte emotionale Fähigkeiten.

Auch in Therapiesituationen treten Kommunikationsschwierigkeiten zwischen Kind und Therapeut*in auf. Viele Kinder mit ASS profitieren deutlich von roboter-unterstützter Intervention, wenn es um den Erwerb von sozialen und emotionalen Fähigkeiten geht. In den letzten Jahren wurde dies auch zunehmend wissenschaftlich erforscht. Roboter werden als soziale Vermittler eingesetzt, um mit den Kindern in Interaktion zu treten. Dies ermöglicht den Kindern eine vorhersagbare und gleichbleibende Umgebung mit vorhersagbaren Regeln, was sie dabei unterstützt, ihre sozialen Fähigkeiten zu verbessern.

Wir wissen nicht genau, warum Kinder mit ASS viel lieber mit Robotern als mit Menschen interagieren. Dies ist jedenfalls auch der Fall, wenn Roboter eingesetzt werden, die Menschen optisch sehr ähneln. Einige Psycholog*innen gehen davon aus, dass dieses Phänomen damit zusammenhängt, dass Menschen mit ASS sich in vorhersehbaren Situationen sehr wohl fühlen. Jedenfalls haben sich soziale Roboter als wertvolle Möglichkeit erwiesen, Kinder mit ASS zur Interaktion zu motivieren. Ist die Interaktion erst einmal im Gange, bietet sich die Gelegenheit einem Kind schrittweise neue soziale und emotionale Fähigkeiten beizubringen und diese Fähigkeiten durch Übung zu festigen.

Die soziale Interaktion mit Robotern könnte noch verbessert werden, in dem man in den Interaktionsprozess eine affektive Schleife implementiert. Um eine solche affektive Schleife herzustellen, müssten die Emotionen der Kinder mit ASS nicht nur beeinflusst, sondern auch vom Roboter erkannt werden können.

Ziel des EMBOA-Projekts ist es, zu erforschen, inwieweit Roboter mit implementierten Emotionserkennungstechnologien als neues Interventionswerkzeug für Kinder mit ASS eingesetzt werden könnten.

1.3. Über das Projekt

Das EMBOA-Projekt (engl. *Affective loop in Socially Assistive Robotics as an intervention tool for children with autism*) setzt sich zum Ziel, Leitlinien für die Anwendung von Emotionserkennungstechnologien in roboter-unterstützter Intervention für Kinder mit ASS zu entwickeln und diese praktisch zu evaluieren. Das Projekt wird über die Erasmus-Plus-Förderschiene der Europäischen Union finanziert.

In dem Projekt kooperieren Hochschulpartner*innen miteinander, die in den Bereichen Therapie von Kindern mit Entwicklungsstörungen, Robotik und automatisierte Emotionserkennung spezialisiert sind.

Das Projektkonsortium stammt von folgenden Universitäten:

- Politechnika Gdańska, Polen,
- University of Hertfordshire, Vereinigtes Königreich,
- İstanbul Teknik Üniversitesi, Türkei,

- Yeditepe Üniversitesi, Türkei,
- Macedonian Association for Applied Psychology, Nordmazedonien,
- Universität Augsburg, Deutschland.

Abbildung 2 zeigt ein Foto vom internationalen EMBOA-Projekt-Meeting zum Thema soziale Roboter.



Abb. 2. Interaktion mit sozialen Robotern, EMBOA-Projekt-Training, Hertfordshire, Vereinigtes Königreich.

Die Innovation des Projekts liegt in der Kombination folgender Aspekte: Soziale Roboter und automatisierte Emotionserkennung, eingesetzt in spezifischen Interventionssettings für Kinder mit ASS. Es wird erwartet, dass die dadurch entwickelte affektive Schleife jene Interventionen positiv beeinflusst, die die Verbesserung der emotionalen Fähigkeiten der Kinder zum Ziel haben.

Das Projekt ist eine hochgradig interdisziplinäre Initiative, die Robotik, Informations- und Kommunikationstechnologie, Kognitionswissenschaften, Entwicklungspsychologie, Pädagogik, Mensch-Maschine Schnittstelle, wie auch andere Disziplinen einschließt, um Technologien so zu optimieren, dass sie die Bedürfnisse von Kindern mit ASS und ihren Familien erfüllen. In den letzten Jahren hat der Einsatz von menschenähnlichen – humanoiden – Robotern in der Therapie für Kinder mit ASS deutlich zugenommen und die ersten Forschungsergebnisse zu ihrer Wirksamkeit sind sehr vielversprechend. Die Unterstützung von Interventionen für Kinder mit ASS mittels Robotertechnologie ist innovativ und wir möchten dieses wertvolle Hilfsmittel mit einer automatischen Emotionserkennung weiter verbessern.

Das Ziel des EMBOA-Projektes war zu untersuchen, inwieweit es möglich ist, roboter-unterstützte Intervention mit Emotionserkennungstechnologien zu kombinieren (Machbarkeitsstudie). Im Speziellen wollten wir erforschen, welche Strategien sich bei der Kombination dieser Technologien bewähren und welche Hindernisse eventuell auftreten können. Die Hauptforschungsfrage unseres Projektes lautete: **Wie können soziale Roboter Emotionen der Kinder effektiv erfassen, darstellen und interpretieren, um bei den Kindern jene emotionalen Zustände zu erzielen, die den Interaktionsprozess positiv beeinflussen?**

Die wichtigsten Aufgaben des Projektes umfassen:

- Systematische Literaturrecherche zum Thema Emotionserkennung bei Mensch-Maschine-Interaktion im Bereich der Autismusintervention, um den aktuellen Stand der wissenschaftlichen Literatur zusammenzufassen;
- Durchführung einer Machbarkeitsstudie mit dem Ziel zu erforschen, inwieweit es möglich ist, roboter-unterstützte Intervention mit Emotionserkennungstechnologien zu kombinieren;
- Identifikation der Strategien, die sich bei der Kombination dieser Technologien bewähren und der Hindernisse, die auftreten können;
- Entwicklung von Leitlinien in Bezug auf die Anwendung von Emotionserkennungstechnologien in roboter-unterstützter Intervention für Kinder mit ASS sowie praktische Evaluation dieser Leitlinien.

2. Zweck und Zielgruppe der Leitlinien

2.1. Zweck der Leitlinien

Automatische Emotionserkennung ist eine relativ neue Disziplin mit einer stetig wachsenden Anzahl an Anwendungsfeldern. Alle Beobachtungskanäle, die für Emotionserkennung genutzt werden können (z.B.: Gesichtsausdrücke, prosodische Merkmale in den sprachlichen Äußerungen, physiologische Parameter) zeichnen sich dadurch aus, dass sie sehr anfällig für Manipulation und Störungen sind. Das Ausmaß dieser Störungen ist zusätzlich abhängig vom jeweiligen Kind und dem Kontext. Es gibt bis dato noch wenige Studien zur automatischen Emotionserkennung bei Kindern mit ASS. Die Kombination von humanoiden und Emotionserkennungstechnologien wurde bislang noch selten versucht bzw. wissenschaftlich erforscht. Es ist ein neuer Ansatz, um Kinder mit ASS dabei zu unterstützen und dazu zu motivieren ihre emotionalen Fähigkeiten zu stärken.

Die Leitlinien sollen die Empfehlungen und Herausforderungen in Hinblick auf die Verwendung von Emotionserkennungstechnologien in roboter-unterstützter Autismusintervention zusammenfassen. Wir hoffen eine affektive Schleife zu generieren, d.h. Methoden zur Erfassung und Beeinflussung der Emotionen der Kinder mit ASS in roboter-unterstützter Intervention zu detektieren. Präzise Emotionserkennung während roboter-unterstützten Interventionen wäre eine große Unterstützung für Kinder und

deren Therapeut*innen. Einige Beispiele hierfür sind: Der Roboter könnte in Echtzeit gezielt auf die kindlichen Emotionszustände reagieren. Dies würde eine natürlichere Kommunikation mit den Kindern erlauben und diese gezielter auf Konversationen mit Gleichaltrigen, ihren Eltern, Therapeut*innen, etc. im wirklichen Leben vorbereiten. Automatische Emotionserkennung könnte das Ausmaß an notwendiger manueller Steuerung des Roboters durch Dritte im Rahmen roboter-unterstützter Interventionseinheiten reduzieren. Dies wiederum würde dem Therapiepersonal erlauben, sich mehr auf das Kind zu konzentrieren, um den Interventionserfolg zu maximieren.

2.2. Zielgruppe

Die Zielgruppe dieser Leitlinien ist sehr vielseitig und umfasst u.a. **Therapeut*innen, Eltern bzw. Erziehungsberechtigte, Forscher*innen** und **Technologie-Anbieter*innen**.

Therapeut*innen und **Eltern bzw. Erziehungsberechtigte** könnten über die Möglichkeit informiert werden, die Therapie von Kindern mit ASS mittels neuer Technologien – soziale Roboter, automatische Emotionserkennung, sowie die Kombination dieser Komponenten – zu verbessern.

Forscher*innen, aus technischen und pädagogischen Fachgebieten, könnten sowohl von den beschriebenen Beobachtungen im Rahmen unserer Studien als auch von den Empfehlungen zu den Themen Studienplanung und Bericht der Ergebnisse profitieren.

Technologie-Anbieter*innen könnten auf interessante Möglichkeiten zur Verbesserung sozialer Roboter mittels automatischer Emotionserkennungstechnologien stoßen. Zusätzlich könnten Anbieter von Emotionserkennungstechnologien von unseren Beobachtungen zu Einschränkungen und Grenzen der Verwendung der Technologien zur Detektion von Emotionen bei Kindern mit ASS profitieren.

3. Entwicklungsablauf

Die Entwicklung der Leitlinien basiert auf mehreren Methoden. Als erstes wurde eine umfangreiche Literaturrecherche durchgeführt, um den aktuellen Wissensstand über roboter-unterstützte Intervention in Zusammenhang mit ASS und über automatische Detektion von Emotionen von Kindern mit ASS zu erfassen. Als zweites wurden zwei Beobachtungsstudien zu roboter-unterstützter Intervention durchgeführt. Schlussendlich wurden die Leitlinien mittels des AGREE II-Instruments evaluiert. Die Leitlinien wurden außerdem in der zweiten Beobachtungsstudie praktisch angewandt.

3.1. Systematische Literaturrecherchen

Zwei systematische Literaturrecherchen unterstützten die Entwicklung der Leitlinien. Die erste Literaturrecherche betraf die Anwendung von automatischer Emotionserkennung bei Kindern mit ASS, während die zweite Literaturrecherche roboter-unterstützte Intervention in der Autismustherapie beleuchtete. Die erste Studie zielte darauf ab, den aktuellen

Wissensstand zur Kombination von Autismustherapie und Technologien der automatischen Emotionserkennung zu erfassen. Konkret ging es darum, wie Emotionen von Kindern mit ASS automatisch erkannt werden können und nicht um die Fähigkeit der Kinder Emotionen von anderen Personen zu erkennen. Über 2000 wissenschaftliche Artikel wurden von sieben Suchmaschinen extrahiert; davon wurden 50 Artikel in die qualitative und 27 Artikel in die quantitative Analyse eingeschlossen. Die Studie erzielte einige Schlussfolgerungen hinsichtlich der Beobachtungskanäle, Modalitäten und der verwendeten Methoden für die Emotionserkennung bei Kindern mit ASS. Qualitative Analysen lieferten wichtige Hinweise hinsichtlich der Zusammensetzung der Studienteilnehmer*innengruppe und der wichtigsten Kombinationen von Modalitäten und Methoden. Die Studie könnte für Forscher*innen sehr interessant sein, die Emotionserkennung anwenden oder an der Verbesserung von Methoden für Affektklassifikation im ASS-Bereich arbeiten. Diese systematische Literaturrecherche deckte auch zahlreiche Herausforderungen auf, die mit der Anwendung von Emotionserkennung in Studien mit Kindern mit ASS einhergehen. Einige hilfreiche methodische Hinweise konnten auch identifiziert werden. Die Studie verwendete das PRISMA-Schema für die Durchführung und die Berichterstattung und wurde schließlich auch publiziert [2].

Die zweite Studie zielte darauf ab, den Wissensstand zur Verwendung von sozialen Robotern in der Intervention für Kinder mit ASS zu erfassen. Die vorhandene Literatur deutet darauf hin, dass Kinder mit ASS von roboter-unterstützter Intervention profitieren. Allerdings sind diese Studien sehr heterogen was Charakteristika der Studienteilnehmer*innen, angewandte Roboter und trainierte Fähigkeiten betrifft. Die Roboter könnte man basierend auf ihrem Aussehen grob in fünf Kategorien einteilen: menschenähnliche bzw. humanoide Roboter, tier- oder kreaturähnliche Roboter, mobile Roboter, ballförmige Roboter, und andere Roboter. Die größte Rolle in der Autismustherapie spielen Roboter, die mit den Kindern interagieren – so genannte soziale Roboter. Einige soziale Roboter sind in Abbildung 1 dargestellt. Die ersten drei sind humanoide Roboter, während es sich beim vierten um einen tierähnlichen Roboter handelt.

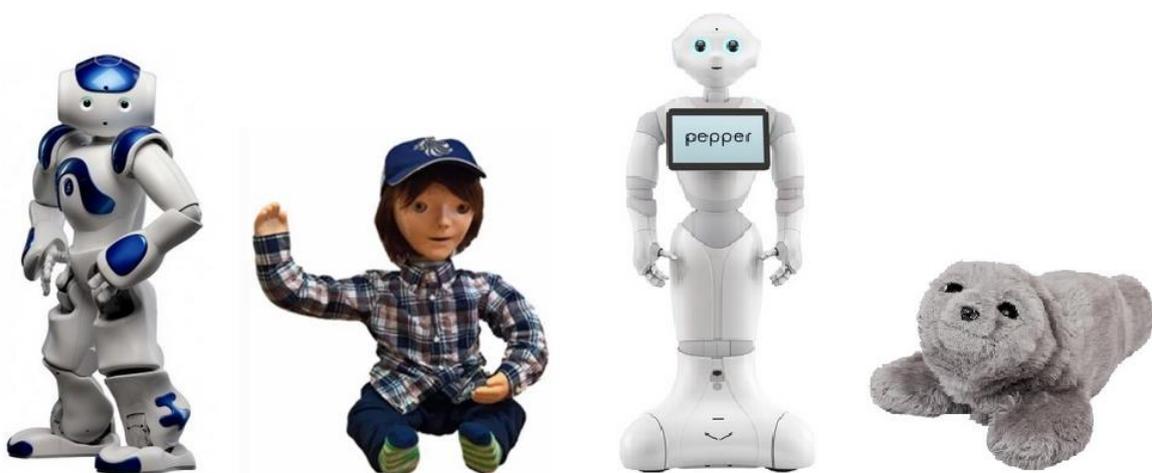


Abb. 1. Beispiele von sozialen Robotern (von links nach rechts): Nao, Kaspar, Pepper, Paro

Für die systematische Literaturrecherche zu roboter-unterstützter Intervention für Kinder mit ASS und den dabei trainierten Fähigkeiten konzentrierten wir uns auf alle relevanten Artikel, die zu dem Zeitpunkt in den Datenbanken Scopus, Web of Science und PubMed gelistet waren. Von insgesamt 609 potenziell relevanten Artikeln erfüllten 60 Artikel – davon 50 Originalarbeiten und 10 Überblicksartikel oder theoretische Arbeiten – die Einschlusskriterien. Nao und ZECA waren die am häufigsten verwendeten Roboter in diesen Artikeln. Die am häufigsten trainierten Fähigkeiten waren das Erkennen von Basisemotionen und das Herstellen von Interaktionen, wobei Freude, Traurigkeit, Angst und Ärger die am häufigsten trainierten Emotionen waren. Die eingeschlossenen Studien berichteten zahlreiche Herausforderungen in Hinblick auf roboter-unterstützte Intervention, die von eingeschränkter Anwendung für spezifische ASS-Subgruppen über Sicherheitsaspekte der Roboter selbst bis hin zur automatischen Erkennung der kindlichen Emotionen durch die Roboter reichen. Der systematische Review beinhaltet auch eine Zusammenfassung und Diskussion hinsichtlich der Anwendung von roboter-unterstützter Intervention für Kinder mit ASS. Auch diese Arbeit wandte das PRISMA-Schema an und wurde publiziert [1].

3.2. Beobachtungsstudien

Die Entwicklung der Leitlinien wurde von zwei Beobachtungsstudien unterstützt, in denen Emotionserkennungstechnologien angewandt wurden. Die Studien wurden im Rahmen des EMBOA-Projekts in Nordmazedonien, Polen, im Vereinigten Königreich und in der Türkei durchgeführt. Häufig verwendete Interaktionsszenarien und Protokolle wurden zur Durchführung der Studien eingesetzt. Die Planung der ersten Beobachtungsrunde erfolgte basierend auf den Erkenntnissen der Literaturrecherchen. Im Anschluss analysierten wir die erhobenen Daten und formulierten darauf basierend Leitlinien. Die zweite Beobachtungsrunde adaptierten wir gemäß den gewonnenen Erkenntnissen.

Dies sind einige generelle Vorgaben, die wir für die Beobachtungssitzungen getroffen haben:

- Alle Vorbereitungen und Einstellungen sollten erfolgen bevor das Kind den Untersuchungsraum betritt;
- sofern möglich, sollten mit jedem Kind zwei oder mehr Sitzungen durchgeführt werden;
- zu Beginn sollte Kind langsam an die neue Situation herangeführt werden, um es an diese Form der Intervention zu gewöhnen;
- es steht eine Liste an häufig eingesetzten Szenarien zur Verfügung, die nach Möglichkeit mit den Kindern durchgeführt werden sollten;
- die Szenarien sollten in die jeweiligen Sprachen übersetzt und bei Bedarf angepasst werden (beispielsweise gibt es in verschiedenen Ländern verschiedene bekannte Kinderlieder);
- die Mitarbeiter*innen wurden dazu angehalten nach Möglichkeit die definierten Szenarien in den Sitzungen mit den Kindern anzuwenden, aber es war möglich, bei der

Auswahl Rücksicht auf das jeweilige Kind zu nehmen, indem nach Bedarf einzelne Szenarien weggelassen oder weitere Szenarien hinzugefügt wurden;

- folgende Daten wurden aufgezeichnet: Video (sofern möglich mit zwei Kameras), Augenbewegungen, Stimme (mit zwei Mikrofonen, wovon eines am Körper des Kindes angebracht werden sollte), physiologische Parameter (Herzfrequenz, Hautleitwert);
- während jeder Sitzung wurden die wichtigsten Beobachtungen mitnotiert; genauere Annotationen erfolgten im Nachhinein basierend auf den Aufnahmen;
- die erhobenen Daten wurden in pseudonymisierter Form gespeichert und innerhalb des EMBOA-Konsortiums geteilt.

Folgende Szenarien wurden in den Sitzungen eingesetzt:

- Scenario 1. Begrüßung und ein Lied („Happy and you know it“ oder ein äquivalentes, länderspezifisch besser geeignetes, Lied) – dieses Szenario beinhaltet grundlegendes Gesprächseinstiegs- und ausstiegsvokalular (hallo, tschüss, wie geht es dir) und ein Lied, das aus mehreren Strophen besteht; dieses Szenario kann eingesetzt werden, um mit den Kindern Begrüßungen und gemeinsames Singen zu üben;
- Scenario 2. Emotionen – beinhaltet das Zeigen der Posen glücklich, traurig, verstecken und überrascht mittels Gesichtsausdrücken und Armbewegungen zusammen mit der entsprechenden verbalen Information. Das Kind wird außerdem dazu aufgefordert, das Verhalten nachzumachen; dieses Szenario kann eingesetzt werden, um mit dem Kind Imitation und Emotionserkennung zu üben;
- Scenario 3. Tiere – das Kind wird dazu aufgefordert Tierlaute zu produzieren bzw. nachzumachen; dieses Szenario kann eingesetzt werden, um mit dem Kind die Produktion von Lautäußerungen, Imitation von Lauten, Sprecher*innenwechsel und das Beantworten von Fragen zu üben;
- Scenario 4. Körperteile – der Roboter Kaspar zeigt anhand seines Körpers verschiedene Körperteile und benennt diese; dieses Szenario kann eingesetzt werden, um mit dem Kind die Imitation von Bewegungen, das Zeigen von Körperteilen (von Kaspar oder vom Kind selbst) und Sprecher*innenwechsel zu üben;
- Scenario 5. Kannst du das nachmachen was Kaspar macht? – Kaspar zeigt verschiedene Posen (z.B.: Hand zur Seite, nach vorne, nach oben strecken) und fordert das Kind dazu auf, diese Bewegungen nachzumachen; dieses Szenario kann eingesetzt werden, um mit dem Kind Imitation von Bewegungen, Sprecher*innenwechsel und Merkfähigkeit zu üben;
- Scenario 6. Nachmachen von Lautäußerungen – Kaspar spricht bestimmte Laute (Vokale), Silben oder einfache Worte vor und fordert das Kind dazu auf, diese Lautäußerungen zu wiederholen; dieses Szenario kann eingesetzt werden, um mit dem Kind Lautäußerungen, einfache Sprachproduktion, Imitation von Lauten, Sprecher*innenwechsel und Merkfähigkeit zu üben.

Zusätzlich beinhaltet jedes Szenario Aufforderungen und bestärkende Äußerungen wie: gut gemacht, bravo, danke, Geräusch einer Trompete + bravo, das tut weh, du bist dran, etc. Diese sollen dazu dienen das Kind dazu zu motivieren mitzumachen und mit Kaspar zu

interagieren. Jede Sitzung startete mit „Hallo“ und endete mit „Tschüss“. Die Reihenfolge der Szenarien konnte an die jeweiligen Bedürfnisse angepasst werden.

Die Partner*innen in vier Ländern führten die Beobachtungsstudien durch und verwendeten dafür zur besseren Vergleichbarkeit der Ergebnisse das gleiche Equipment. Folgendes Equipment wurde eingesetzt:

- Affectiva E4 Armband zur Erhebung von physiologischen Parametern;
- Gazepoint GP3 Eye-Tracking-Gerät zur Erhebung von Blickbewegungen;
- 2 Mikrophone: Zoom H4n Pro und AKG C 417 L Ansteckmikrofon mit Adapter;
- 1 Kamera (erste Beobachtungsstudie) bzw. 2 Kameras (zweite Beobachtungsstudie).

Die jeweiligen Instrumente wurden in der Phase der Projektbeantragung ausgewählt und einige davon erwiesen sich als sehr nützlich für unseren Anwendungsfall. Bitte beachten Sie, dass die Leitlinien keine speziellen Geräte empfehlen, dass wir die von uns verwendeten Geräte aber für zukünftige Referenzierungen (und ggf. Wiederholung) in unserer Studie nennen. Dies waren die von uns definierten Kriterien für die Auswahl der Geräte:

- das Kind sollte durch die Instrumente so wenig wie möglich abgelenkt werden;
- Möglichkeit von Langzeitmessungen;
- möglichst störungsfreie Funktionalität;
- Möglichkeit des Datenexports;
- gutes Preis-Leistungs-Verhältnis.

Im Rahmen der zweiten Beobachtungsstudie haben wir basierend auf den Erkenntnissen aus der ersten Beobachtungsstudie (Leitlinien) folgende Adaptationen getroffen:

- Verwendung von 2 Kameras (eine frontal auf das Kind gerichtet und eine, die die gesamte Szene erfasst);
- Anpassung der Kameraposition und des Aufnahmewinkels und Angabe der Auflösung und des Winkels der Kameralinsen;
- Ggf. Verwendung alternativer Kameras (für ausgewählte Partner*innen);
- Verzicht auf das Eye-Tracking-Gerät, wodurch Kaspar niedriger sitzt – alternativ sollen Blickbewegungsmuster aus dem frontalen RGB-Video-Bild gewonnen werden;
- Verzicht auf das Ansteckmikrofon (da damit keine besseren Ergebnisse erzielt werden konnten als mit dem Mikrofon, das nicht am Körper der Kinder getragen werden muss);
- Isolierte Beispielaufnahmen der kindlichen Stimme, der Raumakustik, der Roboterstimme und der Stimme der Therapeut*innen;
- Verbesserung der Synchronisation der verschiedenen erhobenen Signale.

Insgesamt nahmen ### Kinder mit ASS an unserer Studie teil; sie waren zwischen ### und ### Jahre alt und ### waren männlich und ### waren weiblich. Zusätzlich haben wir folgende Angaben erhoben: Entwicklungsalter, Art und Dauer bisher erhaltener Therapien, Komorbiditäten, Anzahl an Geschwistern, Abschätzung von Sprachverständnis, Sprachproduktion, Wissen über Tiere und Körperteile.

4. Leitlinien

Alle Empfehlungen bzw. Vorschläge, die wir im Rahmen der Literaturrecherche gefunden hatten und alle Erkenntnisse, die wir im Rahmen der Beobachtungsstudien gewonnen hatten, wurden integriert und in Form von Leitlinien zusammengefasst.

Die Leitlinien wurden aus Übersichtsgründen in folgende Kategorien eingeteilt:

1. Leitlinien, die den Einsatz von Beobachtungskanälen betreffen (bezeichnet als CH)
2. Leitlinien, die die akkurate Erhebung von Emotionszuständen betreffen (SYM)
3. Leitlinien zu eingesetzten Technologien und Instrumenten (TECH)
4. Leitlinien, die die Interaktion mit dem sozialen Roboter betreffen; umfasst Aktivitäten, die gut geeignet sind Emotionszustände zu erfassen sowie mögliche Hindernisse (INT)
5. Leitlinien, die die Datenverarbeitung betreffen (PROC)
6. Leitlinien, die sich auf die Emotionserkennung von Kindern mit ASS beziehen (EMO)
7. Leitlinien zum Studiendesign (RES)
8. Leitlinien hinsichtlich der Dokumentation von Studien zu Emotionserkennung, Robotern und Kindern mit ASS (REP)

Die Leitlinien wurden aus Übersichtsgründen mit einem Kategoriekürzel versehen und nummeriert. Zusätzlich zu den oben genannten Kategorien wurden drei allgemeine Leitlinien (GEN) an den Beginn gesetzt.

Leitlinie GEN1:

Folgen Sie den Bedürfnissen des Kindes und den therapeutischen Zielen

Roboter-unterstützte Intervention und die Erhebung von Signalen für die automatische Emotionserkennung erfordert den Einsatz von Technologien. Dies macht das Therapiesetting komplexer, da die Geräte ein gewisses Maß an Aufmerksamkeit von einer bedienenden Person benötigen. Diese Leitlinie besagt, dass der Fokus auf das Kind und die therapeutischen Ziele gelegt werden soll. **Weder das Kind noch die Therapie sollten angepasst werden müssen, um die Bedürfnisse der Emotionserkennungstechnologien zu erfüllen. Stattdessen sollte das Emotionserkennungsssetup an die jeweiligen therapeutischen Ziele angepasst werden.**

Wenn man die vielen Studien zur Entwicklung von Emotionserkennungsanwendungen in der Autismustherapie betrachtet, ergeben sich einige besonders wichtige Fragen, die am Beginn jeder Studie gestellt werden sollten:

- Zu welchem Zweck werden Emotionen erkannt? Sollen die Emotionen erkannt werden, um die Intervention zu unterstützen oder um die Funktionalität der Technologien (Roboter, App) an die Emotionen anzupassen?
- Auf welche Weise könnte Emotionserkennung dabei helfen, die Fähigkeiten der Kinder zu verbessern oder Kinder mit ASS zur Interaktion zu motivieren?

- Welche Fähigkeiten könnten mit Hilfe von automatischer Emotionserkennung geübt werden? [1]

Leitlinie GEN₂:

Legen Sie zu Beginn fest, was konkret Sie über die Emotionszustände des Kindes erfahren möchten

Zu Beginn sollten Sie sich überlegen, was Sie erfassen möchten und was Sie über die kindlichen Emotionen erfahren möchten. Definieren Sie, welche Emotionszustände des Kindes Sie detektieren möchten und wählen Sie dann eine Technologie aus bzw. adaptieren Sie diese entsprechend. Welche Emotionen müssen Sie für welchen Zweck automatisch erkennen können? Es könnte einfacher sein, nur ausgewählte emotionale Zustände zu detektieren.

In der psychologischen Forschung existiert keine einheitliche Definition von menschlicher Emotion [28]. Dennoch besteht weitgehend Konsens über ein diskretes Emotionskonzept. Es gibt verschiedene Modelle, die Emotionen abbilden [29]. Bei der Analyse von wissenschaftlichen Artikeln, die sich mit Emotionserkennung befassen, konzentrierten wir uns auch darauf, welche Emotionen unterschieden und analysiert wurden. Die meisten Artikel verwendeten folgende Emotionsmodelle: Ekman's Basisemotionen (Freude, Ärger, Angst, Ekel, Traurigkeit, Überraschung) [30] und/oder das zweidimensionale Modell von Valenz und Erregung. Einige Artikel, die das Modell von Ekman verwendeten, konzentrieren sich auf ausgewählte Emotionen des Emotionsrepertoires. Bei der Interpretation der Studien sollte beachtet werden, dass die Artikel keine einheitlichen Ausdrücke für die Emotionen verwendeten. Teils konzentrierten sich Studien auch auf weitere Emotionen, die nicht in den beiden Emotionsmodellen enthalten sind. Einige Studien analysierten auch über die klassischen Emotionen hinausgehende Aspekte wie Gemütslagen, kognitive Zustände (z.B.: Konzentration) oder sogar Hunger. Genauere Informationen darüber können Sie unserem systematischen Review [2] entnehmen.

Vielfach kann es eine gute Entscheidung sein mit der Detektion der sechs Basisemotionen (glücklich, ängstlich, verärgert, überrascht, angeekelt, traurig) zu beginnen, aber sie könnten ggf. für gewisse Studien oder Interventionen nicht ausreichend sein. Beispielsweise berichteten einige Studien, dass die Detektion von Bereitschaft des Kindes zur Interaktion mit dem Roboter sowie von frühen Anzeichen von Stress besonders wichtig war [31, 32].

Leitlinie GEN₃:

Die Rechte des Kindes müssen gewahrt werden

Unabhängig vom Zweck der Intervention und/oder Studie müssen die Rechte der Kinder stets gewahrt werden. So hat das Kind zum Beispiel das Recht auf die bestmögliche Therapie.

Ein anderer wesentlicher Aspekt ist die Privatsphäre des Kindes zu schützen. Gemäß der Datenschutz-Grundverordnung (<https://gdpr-info.eu>) versteht man unter persönlichen Daten alle Informationen, durch die eine Person identifiziert werden kann. Persönliche Informationen eines an einer Studie/Intervention teilnehmenden Kindes wie beispielsweise Name, Identifikationsnummern oder Adresse müssen vertraulich behandelt werden. Die Eltern bzw. Erziehungsberechtigte müssen in Form einer Einverständniserklärung einer möglichen Studienteilnahme zustimmen. Auch die Publikation der Ergebnisse, Lautäußerungen oder Fotos der Kinder darf nur erfolgen, wenn die Eltern bzw. Erziehungsberechtigten dem in der Einverständniserklärung zustimmen. Weitere ethische Risiken müssen ebenfalls berücksichtigt werden. Das Kind darf die Interaktion bzw. Studienteilnahme jederzeit beenden, sofern es dies wünscht. Auch mögliche technische Risiken, die die Sicherheit des Kindes gefährden könnten, müssen beachtet werden (siehe TECH1). Bei Bedarf muss ein Ethikvotum beantragt werden.

4.1. Wahl der Beobachtungskanäle (CH)

Emotionen rufen zahlreiche Reaktionen des Nervensystems hervor: Beispielsweise Gesichtsausdrücke, Körpersprache, Lautäußerungen, Herzfrequenz, Hautleitwert und auch periphere Körpertemperatur können sich durch Emotionen verändern. All diese Parameter können gemessen werden und die Werte können Hinweise auf Emotionszustände liefern. Mithilfe eines **Kanals** (z.B. Video) kann ein spezielles Signal (z.B. Gesichtsausdruck, Stimme) aufgezeichnet werden. Unter **Modalität** versteht man eine Art von Information, die beobachtbar ist und die zur Emotionserkennung herangezogen werden kann. Bisherige Studien konzentrierten sich auf folgende Modalitäten:

- Bewegung: Gesichtsausdrücke, Körperbewegungen, Blickbewegungen, Kopfbewegung, Gesten (Handbewegungen), und andere Arten von Bewegung;
- Akustische Information: Lautäußerungen, Prosodie der Sprache;
- Herzaktivität: Herzfrequenz, HRV (Herzratenvariabilität);
- Muskelaktivitäten, die nicht in Verbindung mit Bewegungen stehen: Muskelspannung;
- Schweißsekretion: Hautleitwert;
- Atmung: Intensität und Frequenz;
- Thermoregulation: periphere Körpertemperatur;
- Gehirnaktivität: neuronale Aktivität. [2]

Leitlinie CH1:

Beachten Sie die jeweils geplante Aktivität, die medizinischen Besonderheiten des Kindes und die Umgebung bei der Wahl des Beobachtungskanals

Ob spezifische Beobachtungskanäle für die Emotionserkennung von Nutzen sind, hängt maßgeblich von den zu beobachtbaren Aktivitäten ab.

Bei der Wahl der Kanäle ist es demnach wichtig die Art der geplanten Interaktion zwischen dem Kind und dem Roboter zu berücksichtigen. So ist es beispielsweise möglich, dass eine RGB-Kamera während einer Interaktion, die vom Kind Bewegung erfordert – wie beispielsweise Turnübungen – ggf. nicht ausreichend Daten für Emotionserkennung liefert, da das Kind sich teilweise aus dem Kamerawinkel herausbewegt bzw. das Gesicht des Kindes nicht durchgehend erfassen kann.

Darüber hinaus müssen auch individuelle Gegebenheiten und Bedürfnisse des Kindes beachtet werden. Manche Kinder mit ASS wehren sich beispielsweise gegen Körperkontakt, was das Anbringen von Armbändern und ähnlichen Geräten zur Messung von physiologischen Parametern erschwert. Beachten Sie auch die individuellen Fähigkeiten und Sprachkenntnisse des Kindes.

Bei der Wahl der Beobachtungskanäle sollten auch die Umgebungsbedingungen (z.B. Ort, Dauer) der Studie berücksichtigt werden.

Leitlinie CH₂:

Es ist möglich, dass Beobachtungskanäle nicht während der gesamten Untersuchung brauchbare Daten liefern.

Unabhängig davon welcher Kanal verwendet wird, kann es zu einem zeitweisen Fehlen von Daten kommen. Woran das Fehlen von Daten liegt, kann je nach Kanälen und Geräten verschieden sein. Wenn beispielsweise Bilder des Gesichts mit einer RGB-Kamera aufgezeichnet werden sollen, kann es vorkommen, dass das Kind sich aus dem Aufnahmebereich der Kamera herausbewegt, nach unten oder auf die Seite blickt oder Teile des Gesichts verdeckt werden. Wenn physiologische Daten mit einem Armband erhoben werden, können beispielsweise schnelle Handbewegungen oder Fehler in der Datenaufzeichnung auslösen. Solche Situationen können nicht verhindert werden und treten besonders in Studien mit Kindern häufig auf.

Wichtig ist, dass Sie sich bei der Datenauswertung dieser Risiken bewusst sind. Die Qualität der mit den einzelnen Kanälen erhobenen Daten sollte regelmäßig überprüft werden und fehlerhafte Daten, welche nicht zuverlässig Emotionszustände abbilden können, sollten manuell entfernt werden. Um eine größere Menge an zuverlässigen Daten zu gewinnen, kann es von Vorteil sein, multimodale Untersuchungen (siehe CH₃) durchzuführen oder mehr Daten einer einzelnen Modalität zu erheben, indem beispielsweise mehrere Kameras zum Einsatz kommen (siehe SYM₂).

Leitlinie CH₃:

Multimodale Untersuchungen können dabei helfen, eine größere Menge an Daten zu gewinnen und die Reliabilität der Daten zu erhöhen

Da einzelne Beobachtungskanäle zeitweise keine Daten liefern können (Leitlinie CH₁), könnte der Einsatz mehrerer Beobachtungskanäle oder Geräte zielführend sein.

Man kann zwischen Mehrkanal-Beobachtungen und multimodalen Beobachtungen unterscheiden. Mehrkanal-Beobachtungen bedeuten, dass mehrere Geräte eingesetzt werden, um Daten zu erheben, z.B.: es wird sowohl eine Kamera als auch ein Mikrophon verwendet oder aber mehrere Kameras. Multimodale Beobachtungen bedeutet, dass verschiedene Arten an Daten erhoben werden, die auf Emotionszustände rückschließen lassen, z.B.: Gesichtsausdrücke, prosodische Elemente der gesprochenen Sprache. Beachten Sie hierbei, dass multimodale Beobachtungen (z.B.: Gesichtsausdrücke, Prosodie) entweder mittels eines einzelnen Kanals (beide Modalitäten können auf den Daten basierend ausgewertet werden, die eine einzelne Kamera erhoben hat) oder mittels mehrerer Kanäle (sowohl Kamera als auch Mikrophon kommen zum Einsatz) erfolgen können.

Multimodale (oder Mehrkanal-) Analyse von Emotionszuständen erhöht die Verfügbarkeit der Daten und die Reliabilität der Ergebnisse. Je mehr Geräte man verwendet umso höher ist die Wahrscheinlichkeit, dass zumindest eines der Geräte zu einem bestimmten Zeitpunkt wertvolle Informationen über den Emotionszustand des Kindes aufzeichnet. Des Weiteren erzielen multimodale Emotionserkennungssysteme oftmals eine höhere Klassifikationsgenauigkeit als Systeme, die nur auf einer Modalität beruhen [6]. Daher ist es – sofern möglich – ratsam, mehrere Beobachtungskanäle zu verwenden, um unterschiedliche Modalitäten zu erfassen.

Kommen mehrere Modalitäten oder Kanäle zum Einsatz, kommt es zu Herausforderungen was die Zusammenführung der Daten der verschiedenen Kanäle betrifft (siehe dazu Leitlinien PROC₁ und PROC₂).

Leitlinie CH₄:

Gehen Sie sparsam mit der Anzahl an Geräten um

Kinder mit ASS mögen meist sich wiederholende, möglichst gleiche Situationen und Umgebungen. Alle Eingriffe bzw. Veränderungen – wie beispielsweise Geräte, die an den Körper des Kindes angebracht werden, Geräte im Raum und zusätzlich anwesende Personen – können dazu führen, dass das Kind nicht an der Therapie teilnehmen möchte oder könnten es dazu veranlassen, sich anders als sonst zu verhalten. Zur Veranschaulichung soll eine Studie zur Erkennung von Lächeln genannt werden, für diese im Gesicht des Kindes kabelloses EMG angebracht werden musste. Es wurde berichtet, dass 70% der Kinder damit einverstanden waren dieses EMG-Gerät zu tragen [33], was bedeutet, dass es bei manchen Kindern nicht möglich war, die entsprechenden Daten zu erheben. Die Studie analysierte den Effekt nicht, den das Tragen des Gerätes auf das kindliche Verhalten hatte.

Je mehr Geräte man einsetzt, umso wahrscheinlicher wird es, dass ein Kind aufgrund der Neuheit und Komplexität der Umgebung (Kameras, Computer, Kabel, zusätzliche Personen im Raum) nicht mehr an der Studie oder Therapie teilnehmen möchte. Es ist uns bewusst, dass diese Leitlinie der vorherigen (CH₂) widerspricht. Es ist wichtig ein

gutes Mittelmaß zwischen Reliabilität und Verfügbarkeit der Daten und veränderten Umgebungsbedingungen zu finden.

Eine Möglichkeit wäre es, mit einem einzigen Gerät mehrere Modalitäten zu erfassen. So könnte beispielsweise ein RGB Video Gesichtsausdrücke, Körperbewegungen und Gesten erfassen [2]. Man könnte auch darüber nachdenken, Blickbewegungen vom Video abzuleiten anstatt teure und schwer kalibrierbare Eye-Tracking-Geräte einzusetzen. Auch akustische Informationen könnten aus den Aufnahmen einer hochqualitativen Videokamera gewonnen werden anstatt möglicherweise ablenkende Mikrophone auf dem Tisch vor dem Kind oder auf der Kleidung des Kindes anzubringen.

Verwendet man beispielsweise eine RGB-Kamera und ein Armband, das Biosignale aufzeichnen kann, geht man sehr sparsam mit Geräten um, kann aber dennoch einige Modalitäten erfassen. So lassen sich mit der Kamera Bewegungen (z.B.: Gesichtsausdrücke, Blickbewegungen, Kopfbewegungen) und gesprochene Sprache (z.B.: prosodische Merkmale der Sprache, Lautäußerungen) erfassen. Mit dem Armband können verschiedene Biosignale (z.B.: Herzfrequenz, Hautleitwert, Temperatur) aufgezeichnet werden [2].

4.2. Zuverlässige Erfassung von Parametern, die auf den Emotionszustand schließen lassen (SYM)

Wir stellten fest, dass, auch wenn wir die gesamte Therapieeinheit mit Geräten aufzeichnen, die Daten möglicherweise trotzdem nicht ausreichend Informationen enthalten, um auf Emotionszustände rückschließen zu können. Das trifft beispielsweise auf die Erfassung von Augenbewegungen mittels Eye-Tracking zu, das eine aufwändige Kalibrierung erfordert, aber auch auf andere Modalitäten. Die Erfassung von Parametern, die auf den Emotionszustand schließen lassen, wird von vielen Störfaktoren beeinflusst, die die Zuverlässigkeit der Ergebnisse reduzieren können. Daher ist es sinnvoll, sich näher mit den Aufnahmebedingungen der verschiedenen Parameter auseinanderzusetzen, die auch von der spezifischen Modalität abhängig sind.

Leitlinien zur Erfassung der Gesichtsausdrücke des Kindes

Leitlinie SYM₁:

Passen Sie den Abstand zwischen dem Gesicht des Kindes und der Kamera an die Auflösung der Kamera an.

Um Emotionen aus kindlichen Gesichtsausdrücken ableiten zu können ist es notwendig, dass das gesamte Gesicht des Kindes im Videobild sichtbar ist. Optimal wäre es, wenn das Gesicht des Kindes zumindest 10% des gesamten Bildes einnimmt. Ein Beispiel: Wird eine gewöhnliche kleine Internetkamera verwendet, sollte das Gesicht des Kindes ca. 1-1,5 Meter entfernt von der Kameralinse sein. Die Entfernung sollte an die Auflösung und den Winkel der Kamera angepasst werden. Auch der Einsatz von Teleobjektiven kann

hilfreich sein. Werden diese Hinweise nicht berücksichtigt, besteht die Gefahr, dass die Emotionserkennung basierend auf Gesichtsausdrücken weniger zuverlässig funktioniert.

Wie bereits erwähnt ist es bei manchen Emotionserkennungssystemen erforderlich, dass das Gesicht einen gewissen Mindestanteil des Szenenbildes ausfüllt. Allerdings ist es möglich in einem Nachbearbeitungsschritt das Gesicht entsprechend zuzuschneiden und zu vergrößern. Dies ist jedoch zeitaufwändig und könnte zu einer verringerten Bildqualität führen, wenn Videodaten mit niedriger Auflösung zugrunde liegen. Daher ist es empfehlenswert, die Kamera nahe am Gesicht des Kindes zu platzieren und die Kameraauflösungseinstellungen anzupassen, um die bestmögliche Aufnahmequalität zu gewährleisten.

Es kann vorkommen, dass der Abstand zwischen Kamera und Kind zu gering ist, wenn der Kamerawinkel klein ist. Falls das Kind sich dann zu weit nach vorne oder seitlich bewegt (was sich nicht verhindern lässt), kann es sein, dass nur ein Teil des Gesichts erfasst wird, wodurch sich wieder Probleme bei der automatischen Emotionserkennung ergeben könnten. Die Verwendung von Weitwinkelkameras kann sich daher als hilfreich erweisen.

Überprüfen Sie bitte vor dem Start der Aufnahme, ob das gesamte Gesicht erfasst wird und wie die Größe des Gesichts in Relation zum ganzen Bild ist, um bestmögliche Erkennungsergebnisse zu erzielen.

Leitlinie SYM₂:

Verwenden Sie mehr als eine Kamera und richten Sie zumindest eine Kamera frontal auf das Kind

Wenngleich wir in Leitlinie CH₃ empfohlen haben mit einem einzelnen Gerät mehrere Modalitäten zu erfassen, möchten wir vor den Risiken warnen, die mit der Verwendung von nur einer Kamera einhergehen. Beispielsweise ist eine einzelne Kamera nicht ausreichend, um kontinuierlich über die gesamte Therapieeinheit hinweg Informationen über Gesichtsausdrücke zu erfassen, wenn das Kind sich im Raum bewegt oder sich während es sitzt auf die Seite dreht. Ziehen Sie also in Betracht mehrere Kameras im Raum zu verteilen, um die Zeitspanne zu verlängern, in der das Gesicht des Kindes aus dem richtigen Winkel erfasst werden kann. Eine einzelne Kamera, die zu hoch oder seitlich platziert ist, kann Ihnen einen Szenenüberblick liefern, jedoch nicht in ausreichender Qualität das Gesicht erfassen. Ein schräger oder gekippter Blick auf das Gesicht kann es schwierig machen die Gesichtsausdrücke zu analysieren. Bei der Verwendung von mehr als einer Kamera ist es wichtig auf die Synchronisation der Geräte zu achten (siehe PROC₁ und PROC₂).

Die Emotionserkennung basierend auf Gesichtsausdrücken wird maßgeblich davon beeinflusst, wie die Kamera auf das Gesicht gerichtet wird [7]. Positioniert man die Kamera beispielsweise über dem Gesicht, ist die Wahrscheinlichkeit größer, dass die Emotion Ärger erkannt wird. Positioniert man die Kamera zu niedrig, ist es

wahrscheinlicher, dass die Emotion Überraschung erkannt wird. Ähnliche Effekte zeigen sich was die vertikale Ausrichtung betrifft. Aus diesem Grund ist es insbesondere bei der Verwendung von nur einer Kamera sehr wichtig, diese so frontal wie möglich auf das Gesicht des Kindes zu richten.

Werden mehrere Kameras verwendet, können zusätzliche von der Seite gerichtete Kameras Verhaltensmuster wie Körperhaltung und Gesten erfassen.

Um bessere automatische Emotionserkennungsergebnisse zu erzielen ist es wichtig, dass möglichst wenige andere Personen zugleich mit dem Kind von der Kamera erfasst werden.

Leitlinie SYM₃:

Masken, Brillen und andere Accessoires, die das Gesicht verdecken, machen es schwieriger oder manchmal auch unmöglich zuverlässige Gesichtsausdrücke zu erfassen auf denen basierend automatische Emotionserkennung erfolgen kann.

Wird das Gesicht teilweise verdeckt, kann es zu Problemen bei der Emotionserkennung basierend auf Gesichtsausdrücken kommen. Zu den häufigsten dieser Verdeckungen zählen Brillen, insbesondere solche mit dicken Rahmen. Auch Haarschnitte mit langem Pony, der die Augen verdeckt, beeinflusst die Erkennung des Gesichts und damit die Emotionsausdrücke. Aufgrund der COVID-19-Pandemie sind in letzter Zeit auch Masken häufig notwendige Accessoires, die das Gesicht teilweise verdecken.

Sorgen Sie wann immer möglich dafür, dass das Gesicht möglichst frei ist, z.B. indem in einer sicheren Umgebung auf Masken verzichtet wird oder das Haar aus dem Gesicht frisiert wird. Natürlich wird das nicht immer möglich sein, z.B. bei Brillenträger*innen oder in einigen Fällen auch bei Masken. Daher ist es wichtig, diese Situationen zu dokumentieren, um sich bei der weiteren Verarbeitung der Daten dessen bewusst zu sein, dass die Genauigkeit der Emotionserkennung basierend auf dieser Modalität vermutlich gering ist.

Leitlinie SYM₄:

Achten Sie auf die passende Beleuchtung des kindlichen Gesichts.

Bei Videoaufnahmen ist es wichtig auf geeignete Beleuchtungsbedingungen zu achten. Zu starke Beleuchtung führt zu Überbelichtung, was wiederum einen Verlust der Bildqualität zur Folge hat; z.B. sind weniger Merkmale des Gesichts erfassbar, welche essentiell für die Emotionserkennung sind. Zu schwache Beleuchtung hingegen führt zu Unterbelichtung, was beispielsweise einen Verlust von Schatten zur Folge hat, die wichtig für die Erkennung von Gesichtsmarkmalen sind. Überbelichtung, Unterbelichtung und ungleichmäßige Beleuchtung wirken sich sehr ungünstig auf die Emotionserkennung basierend auf Gesichtsausdrücken aus.

Um die Beleuchtungsbedingungen zu überprüfen, können Sie vorab ein Testvideo aufnehmen. Passen Sie die Beleuchtung entsprechend an, um Details des Gesichts wie Nase, Augenbrauen und Mund gut erfassen zu können. Achten Sie dabei auch auf sich im Raum befindliche Fenster (welche kein guter Hintergrund sind, da sie je nachdem wo genau das Kind sitzt eine ungleichmäßige Beleuchtung des Gesichts verursachen können).

Leitlinien für die Erhebung kindlicher Lautäußerungen

Leitlinie SYM5:

Wählen Sie einen möglichst ruhigen Raum – Achten Sie dabei insbesondere auf mögliche Hintergrundgeräusche und vermindern Sie Störgeräusche, die durch Einrichtungsgegenstände im Raum hervorgerufen werden können.

Um hochqualitative Aufnahmen der kindlichen Stimme zu erheben, sollten die Interventionseinheiten in einer ruhigen Umgebung mit möglichst geringen Hintergrundgeräuschen stattfinden. Vermeiden Sie also beispielsweise Straßenlärm, Geräusche von nahegelegenen Spielplätzen, etc.

Geräusche am Gang vor dem Therapieraum kann man oftmals reduzieren, indem man ein Informationsschild vor die Tür stellt auf dem zu lesen ist, dass aufgrund einer laufenden Studie/Therapie um Ruhe gebeten wird. Geräusche von der Straße können verringert werden, indem die Fenster während der Datenerhebung geschlossen bleiben.

Wählen Sie für Ihre Studie/Therapie bitte einen möglichst reflexionsarmen Raum (geringe Nachhallzeit). Es kann hierbei auch hilfreich sein, Vorhänge aufzuhängen, schallabsorbierende Einrichtungsgegenstände zu ergänzen und/oder den Raum mit Stofftieren oder Pölstern zu dekorieren (setzen Sie Dekogegenstände jedoch sparsam ein, um die Kinder nicht abzulenken). Stofftiere/Pölder können dem Kind auch dabei helfen sich wohler im Raum zu fühlen.

Versuchen Sie während der Datenerhebung Störgeräusche von beispielsweise Stühlen, Tischen, Türen, Einrichtungsgegenständen und Fenstern möglichst zu vermeiden. Hier einige Vorschläge wie das gelingen kann: Schließen Sie die Fenster vor dem Start der Datenerhebung. Legen Sie alle benötigten Gegenstände bereits vor Beginn der Datenaufzeichnung griffbereit auf einen Tisch oder lagern Sie diese in einer Tasche neben sich, sodass Sie während der Therapie nicht aufstehen müssen, um sie aus einem Schrank zu holen. Platzieren Sie Teppiche unter Tisch, Stühle und Füße von allen Personen, die sich während der Datenerhebung im Raum befinden.

Sie sollten auch Therapeut*innen und Eltern darum bitten, während der Datenerhebung möglichst wenige Geräusche mit Schuhen, Händen bzw. Objekten zu produzieren.

Auch bei der Wahl der richtigen Position für das Mikrofon sollten Sie mögliche Hintergrundgeräusche beachten: Es ist beispielsweise nicht empfehlenswert das Mikrofon auf Tischen oder Kleidungsstücken anzubringen, die laute Geräusche

produzieren, wenn sie berührt oder bewegt werden, da so vom Spracherkennungssystem weniger kindliche Lautäußerungen detektiert werden könnten.

Leitlinie SYM6:

Achten Sie darauf, dass sich möglichst selten kindliche Lautäußerungen mit Stimmen von anderen Personen überlagern.

Die Ergebnisse einer audiobasierten Emotionserkennung können sich verbessern, wenn vorab eine Stimmerkennung durchgeführt wird [9]. Die Stimme des Kindes automatisiert zu erkennen ist jedoch schwierig, wenn sie von Stimmen anderer Personen überlagert wird [10]. Daher empfehlen wir, dass möglichst wenige Personen während der Therapie im Raum sind. Wahrscheinlich ist es nicht möglich, dass der*die Therapeut*in versucht weniger zu sprechen, aber man könnte andere anwesende Personen wie zum Beispiel die Eltern bitten während der Therapie möglichst nicht zugleich mit dem Kind zu sprechen. Auch Therapeut*in und Eltern sollten versuchen nicht zugleich zu sprechen und sollten dazu ermutigt werden jegliche Konversationen, die nicht direkt mit dem Ablauf der Therapieeinheit zu tun haben, soweit möglich auf später zu verschieben.

Zu Studienzwecken empfehlen wir zu notieren wie viele Erwachsene sich während der Therapieeinheit im Raum befinden.

Leitlinie SYM7:

Machen Sie den Standort des Mikrophons davon abhängig, wo im Raum sich das Kind und andere Personen befinden.

Um hochqualitative Aufnahmen der kindlichen Stimme zu erheben, ist es empfehlenswert das Mikrophon näher zum Kind als zu anderen Personen bzw. dem Roboter zu stellen, damit die kindliche Stimme im Vergleich zu den anderen Stimmen entsprechend lauter aufgenommen wird. Der*die Therapeut*in sollte – da er*sie vermutlich oftmals während der Therapie spricht – so weit wie möglich entfernt vom Mikrophon sitzen. Motivieren Sie das Kind dazu lauter zu sprechen, falls es besonders leise spricht. Sie könnten dem Kind beispielsweise sagen, dass der Roboter es besser versteht, wenn es lauter spricht.

4.3. Technologien und Hardware (TECH)

Es war herausfordernd für uns, die Hardware zur Erhebung der emotionalen Zustände im Rahmen unserer Beobachtungsstudien auszuwählen, anzuwenden und richtig zu konfigurieren. Demnach beziehen sich die nachfolgenden Leitlinien oftmals auf spezifische Geräte und lassen sich möglicherweise nicht eins zu eins auf andere Hardware übertragen.

Leitlinie TECH₁:**Achten Sie auf eine sichere und nicht ablenkende Umgebung für das Kind.**

Die Therapieumgebung sollte sicher sein; dies betrifft natürlich in erster Linie das Kind, aber auch den Roboter und die Aufnahmegereäte. Bauen Sie die Hardware bereits vor Therapiebeginn auf. Der Roboter sollte stabil und sicher platziert werden, sodass er während der Interaktion mit dem Kind nicht versehentlich umfallen kann. Achten Sie auch auf Kabel der angeschlossenen Geräte. Wenn möglich sollten diese nicht neben dem Kind platziert werden, um zu vermeiden, dass das Kind stolpert oder den Messaufbau zerstört, indem es am Kabel zieht. Beachten Sie, dass Kinder mit ASS sehr intensiv darauf reagieren könnten, wenn der Roboter oder ein Messgerät umfällt.

Neben der Sicherheit ist auch die Ablenkung des Kindes ein wichtiger Aspekt, den es zu beachten gilt. Daher sollte darauf geachtet werden, dass das Kind so wenige Elemente der Hardware wie möglich wahrnimmt. Computer und Aufnahmegereäte könnten hinter einer Platte, einem Vorhang oder sogar in einer Schachtel versteckt werden und Kabel könnten am Boden verklebt werden. Die Kabel können darüber hinaus mit Teppichen versteckt werden.

Leitlinie TECH₂:**Beachten Sie bei der Auswahl der Hardware den Komfort für das Kind sowie den Aufbau- bzw. Einstellungsaufwand.**

Einige Geräte benötigen ein aufwändiges Einstellungsprozedere, um die Daten aufnehmen und entsprechend verarbeiten zu können. Außerdem könnten einige Geräte vom Kind als störend empfunden werden (siehe TECH₄).

Es ist möglich, dass eine Stimmerkennung nicht mit allen Arten an Mikrofonen gut funktioniert. Testen Sie die Genauigkeit der Stimmerkennung insbesondere dann sorgfältig, wenn Sie Mikrophone verwenden möchten, die Kinder mit ASS unter Umständen stören könnten, wie Mikrophone, die an der Kleidung montiert werden. In unserer ersten Beobachtungsstudie fanden wir heraus, dass die Stimmerkennung basierend auf den vom Knopflochmikrofon erhobenen Daten nicht besser funktionierte als basierend auf den Daten des am Tisch platzierten Mikrophons [10], dass das Knopflochmikrofon aber von manchen Kindern als störend empfunden wurde. Daher entschieden wir, es in der zweiten Beobachtungsstudie nicht einzusetzen.

Ein weiteres Beispiel sind Eye-Tracking-Geräte, die es oftmals erfordern, dass das Kind still sitzt oder steht und Punkte fixiert, die nach und nach an verschiedenen Stellen auf einem Bildschirm dargeboten werden. Es ist für ein Kind häufig schwierig, einem solchen Kalibrierungsvorgang zu folgen, insbesondere wenn es sich um ein Kind mit ASS handelt. Wir empfehlen daher folgendes: (1) Finden Sie ein Eye-Tracking-Gerät, das entweder keine Kalibrierung erfordert oder mit gelegentlichen Fehlern in der Kalibrierung umgehen kann; (2) versuchen Sie das Kalibrierungsprozedere für das Kind

zu vereinfachen; (3) verwenden Sie das Eye-Tracking-Gerät nur mit älteren Kindern bzw. Kindern, die dazu in der Lage sind, die Kalibrierung erfolgreich zu absolvieren.

Beachten Sie auch, dass manche Eye-Tracking-Geräte (nicht alle) Daten nur in einem relativ kleinen Bereich aufnehmen und analysieren können. Um qualitativ hochwertige Daten zu erheben, kann es notwendig sein, dass das Kind sich nicht nach vorne, hinten oder seitwärts bewegt, was schwierig zu kontrollieren ist. Daher ist es empfehlenswert das Gerät vor Therapiebeginn umfangreich zu testen. In unseren Beobachtungsstudien stellte sich zumeist die Kalibrierung des Eye-Tracking-Geräts und auch die Analyse der Daten als schwierig heraus.

Überlegen Sie, ob es Alternativen gibt, um Daten einer bestimmten Modalität zu erheben – so können gewisse Blickbewegungsmuster mittels spezieller Algorithmen auch von Videoaufnahmen gewonnen werden [34].

Leitlinie TECH₃:

Wenn Hardware am Kind selbst angebracht werden muss (z.B. an den Handgelenken oder an der Kleidung), kann eine Gewöhnungsphase vor Therapiebeginn hilfreich sein.

Beachten Sie, dass einige Kinder sich in Gegenwart der Aufnahmegeräte möglicherweise unwohl fühlen; dies ist insbesondere dann der Fall, wenn diese an ihrem Körper oder an ihrer Kleidung getragen werden müssen. Versuchen Sie daher Hardware zu finden, die nicht am Körper oder an der Kleidung getragen werden muss. Ist dies nicht möglich, empfehlen wir eine Gewöhnungsphase vor Therapiebeginn, die auch an die individuellen Bedürfnisse der Kinder angepasst werden kann.

Beispielsweise kann es für einige Kinder mit ASS schwierig sein, ein Knopflochmikrofon an ihrer Kleidung zu akzeptieren. Überlegen Sie daher genau, ob ein solches Mikrofon zur Erreichung ihrer Studien- bzw. Therapieziele erforderlich ist. Sofern dies der Fall ist, überlegen Sie sich, wie Sie die Kinder an das Mikrofon gewöhnen könnten. Es kann zum Beispiel hilfreich sein, wenn Roboter, Therapeut*in, oder Eltern ähnliche Mikrofone (oder Attrappen) tragen. Bei manchen Kindern könnte es auch helfen, wenn man ihnen erklärt, dass der Roboter sie nur verstehen kann, wenn sie dieses Mikrofon tragen.

Wählen Sie kabellose Varianten für Hardware mit der Sie physiologische Parameter aufzeichnen möchten (so wie das E4-Armband, das wir in unseren Studien verwendet haben). Armbänder sind eine gute Möglichkeit, da sie optisch einer Uhr ähneln und damit womöglich einfacher vom Kind akzeptiert werden. Einigen Kindern könnte es auch helfen, wenn Forscher*in bzw. Therapeut*in oder ein Elternteil ihnen das Gerät zuerst zeigt und zum Beispiel einen Knopf drücken lässt. Andere Kinder profitieren eventuell davon, das Armband in ein leicht farbiges Tuch zu wickeln. Man könnte auch Kaspar, Therapeut*in und Eltern mit ähnlichen Armbändern (oder Attrappen) ausstatten, um den Kindern zu vermitteln, dass Armbänder zu dieser speziellen Therapiesituation einfach dazugehören.

Seien Sie sich dessen bewusst, dass Kinder Geräte, die sich in Reichweite befinden, gerne mit ihren Händen untersuchen (z.B. Armbänder, Knopflochmikrophone). Das passiert besonders dann, wenn man ihnen am Beginn gezeigt hat, wie das Gerät ein- oder ausgeschaltet werden kann. Zeigen Sie daher dem Kind nur dann wie ein Gerät funktioniert, wenn keine andere Gewöhnungsstrategie hilft.

Leitlinie TECH₄:

Wählen Sie Audioeinstellungen am Mikrophon, mit denen die kindliche Stimme deutlich erfasst werden kann.

Wenn das Mikrophon so eingestellt ist, dass es Audiodaten nur sehr leise aufzeichnet, funktioniert die Stimmerkennung gegebenenfalls nicht gut und es werden nur wenige oder keine kindlichen Lautäußerungen detektiert. Stellen Sie das Mikrophon daher vor dem Therapiebeginn sorgfältig ein. Passen Sie die Mikrophonkonfiguration an den Raum, mögliche Hintergrundgeräusche, Position des Mikrophons im Raum und die Lautstärke der kindlichen Stimme an. Wir empfehlen, vor Therapiebeginn eine Testaufnahme (ohne Kind) durchzuführen.

4.4. Aktivitäten in der Interaktion mit sozialen Robotern (INT)

Einige Untersuchungen beschrieben bereits das große Potenzial von Robotern in der Therapie von Kindern mit ASS was die Entwicklung ihrer sozialen und emotionalen Fähigkeiten betrifft. Roboter gelten als soziale Vermittler, die Kinder dazu motivieren sich an Aufgaben zu beteiligen, und dabei für eine voraussagbare, gleichbleibende Umgebung sorgen; so sind voraussagbare Regeln eine wichtige Voraussetzung um soziales Verhalten voranzubringen. Wir wissen nicht genau, warum Kinder mit ASS lieber mit Robotern interagieren, die Menschen ähnlich sehen, als mit Menschen selbst. Unabhängig vom Grund haben sich soziale Roboter als eine Möglichkeit erwiesen, die sozialen Hindernisse der Kinder zu überwinden und sie in Interaktionen zu involvieren. Sobald die Interaktion dann schließlich stattfindet, haben wir die einzigartige Gelegenheit dem Kind nach und nach soziale und emotionale Fähigkeiten beizubringen. [1]

Leitlinie INT₁:

Passen Sie die Ziele und Technologien der Emotionserkennung an die Therapie an.

Wir analysierten auf welche Fähigkeiten frühere Studien zu roboter-unterstützten Intervention fokussierten. Die folgende Liste soll einen Überblick über die detektierten Fähigkeiten geben und Ihnen zeigen für welche Zwecke soziale Roboter eingesetzt werden könnten [1]:

- Soziale Konvention (begrüßen, zusammen singen, etwas mitteilen, eine Interaktion beenden, danken, bitten);
- Soziale Interaktion (eine Interaktion beginnen/initiieren, Sprecher*innenwechsel, Bewegungen folgen/imitieren, soziale Aufmerksamkeit, sich an einer Aktivität beteiligen, sensorische Verarbeitungsfähigkeiten, Augenkontakt, Konversation, sozio-emotionales Verhalten, zusammen ein Spiel spielen, ein Objekt verlangen);
- Soziales Antworten (dem Blick von anderen folgen, eine Aufforderung befolgen, auf den eigenen Namen reagieren);
- Emotionale Fähigkeiten (Erkennen der Basisemotionen, Emotionen und Geräusche miteinander verknüpfen, Kontext-Emotions-Assoziation, zwischen Gedanken und Emotionen unterscheiden, Emotionen von anderen erkennen, Emotionen, die im Rahmen des Robotertrainings erworben wurden nachmachen);
- Kontrollieren (den Roboter berühren um ihn wieder glücklich zu machen, eigenes Verhalten in Situationen, in denen man verärgert oder traurig ist, anpassen, eigene negative Emotionen kontrollieren in sozialen Situationen);
- Andere Fähigkeiten (Selbstfürsorge, kognitive Fähigkeiten, Lernkompetenz verbessern, Trommelspiele, Synchronität zwischen Ober- und Unterkörper bei rhythmischen Bewegungen, sich nach Zahlen bewegen, im Rhythmus bewegen, bestimmte Farben auswählen).

Denken Sie sorgfältig darüber nach wie genau die Intervention und die Roboter-Kind-Interaktion von Emotionserkennung profitieren könnte. Dafür empfehlen wir, die Ziele der Emotionserkennung für ein spezielles Kind und einen Therapiezweck zu definieren. Eine genaue Emotionserkennung ist nicht in jedem Fall erforderlich.

Automatische Emotionserkennungstechnologien sind vor allem dann wichtig, wenn dem Kind emotionale Fähigkeiten beigebracht werden sollen. Welche Emotionen erkannt werden sollen könnte an den Interventionszweck und die kindlichen Fähigkeiten angepasst werden. So wäre es beispielsweise denkbar, während dem Training von Lächeln nur die Emotion "glücklich" zu detektieren.

Auf der anderen Seite könnte es interessant sein den allgemeinen Gemütszustand und die Einstellung eines Kindes während einer Intervention zu erfassen. In dem Fall könnte man zum Beispiel versuchen zwischen Angst, Ärger, Langeweile und Bereitschaft zur Interaktion zu unterscheiden.

In einigen Fällen könnte es auch ausreichen automatisch zu erkennen, ob ein Kind weint oder nicht bzw. zwischen positiven und negativen Reaktionen zu differenzieren.

Leitlinie INT₂:

Planen Sie eine Gewöhnungsphase ein, damit das Kind den Roboter und andere verwendete Hardware kennenlernen kann.

Für die meisten kleinen Kinder, speziell, wenn sie bisher noch keine Erfahrung mit roboter-unterstützter Intervention machen konnten, kann eine Gewöhnungsphase mit

dem Roboter und anderer Hardware vor der eigentlichen Datenerhebung hilfreich sein (dies betrifft vor allem Geräte, die am Körper des Kindes angebracht werden müssen – siehe TECH₄). Manche Kinder müssen sich vorab eventuell auch erst an den unbekanntem Raum und möglicherweise fremde Personen gewöhnen.

Es ist wichtig, dass das Kind die Interaktion jederzeit beenden darf (siehe GEN₃). Therapeut*in bzw. Forscher*in sollten versuchen das Kind zur Interaktion mit dem Roboter zu motivieren, dürfen aber keinen Druck ausüben.

Leitlinie INT₃:

Planen Sie das Interventionsprogramm im Vorfeld, aber ändern sie es wenn notwendig während der Therapie ab.

Die Intervention sollte dem Kind folgen, d.h. es ist gut im Vorfeld einen ungefähren Plan der Intervention zu haben, aber Forscher*innen/Therapeut*innen sollten flexibel genug bleiben, um die konkreten Interventionsaktivitäten bei Bedarf an die aktuellen Bedürfnisse des Kindes anpassen zu können.

Am Beginn der roboter-unterstützten Intervention empfehlen wir „Eisbrecher“ einzusetzen. Dies kann zum Beispiel die Vorstellung des Roboters sein oder ein gemeinsam gesungenes Lied. Es kann sinnvoll sein, verschiedene Arten von Eisbrechern parat zu haben, um den Eisbrecher an das Alter, die Interessen und die verbalen Fähigkeiten des jeweiligen Kindes anpassen zu können.

Bei der Intervention selbst muss beachtet werden, dass ASS ein sehr heterogenes Störungsbild ist. Daher sollte die Intervention an die Aufmerksamkeitsspanne und andere Charakteristika des Kindes, wie verbale Fähigkeiten, kognitive Fähigkeiten, mögliche Komorbiditäten, Alter und Vorlieben angepasst werden. Dies kann beispielsweise über Flexibilität hinsichtlich Länge der Intervention, Auswahl, Reihenfolge und Länge der konkreten Interventionsaktivitäten und Anzahl und Länge der Pausen erfolgen.

Leitlinie INT₄:

Bereiten Sie verschiedene Aktivitätsarten und Schwierigkeitsstufen vor.

Wir empfehlen das Aktivitätsprogramm des Roboters an das jeweilige Alter, verbale und kognitive Fähigkeiten, Aufmerksamkeitsspanne, individuelle Interessen und therapeutische Ziele anzupassen. Es ist sinnvoll Aktivitäten in verschiedenen Schwierigkeitsstufen vorzubereiten, um jedem Kind mit ASS die am besten geeignete Intervention zukommen zu lassen. Damit man schließlich schnell die passenden Aktivitäten auswählen kann, ist es hilfreich, diese beispielsweise nach ihren Schwierigkeitsstufen zu annotieren.

Bei der Auswahl des Interventionsprogramms sollten die Interessen der Kinder beachtet werden. Vielen Kindern gefällt zum Beispiel Singen sehr gut. Daher könnte das Singen

eines Liedes sich gut als Eisbrecher eignen und/oder zwischendurch eingesetzt werden, um das Kind zu beruhigen, wenn es gestresst wirkt. Andererseits ist das Singen eines Liedes vielleicht nicht für alle Kinder jedes Alters geeignet. Daher wäre es hilfreich bereits vor der roboter-unterstützten Interventionseinheit die Interessen des jeweiligen Kindes herauszufinden, um jene Aktivitäten einzuplanen, die dem Kind wahrscheinlich gefallen. Ein Beispiel für Kinder mit guten verbalen Fähigkeiten könnte ein Vokal-Memory-Spiel sein.

Leitlinie INT5:

Führen Sie mehrere Therapieeinheiten mit einem Kind durch.

Um den Interventionserfolg von roboter-unterstützter Therapie im Vergleich zu standardmäßig eingesetzten therapeutischen Ansätzen zu evaluieren, sollte man das Kind über einen längeren Zeitraum begleiten; d.h. mit dem Kind mehrere roboter-unterstützte Therapieeinheiten durchführen. Es könnte eine Herausforderung darstellen, Kinder für Langzeitstudien zu rekrutieren. In diesem Fall kann es hilfreich sein, roboter-unterstützte Therapieeinheiten in den regulären Therapieplan der Kinder zu integrieren und/oder Familien mit Aufmerksamkeiten für die Studienteilnahme zu belohnen, wie etwa Aufwandsentschädigungen oder Spielzeug für das Kind.

4.5. Datenverarbeitung (PROC)

Im Rahmen unserer Literaturrecherche stießen wir auf einige Herausforderungen, die frühere Studien hinsichtlich der Verarbeitung der Daten der verschiedenen Modalitäten hatten. Diese Herausforderungen können in drei Kategorien eingeteilt werden: (1) die erhobenen Daten erfüllen nicht die notwendigen Qualitätskriterien (sie beinhalten keine wesentlichen Informationen, die auf kindliche Emotionszustände schließen lassen), (2) einzelne Modalitäten sind aufgrund von atypischen Mustern in den Daten schwierig auszuwerten, (3) die erhobenen Daten, die auf kindliche Emotionen rückschließen lassen, sind widersprüchlich.

Leitlinie PROC1:

Achten Sie auf die Synchronisation der Daten, wenn Sie mehrere Modalitäten erheben.

Multimodale Emotionserkennungssysteme ermöglichen eine höhere Klassifikationsgenauigkeit als Systeme, die nur auf einer einzelnen Modalität beruhen [6]. Die am meisten verwendeten Kanäle sind: RGB Video, Tiefenbildvideo, Audiosignale und physiologische Signale. Werden allerdings verschiedene Geräte verwendet um die Daten dieser Kanäle zu erheben, kann es eine Herausforderung darstellen, die Daten zu

synchronisieren. Außerdem ist es nicht so einfach, die Aufnahme aller Geräte zugleich zu starten.

Daher ist es wichtig, eine geeignete Synchronisationsstrategie zu erarbeiten und vorab zu testen. Man kann beispielsweise den exakten Startzeitpunkt jedes Geräts notieren. Manche Geräte bieten auch die Möglichkeit, einen Zeitstempel festzulegen. Für audiovisuelle Kanäle kann man eine Filmklappe verwenden. Darüber hinaus ermöglichen spezielle Programme den simultanen Start aller Geräte. Egal welche Methode Sie schließlich wählen, ist es für die Weiterverarbeitung der erhobenen Daten sehr wichtig, dass sie die Synchronisation der Kanäle ermöglicht.

Leitlinie PROC₂:

Achten Sie bei der Erhebung von Daten mehrerer Modalitäten auf inkonsistente Ergebnisse

Bei der Anwendung von multimodalen oder mehrkanaligen Emotionserkennungssystemen (siehe CH₃ und REP₂) sollten Sie auf inkonsistente Ergebnisse der Einzelmodalitäten achten.

Zu den Herausforderungen von multimodaler Integration zählt die Eingliederung der Ergebnisse in unterschiedliche Emotionsrepräsentationsmodelle und der Umgang mit inkonsistenten Ergebnissen.

Wir beobachteten Inkonsistenzen in den Erkennungsergebnissen von Experimenten, sowohl was die Analyse unterschiedlicher als auch gleicher Inputkanäle betrifft. [12] Zum Beispiel wurde in einem Experiment zu einem didaktischen Management-Spiel eine signifikante Diskrepanz zwischen den selbstevaluierten und den detektierten Emotionszuständen berichtet [13]. In einer anderen Studie fanden wir Unterschiede in den detektierten Emotionszuständen basierend auf Gesichtsausdrücken, die mit zwei Kameras aufgenommen wurden: eine war unter und eine über dem Monitor platziert [7].

Bei der Analyse des gleichen Beobachtungsfensters mit derselben Person und Umgebung könnten unterschiedliche Algorithmen unterschiedliche Ergebnisse hinsichtlich der Emotionserkennung generieren.

Wir empfehlen vor der tatsächlichen Datenverarbeitung und Analyse zu evaluieren, ob Kanäle während der Aufnahme temporär nicht verfügbar waren und die Datenqualität zu überprüfen. Darüber hinaus ist es aus unserer Sicht essenziell, bei der Dateninterpretation auf mögliche inkonsistente Emotionserkennungsergebnisse zu achten und diese zu berichten.

Leitlinie PROC₃:

Annotieren Sie wichtige Ereignisse während der Datenerhebung (wenn möglich) oder im Zuge der Datenverarbeitung.

Im Zuge einer Roboter-Kind-Interaktionssitzung können verschiedene Ereignisse eintreten, die die erhobenen Daten beeinflussen und somit zu Falschinterpretationen führen können. Daher ist es wichtig, diese Ereignisse sehr genau mitzunotieren. Die Notizen sollten Unterbrechungen erfassen, z.B. eine weitere Person betritt den Raum, neu auftretende Geräusche von außerhalb des Raums (oder auch im Inneren des Raums – z.B. etwas fällt hinunter). Sie sollten außerdem atypisches Verhalten des Kindes notieren wie auch Ereignisse, die mit der Datenerhebungshardware zu tun haben, wie beispielsweise eine leere Batterie, ein Sensorfehler oder dass das Kind einen Sensor abnimmt.

Wenn möglich sollten diese Notizen in Echtzeit erfolgen (manchmal benötigt man dafür eine zusätzliche Person als Beobachter*in im Raum). Es ist uns bewusst, dass es manchmal schwierig sein kann, alle Ereignisse festzuhalten. Eine Annotation der Aufnahmen im Nachhinein (wenn Sie die gesamte Szene mit einer Videokamera festgehalten haben) ist auch möglich, allerdings zeitaufwändig.

Eine Annotation im Nachhinein ist wichtig für die Interpretation der Ergebnisse und kann wertvoll sein was die Veröffentlichung eines Datensatzes für zukünftige Studien betrifft (siehe RES₂). Für eine akkurate Annotation ist es wichtig, Ereignisse, Zustände und Aufgaben genau zu definieren, die Annotation von mehreren Personen durchführen zu lassen (inklusive der Überprüfung der Übereinstimmung der Annotationen), und eine wiederholbare Methode (bzw. Werkzeug) für den Annotationsprozess anzuwenden.

Leitlinie PROC₄:

Was den Audiokanal betrifft ist es hilfreich, wenn die kindliche Stimme zuerst von anderen Geräuschen unterschieden wird.

Um den Emotionszustand basierend auf der Stimme des Kindes zu erkennen, müssen wir zuerst jene Segmente detektieren, in denen kindliche Lautäußerungen auftreten. Dafür ist es wichtig, die kindliche Stimme von anderen Stimmen und Hintergrundgeräuschen zu unterscheiden. Wir beobachteten, dass jene Segmente, in denen kindliche Stimme detektiert wurde, meist laute, klar artikulierte kindliche Lautäußerungen beinhalteten und zugleich kaum Hintergrundgeräusche vorhanden waren [10]. Im Falle von unvermeidbaren Hintergrundgeräuschen – wie beispielsweise Geräusche aus einem nahe gelegenen Warteraum – ist es hilfreich, wenn die kindliche Stimme deutlich lauter aufgenommen wird als das Hintergrundgeräusch. Außerdem ist es hilfreich, dem Stimmerkennungsdetektor eine große Menge an verschiedenartigem Trainingsmaterial zur Verfügung zu stellen. Wenn das Trainingsmaterial zum Beispiel keine ausreichende Menge an Gesang, nachgeahmten Tierlauten, oder dem Weinen von Kindern beinhaltet, ist es für den Stimmerkennungsdetektor schwieriger, in solchen Daten eine kindliche Stimme zu detektieren [10]. Natürlich sollte man auch im Hinterkopf behalten, dass sich bestimmte Stimmerkmale mit dem Alter verändern.

Bei der Studienplanung muss beachtet werden, dass die Genauigkeit eines Stimmerkennungsdetektors nur evaluiert werden kann, wenn jedes Geräusch wie kindliche Stimme, Stimmen von anderen Personen, Geräusche vom Roboter und

Hintergrundgeräusche annotiert werden. Es kann hilfreich sein, eine Referenz-Aufnahme der Stimme jedes Kindes zu machen (siehe PROC₅).

Leitlinie PROC₅:

Für manche Modalitäten sollte eine Referenz-Aufnahme des Kindes erfolgen.

Eine Referenz- oder Baseline-Aufnahme könnte für die Datenverarbeitung und -analyse hilfreich sein. Unter einer Baseline versteht man einen fixen Referenzpunkt, der für Vergleichszwecke herangezogen werden kann. Für den Audiokanal beispielsweise bedeutet das, dass unabhängig von der Therapie selbst eine kurze Aufnahme der kindlichen Stimme ohne jegliche Störgeräusche erfolgen könnte. Für physiologische Daten bedeutet es, dass eine Referenzmessung in Ruhe erfolgen könnte (bevorzugt in einem Raum, in dem sich der Roboter nicht befindet).

Menschen verhalten sich sehr unterschiedlich was die Reaktionen ihres Nervensystems auf Emotionen betrifft. Zum Beispiel gibt es Menschen, die sehr intensiv, mittelmäßig oder nur wenig auf Emotionen reagieren. Physiologische Reaktionen verändern sich auch mit dem Alter. Kinder mit ASS sind sogar noch divergenter in ihren Reaktionen was mit ihren individuellen Defiziten zu tun hat. Eine Referenz-Aufnahme (im Ruhezustand) stellt eine Möglichkeit dar die Diversität bei der Datenverarbeitung zu berücksichtigen.

Studien zeigten, dass die Referenzwerte für Kinder mit ASS sich von denen typisch entwickelter Kinder unterscheiden. Zum Beispiel beschrieb Studie [16] wichtige Beobachtungen hinsichtlich der Sensoren und Technologien, die für automatische Emotionserkennung verwendet werden können: (1) Kinder mit ASS wiesen bei Referenzmessungen eine signifikant niedrigere Amplitude der respiratorischen Sinusarrhythmie sowie eine schnellere Herzfrequenz als typisch entwickelte Kinder auf, was auf eine insgesamt niedrigere Regulation der Herzfrequenz durch den Vagusnerv hindeutet; (2) ein großer Anteil der Kinder mit ASS wiesen ungewöhnlich hohe sympathische Aktivität auf, d.h. Hautleitfähigkeitsreaktion. Eine andere Studie [17] verglich galvanische Hautreaktionsdaten von Kindern mit ASS und typisch entwickelten Kindern. Die Studie zeigte, dass Kinder mit ASS unregelmäßigere Hautleitfähigkeitsmuster aufwiesen als typisch entwickelte Kinder.

Was den Audiokanal betrifft könnten Sie vor oder am Anfang der Intervention eine Stimmprobe des Kindes aufnehmen, um den Stimmerkennungsdetektor so zu optimieren, dass er die Stimme des Kindes besser erkennt. Dafür wäre es ideal Stimmmaterial im Umfang von etwa einer Minute aufzunehmen. Hier geben wir einige Vorschläge wie man ein solches Referenzmaterial erheben kann: (1) Ältere Kinder könnten zum Beispiel über einen typischen Schultag, ihren letzten Urlaub oder einen kürzlich gehabt Traum erzählen (ohne dabei von anderen im Raum unterbrochen zu werden); (2) Jüngere Kinder, die damit einverstanden sind Kopfhörer zu tragen, könnten einige Nachahmungsaufgaben absolvieren, z.B.: „Bitte sprich den folgenden Satz nach: ‚Mein Freund hat einen Apfelbaum in seinem Garten.‘, ‚Kaspar liebt Musik und neue Freunde kennenzulernen.“; (3) Jüngere Kinder, die keine Kopfhörer tragen möchten

könnten diese Nachahmungsaufgaben vor Ort von einem*er Therapeut*in oder Forscher*in gestellt bekommen, der*die die Aufnahme pausiert während er*sie selbst spricht. Die Erhebung dieses Stimmmaterials könnte sogar in eine Art Spiel oder Routine implementiert sein, die dazu dient, den Roboter zu „aktivieren“: „Bitte wiederhole alle Sätze, die ich dir gleich sagen werde, um Kaspar aufzuwecken, sodass er mit dir/uns lustige Spiele spielt.“

Natürlich funktioniert die Erhebung solches Stimmreferenzmaterials nicht für nonverbale/minimal verbale Kinder und auch nicht für sehr schüchterne/unglückliche Kinder oder Kinder die (noch) keine verbalen Imitationsfähigkeiten haben oder (noch) keine kurzen Geschichten erzählen können.

Leitlinie PROC6:

Kombinieren Sie Methoden, um die Daten richtig mit Emotionszuständen zu annotieren.

Da Kinder mit ASS oft unterschiedliche, teilweise einzigartige, Arten haben ihre Emotionen auszudrücken, kann es schwierig sein, ihre Emotionszustände richtig zu annotieren. Daher wird die Kombination von verschiedenen Methoden empfohlen wie beispielsweise das Kind zu fragen wie es sich fühlt, Expert*innen die Daten annotieren zu lassen, Eltern bzw. Bezugspersonen, die das Kind gut kennen, um Überprüfung der Annotationen bitten und Annotationen basierend auf speziellen Stimuli zu machen.

Die Daten korrekt zu annotieren ist generell eine Herausforderung bei Emotionserkennung, nicht nur wenn es Kinder mit ASS betrifft. Es ist schwierig den inneren Emotionszustand (Grundwahrheit) einer Person auf kontinuierliche Weise zu bestimmen; dies gilt sogar für die Person selbst. Einige Strategien könnten angewandt werden, um die Daten zu annotieren. In Studie [18] annotierten trainierte Therapeut*innen Aufnahmen von mehreren Therapieeinheiten, was eine häufig angewandte Methode ist. Studie [19] ließ Eltern subjektiv Emotionszustände bewerten und verglich die Ergebnisse mit den Berichten von Therapeut*innen; die Übereinstimmung betrug ungefähr 83%. Studie [20] verwendete Selbst-Berichte und kombinierte diese mit den Evaluationen von Therapeut*innen. Die Selbst-Berichte der Kinder mit ASS war nur teilweise übereinstimmend mit den Annotationen der Therapeut*innen. Die Berichte der Therapeut*innen wurden als „Grundwahrheit“ für die Klassifikation angenommen. Die Autor*innen berichteten, dass die „klassische“ Methoden der Emotionsannotation aufgrund der Kommunikationsschwierigkeiten der Kinder mit ASS schwierig anzuwenden sind. Um die Reliabilität der Annotationen zu erhöhen, schlugen sie vor, sowohl klinische Beobachter*innen als auch enge Bezugspersonen der Kinder in Studien einzuschließen. [20]

Wenn man eine Studie zu Emotionserkennung während Roboter-Kind-Interaktion plant, sollte man sorgfältig über die Annotationsstrategie nachdenken, um die entsprechenden Studienziele zu erreichen.

Leitlinie PROC7:

Es könnte schwierig sein, allgemeine Modelle anzuwenden, da Charakteristika und Schwierigkeiten von Kindern mit ASS sehr heterogen sind.

Autismus-Spektrum-Störung ist eine heterogene Erkrankung mit verschiedenen Symptomen und Schweregraden. So haben beispielsweise einige Kinder gute sprachliche Fähigkeiten, wohingegen andere nonverbal sind. Außerdem kann roboter-unterstützte Intervention eine wertvolle Alternative zu standardmäßig durchgeführten Therapieansätzen für Kinder verschiedenen Alters sein. Natürlich sollten die Interventionsaktivitäten aufgrund verschiedener Interessen und Fähigkeiten für 3-jährige Kinder nicht gleich sein wie für 12-jährige Kinder. Es wäre notwendig, Emotionserkennungstechnologien zu implementieren, die passend oder adaptierbar sind für verschiedene Sprachen, kulturelle Hintergründe und das Geschlecht des Kindes. Im besten Fall sollten diese Technologien auch für unterschiedliche akustische Umgebungen, Mikrofontypen und -positionen, Eye-Tracking-Geräte, Arten von Robotern, etc. geeignet sein. Weitere Studien werden gebraucht, um zu evaluieren, wie unterschiedliche Einstellungen und Geräte die Emotionserkennungsergebnisse beeinflussen. Eine große Menge an annotierten Daten – erhoben in verschiedenen Räumen, Situationen, Sprachen, mit verschiedenen Robotern, etc. – wäre hilfreich, um vorhandene Emotionserkennungsmodelle zu optimieren.

4.6. Emotionserkennung bei Kindern mit Autismus-Spektrum-Störung (EMO)

Zu den Defiziten von Kindern mit ASS zählen Schwierigkeiten beim Ausdrücken und Erkennen von Emotionszuständen. Hinsichtlich automatischer Emotionserkennung können atypische Arten Emotionen auszudrücken zu Falschannahmen hinsichtlich der inneren Emotionszustände eines Kindes führen. Die meisten Leitlinien in diesem Kapitel sind nur informativ, da man nicht ändern kann, wie Emotionsausdrücke entstehen.

Leitlinie EMO1:

Kinder mit ASS drücken ihre Emotionen anders aus.

Kinder mit ASS zeigen ungefähr den gleichen Intensitätslevel eines emotionalen Gesichtsausdrucks, aber ihre Muster können atypisch sein. Beispielsweise wurden signifikante Unterschiede zwischen Kindern mit hochfunktionalem Autismus und typisch entwickelten Kindern hinsichtlich ihrer Gesichtsausdrücke für Ekel und Traurigkeit über das gesamte Gesicht, Freude in oberen und unteren Gesichtsteilen und Überraschung nur in den unteren Gesichtsteilen gefunden. Im Gegensatz dazu wurden keine signifikanten Unterschiede hinsichtlich der Gesichtsausdrücke für Ärger und Angst gefunden [11]. Daher sind gewöhnliche Emotionserkennungsmodelle, die auf Daten von typisch entwickelten Kindern trainiert wurden, nur teilweise geeignet. Wir empfehlen, spezifische

Klassifikatoren anzuwenden, die mit Daten von Kinder mit ASS trainiert wurden, anstatt für alle Kinder die gleichen Klassifikatoren zu verwenden.

Leitlinie EMO₂:

Kinder mit ASS zeigen manchmal eine atypische Synchronisation von emotionaler Stimme/Gesichtsausdrücken/Gesten.

Einige Studien zeigten, dass Kinder mit ASS ihre Emotionen auf atypische Art und Weise zeigen. Eine Studie über die Koordination der Emotionsausdrücke mehrerer Modalitäten zeigt, dass die Koordination in der ASS-Gruppe kleiner ist als in der Gruppe der neurotypischen Kinder [21]. Kinder mit ASS und neurotypische Kinder unterscheiden sich nicht hinsichtlich der Ausdrucksstärke ihrer emotionalen Gesichtsausdrücke und verbalen Ausdrücke, aber die Gesichtsausdrücke und verbalen Ausdrücke sind bei Kindern mit ASS weniger miteinander koordiniert. Laut dem Literaturüberblick dieses Artikels tritt eine modalitätenübergreifende Koordination auch bei Gesichtsausdrücken und Gesten auf. Es wird auch berichtet, dass Kinder mit ASS eine atypische zeitliche Abfolge und Synchronisation von Bewegungen verschiedener Gesichtsregionen, reduzierte Intensität der oberen Gesichtsbewegungen, reduzierte Vielfalt der Gesichtsbewegungen, und mehr Ambiguität zeigen, da Ausdrücke für positive und negative Valenz sich nicht so unterscheiden wie für typisch entwickelte Kinder im gleichen Alter. Eine andere Studie bestätigte diese Beobachtungen [22], indem sie weniger Synchronisation der Bewegungen zwischen Gesichtsausdrücken, weniger komplexe Gesichtsdynamik und mehr Ambiguität berichtete. Die Auffälligkeiten in den Gesichtsausdrücken wurden als unabhängig vom Emotionstyp beschrieben (analysiert wurden Freude, Ärger, Traurigkeit und der neutrale Zustand).

Leitlinie EMO₃:

Manche Kinder mit ASS verwenden keine oder nur wenige sprachliche Ausdrücke bzw. Lautäußerungen während sie mit einem Roboter interagieren.

Wie in der Literatur vielfach beschrieben, haben die meisten Kinder mit ASS – teils große – Schwierigkeiten mit dem Spracherwerb [14, 15]. Daher ist es wahrscheinlich, dass zumindest einige von ihnen keine oder nur wenig Sprache verwenden, wenn sie mit einem sozialen Roboter interagieren. Milling und Kolleg*innen [10] setzten kürzlich einen Deep Learning basierten Spracherkennungsdetektor ein und zeigten, dass dieser nur in etwa 4% der insgesamt aufgenommenen Zeit der roboter-basierten Interventionseinheiten kindliche Vokalisationen erkennen konnte. In dieser Studie wurden zwischen 0,6% und 20% der Interventionszeit kindliche Lautäußerungen detektiert. Emotionserkennung basierend auf Sprachsignalen könnte bei Kindern mit besseren lautsprachlichen Fähigkeiten besser funktionieren als bei Kindern mit geringeren lautsprachlichen Fähigkeiten.

Leitlinie EMO4:

Kinder mit ASS schauen oftmals auf die Seite oder nach unten während sie sprechen.

Einige Studien berichten, dass Kinder mit ASS atypische Verhaltensmuster zeigen, die einen Einfluss auf die Aufnahmen haben könnten. Das trifft insbesondere auf interaktive Aktivitäten zu. Zum Beispiel schauen Kinder während einer Konversation oftmals auf die Seite oder nach unten. Diese Verhaltensweise beeinflusst sowohl die Augenbewegungen als auch die Kopfbewegungen und hat einen Einfluss auf die Daten, die erhoben werden.

Da das kindliche Verhalten mit dem Störungsbild verknüpft ist und wir es nur begrenzt beeinflussen können, beschränkt sich die Empfehlung darauf, solche Verhaltensweisen zu dokumentieren.

Leitlinie EMO5:

Beachten Sie, dass Sie nur Ausdrücke von Emotionen erfassen.

Die Aktivierung des menschlichen Nervensystems induziert Veränderungen in Aktivitäten, die als Ausdrücke von Emotionen interpretiert werden können. Allerdings sind alle Aktivitäten und Modalitäten, die als Grundlage für Emotionserkennung herangezogen werden können, anfällig für Störungen und Falschinterpretationen.

Der innere emotionale Zustand (Grundwahrheit) ist schwer zu ermitteln, manchmal sogar für die Person selbst, die die Emotion empfindet. Es gibt keine Möglichkeit den aktuellen Emotionszustand eines Menschen zu bestimmen und sogar Selbstberichte könnten verzerrt sein.

Es ist wichtig zu beachten, dass die „Grundwahrheit“ der inneren Phänomene von Emotionen unbekannt bleibt und dass alle Emotionsannotationen verzerrt sind [12]. Menschen können auch einige emotionale Ausdrücke absichtlich verändern. Kinder mit ASS lächeln oft, wenn sie alleine gelassen werden, was aber nicht bedeutet, dass sie dabei Freude empfinden.

Daher sollten Interpretationen der Ergebnisse (Einschätzungen der Algorithmen) bei Emotionserkennung sich auf emotionale Ausdrücke beziehen und nicht auf die Emotionen, die ihnen zugeschrieben werden. Es wäre korrekter von emotionalen Ausdrücken und nicht vom Emotionszustand einer Person zu sprechen, wenn man beachtet, dass wir nur emotionale Ausdrücke beobachten und messen können. Automatische Lösungen erfordern weiterhin einen Menschen in der Schleife.

4.7. Design von Forschungsstudien (RES)

Diese Leitlinien wurden nicht nur für Therapeut*innen, sondern auch für Forscher*innen erstellt. Dieser Abschnitt bezieht sich insbesondere auf Forschungsstudien.

Leitlinie RES₁:

Beachten Sie die Zusammensetzung der Studien- und der Kontrollgruppe.

Beim Zusammenstellen von Studien- und Kontrollgruppe müssen einige Herausforderungen bewältigt werden. Zuallererst ist es schwierig, in den Gruppen ein ausgewogenes Geschlechterverhältnis herzustellen, da Autismus viel häufiger bei Jungen als bei Mädchen diagnostiziert wird. Ein ähnliches Problem besteht bei der Balancierung von Kindern hinsichtlich ihrer Fähigkeiten. Kinder mit mehr Schwierigkeiten könnten – je nach Art der Aktivität – öfter nicht an der Studie teilnehmen wollen. Bei der Zusammenstellung der Kontrollgruppe sollten die intellektuellen Fähigkeiten der Kinder berücksichtigt werden. Je nach Forschungsfrage sollte die Zusammenstellung der Kontrollgruppe gegebenenfalls eher anhand des Entwicklungsalters und der Fähigkeiten als anhand des chronologischen Alters der Kinder mit ASS erfolgen [2].

Leitlinie RES₂:

Legen Sie einen frei zugänglichen und gut beschriebenen Datensatz für zukünftige Studien an.

Das Forschungsfeld zur automatisierten Emotionserkennung würde von vielen erhobenen Datensätzen, die frei mit anderen Forscher*innen geteilt werden, profitieren. Die Datensätze sollten extrahierte Merkmale oder Rohdaten beinhalten, die anonymisiert sind und nach Emotionen annotiert wurden. Allerdings ist bislang nur ein einziger Datensatz verfügbar, der auf Emotionserkennung von Kindern mit ASS fokussiert. Er wurde im Rahmen des DE-ENIGMA-Projekts erhoben und kann für zukünftige Studien verwendet werden [23]. Der Datensatz beinhaltet Aufnahmen von 35 Kindern mit ASS.

Bei der Erstellung und Publikation eines Datensatzes ist es empfehlenswert, den FAIR-Richtlinien zu folgen [35]. FAIR ist ein Akronym, das für folgende Eigenschaften eines Datensatzes steht: Auffindbar – zugänglich – kompatibel – wiederverwendbar. Um kompatibel zu sein, sollten die Daten entsprechend breit zugänglich aufbereitet sein, die Metadaten sollten verständlich und korrekt beschrieben sein, und Referenzen zu anderen in Verbindung stehenden Quellen sollten angegeben sein. Um wiederverwendbar zu sein, sollten die Metadaten ausführlich mit einer Vielfalt von zutreffenden und relevanten Eigenschaften beschrieben werden, mit einer klaren und zugänglichen Datenverwendungslizenz zur Verfügung gestellt werden, und den allgemeinen Standards der jeweiligen Zielgruppe entsprechen. Um auffindbar zu sein, sollte der Datensatz mit einer weltweit eindeutig identifizierbaren und einheitlichen Kennzeichnung versehen werden und in einer Datenbank registriert oder indexiert sein.

Um zugänglich zu sein, sollte der Datensatz mithilfe eines standardisierten Protokolls abrufbar sein (das Protokoll soll frei zugänglich und universell implementierbar sein).

Wir möchten betonen, dass das Teilen von Daten es erfordert, eine entsprechende schriftliche Einverständniserklärung der Studienteilnehmer*innen bzw. deren Erziehungsberechtigten einzuholen. Das Recht auf Privatsphäre der Studienteilnehmer*innen ist zu beachten. Auch das Votum einer Ethikkommission sollte eingeholt werden (siehe GEN₃).

Leitlinie RES₃:

Berichten Sie die Charakteristika der Kinder und der Studienkohorten so detailliert wie möglich.

Da die mit der Autismus-Spektrum-Störung einhergehenden Defizite sich individuell sehr verschieden manifestieren, ist die Generalisierbarkeit der Ergebnisse eine Herausforderung, die alle Studien betrifft. Daher ist es für jede Studie – egal wie die Gruppen zusammengesetzt sind – wichtig, die Informationen über die Studienteilnehmer*innen so detailliert wie möglich anzugeben. Folgende Informationen sind essenziell: chronologisches Alter und Entwicklungsalter, Geschlecht, sowie Informationen über Fähigkeiten, Komorbiditäten und Therapiedauer.

Leitlinie RES₄:

Berichten Sie Gewöhnungseinheiten und Einheiten, die fehlgeschlagen sind.

Seien Sie darauf vorbereitet, dass Daten für manche Kinder und/oder Therapieeinheiten nicht verfügbar sind. Dafür kann es viele Gründe geben – diese können technische Probleme sein (z.B.: Verbindungsschwierigkeiten des Sensors, Fehlfunktion eines Geräts, niedriger Ladezustand), Bedienungsfehler (z.B.: Gerät wurde nicht eingeschaltet, Kalibrierungsfehler), spezielles kindliches Verhalten während der Einheit. Ein Kind könnte sich zum Beispiel weigern ein Armband zu tragen oder mit dem Roboter zu interagieren, es könnte besonders ängstlich sein, oder es könnte einfach eine Position einnehmen, die die Datenerhebung unmöglich macht (zum Beispiel zur Seite schauen).

Wann immer eine Interaktion fehlschlägt sollte es zusammen mit dem jeweiligen Grund berichtet werden. Dadurch wird die Studie transparent, glaubwürdiger und wertvoller für zukünftige Forschung. Zusätzlich ist es sehr hilfreich, wenn Forscher*innen beschreiben, was nicht funktioniert hat, da es erlaubt Irrwege und zwecklose Geräte und Methoden sowie Herausforderungen, die in ähnlichen Studien auftreten könnten, zu identifizieren.

Leitlinie RES5:

Wählen Sie geeignete Stimuli, wenn Sie in Ihrer Studie Emotionen evozieren möchten.

Einige Studien könnten es erfordern, Emotionszustände hervorzurufen, anstatt natürliche Emotionen zu beobachten. Studien verwendeten viele Strategien, um Emotionen hervorzurufen und zu markieren. Hinsichtlich der Stimuli wurden sowohl evozierte Emotionen als auch ein natürlicher Interaktionsansatz verwendet und die Herausforderung wurde thematisiert, einen entsprechenden Trainingsdatensatz von Kindern mit ASS zu erheben [24]. Die in den Studien am häufigsten verwendeten Stimuli waren Bilder und Videos. Allerdings zeigte eine Studie [17] zu galvanischer Hautreaktion, dass Bilder keine geeigneten Stimuli sind, um Emotionen bei Kindern mit ASS zu evozieren. Eine andere Studie analysierte Augenbewegungsmuster als Reaktion auf Videostimuli, die menschliche Gesichter enthielten. Obwohl Videostimuli am häufigsten verwendet wurden, setzten andere Studien folgende Stimuli ein: einen spielbasierten Ansatz [25], computer-basierte Interventionstools [26], oder Beobachtung einer natürlichen Mensch-Roboter-Interaktion [20] [27].

4.8. Wissenschaftliche Artikel über Kinder mit ASS (REP)

Im Zuge unserer Literaturrecherchen bemerkten wir, dass einige Begriffe und Phrasen nicht einheitlich verwendet wurden. Dieser Abschnitt befasst sich mit diesen Begrifflichkeiten.

Leitlinie REP1:

Unterscheiden Sie zwischen den zwei möglichen Bedeutungen von „Emotionserkennung“.

Im Zusammenhang mit Emotionserkennungstechnologien, die eingesetzt werden, um die Therapie von Kindern mit ASS zu unterstützen, hat der Begriff „Emotionserkennung“ zwei mögliche Bedeutungen: Kinder mit ASS erkennen Emotionen und die Emotionen der Kinder mit ASS werden erkannt. Die erste Bedeutung bezieht sich auf die Fähigkeit der Kinder, die Emotionen von anderen zu erkennen. Die zweite Bedeutung bezieht sich auf Emotionserkennungstechnologien, die eingesetzt werden, um die inneren Emotionszustände eines Kindes zu erkennen und zu analysieren.

Leitlinie REP2:

Seien Sie genau bei der Beschreibung von Geräten, Kanälen und Modalitäten und

Seien Sie sich der Unterschiede zwischen diesen Begriffen bewusst.

Wenn man von Inputs für Emotionserkennung spricht, ist es wichtig, zwischen Aktivitäten/Handlungen, Beobachtungskanälen und Modalitäten zu unterscheiden.

Emotionserkennung erfolgt basierend auf bewussten und unbewussten **Aktivitäten/Handlungen** eines Menschen, die emotionale Ausdrücke widerspiegeln, die mit Emotionserkennung weiter analysiert werden können. Die folgenden Aktivitäten/Handlungen wurden in den ausgewählten Artikeln analysiert: verschiedene Arten von Bewegungen, von Menschen produzierte Geräusche, physiologische Parameter wie Herzfrequenz, unbewusste Muskelaktivität, Atmung und Thermoregulation. Die Aktivierung des menschlichen Nervensystems induziert Veränderungen dieser Aktivitäten/Handlungen, die wiederum als Ausdrücke von Emotionen interpretiert werden können.

Die Handlungen/Aktivitäten können mittels **Beobachtungskanälen** aufgezeichnet werden; es werden Signale aufgezeichnet, die Informationen über beobachtbare emotionale Ausdrücke beinhalten. Der Kanal bezieht sich eher auf einen Signaltyp als auf ein physisches Medium. Folgende Kanäle wurden in Studien zu Emotionserkennung bei Kindern mit ASS verwendet: RGB Video, Tiefenbildkamera (hauptsächlich Kinect), Audio, EKG (Elektrokardiographie), BVP (Blutvolumenpuls), Größe des Brustkorbs, EMG (Elektromyographie), fMRT (funktionelle Magnetresonanztomographie), EDA (elektrodermale Aktivität), und Temperatur.

Die Aktivitäten/Handlungen generieren **Modalitäten**, die als beobachtbare Information verstanden und zur Emotionserkennung herangezogen werden können. In unserer Studie werden die Modalitäten anhand der Aktivitäten/Handlungen gruppiert:

- Bewegung: Gesichtsausdrücke, Körperhaltungen, Augenbewegungen, Kopfbewegungen, Gesten (auch Handbewegungen genannt) und andere noch nicht klassifizierte Bewegungen;
- Geräusche: Lautäußerungen, Prosodie der Sprache;
- Herzaktivität: Herzfrequenz, HRV (Herzratenvariabilität);
- Muskuläre Aktivität, die nicht mit Bewegung in Verbindung steht: Muskelspannung;
- Schwitzen: elektrodermale Aktivität;
- Atmung: Intensität und Frequenz;
- Thermoregulation: periphere Temperatur;
- Gehirnaktivität: neuronale Aktivität. [2]

Wir empfehlen die Beobachtungskanäle zu definieren und detaillierte Informationen über die Geräte anzugeben, die zur Datenerhebung verwendet werden sollen. Außerdem sollen die Modalitäten berichtet werden, die mittels der Beobachtungskanäle erfasst werden sollen.

Leitlinie REP₃:

Wählen Sie die Begriffe, mit denen Sie sich auf die Kinder beziehen, mit Bedacht aus.

Inklusive Sprache berücksichtigt Diversität und ist allen Menschen gegenüber respektvoll. Seien Sie bei der Wahl der Begriffe für die Kinder vorsichtig, sodass Sie nicht implizit ein Individuum einem anderen gegenüber aufgrund des Gesundheitszustands oder des Störungsbildes überlegen darstellen. Im Zuge unserer Literaturrecherche bemerkten wir, dass einige Studien Kinder mit ASS als „autistische Kinder“ [engl: „autistic children“] bezeichneten, während die Kinder der Kontrollgruppe als „normale Kinder“ [engl. „normal children“] bezeichnet wurden, was zu vermeiden ist [1, 2]. Wir empfehlen stattdessen die Verwendung von angemesseneren Bezeichnungen wie „sich typisch entwickelnde Kinder“ [eng.: „typically developing children“] oder „neurotypisch“ [engl. „neurotypical“]. Bezieht man sich auf Kinder mit ASS, wird vielfach empfohlen, zuerst die Person zu nennen, z.B.: „ein Kind mit Autismus“ [engl. „a child with autism“] oder „Personen mit der Autismus-Spektrum-Störung“ [engl. „individuals on the autism spectrum“].

Leitlinie REP₄:

Benennen und definieren Sie die zu untersuchenden Emotionszustände.

Da von der Psychologie Emotionen nicht einheitlich definiert werden, verwenden Forscher*innen, Lehrer*innen und Trainer*innen verschiedene Begriffe, um sie zu benennen. Einige Autor*innen beziehen sich bei Emotionszuständen auf die Basisemotionen, während andere ihre eigenen Bezeichnungen verwenden, und diese Bezeichnungen werden von einer psychologischen Perspektive aus nicht genauer definiert. Einige verwendete Begriffe könnten gruppiert werden. Beispielsweise wird in den meisten Studien, die die Begriffe glücklich, Glück, Freude und Lächeln verwendet, der gleiche Emotionszustand gemeint. Diese Leitlinie besagt, dass Emotionszustände genau benannt und definiert werden sollen. Wenn man beispielsweise ein Lächeln erkennt, sollte man es als Lächeln bezeichnen und nicht als Glück oder Freude. Ein anderes Beispiel sind Emotionen, die mit Angst zu tun haben wie Angst, Sorge oder Beklemmung. Keiner dieser Begriffe wird in den Studien genau definiert.

Einige Zustände, die in den Studien [36-38] adressiert werden, haben mehr mit Aufmerksamkeit zu tun als mit Emotionen (Beschäftigung [engl. „engagement“], Beteiligung [engl. „involvement“]) – dies sind Aspekte, die für Autismusstudien sehr interessant sein können. Auch das Konzept von Mischemotionen wie ängstlich-überrascht oder freudig-angeekelt wurde in Bezug auf Kinder mit ASS diskutiert.

Da Psycholog*innen bestimmte diskrete Emotionen unterschiedlich definieren, empfehlen wir, eine dieser Definitionen für die jeweilige Studie auszuwählen und diese auch zu berichten.

5. Evaluation der Leitlinien

Es wurden mehrere Methoden angewandt, um diese Leitlinien – die ER-RIA Leitlinien für Emotionserkennung in roboter-unterstützten Interventionen bei Autismus – zu evaluieren. Diese Methoden sind:

- Fragebogen, um quantitative Daten zu erheben;
- Fokusgruppen, um qualitative Daten zu erheben;
- Expert*innenevaluation mit Hilfe des AGREE-Instruments – sowohl qualitativ als auch quantitativ.

5.1. Fragebogen

Die erste Version der Leitlinien wurde evaluiert und dann verbessert (Liste mit den Änderungen – siehe Abschnitt 5.4). Im Fragebogen wurde jede Leitlinie extra betrachtet. Jede Leitlinie wurde hinsichtlich folgender Kriterien evaluiert:

- Umfang der Beschreibung mit einer symmetrischen 5-Punkte-Skala, die von zu wenig umfangreich bis zu umfangreich reicht (3 entspricht der bestmöglichen Bewertung);
- Verständlichkeit der Leitlinie und ihrer Beschreibung – mit einer 5-Punkte Skala, die von Zustimmung bis Ablehnung reicht (5 entspricht der bestmöglichen Bewertung);
- Anwendbarkeit der Leitlinie - mit einer 5-Punkte Skala, die von Zustimmung bis Ablehnung reicht (5 entspricht der bestmöglichen Bewertung).

Der Fragebogen wurde an 49 Teilnehmer*innen ausgehändigt, die gebeten wurden, alle Leitlinien nacheinander durchzulesen und die drei Fragen für jede Leitlinie zu beantworten. Die Teilnehmer*innen waren Studierende, die an einem Training über die affektive Schleife in der Roboter-Kind-Interaktion im Rahmen von Autismustherapie teilnahmen.

Dies ist eine Zusammenfassung der Ergebnisse der Fragebogenuntersuchung:

- Umfang der Leitlinien: der durchschnittliche Wert betrug 3,25 +- 0,68 (1 bedeutet zu wenig umfangreich, 3 bedeutet neutral, 5 bedeutet zu umfangreich);
- Zu wenig umfangreiche Beschreibung (mit ≥ 10 Personen, die 1 oder 2 angaben) betraf die folgenden Leitlinien: GEN₃, CH₁, INT₂, EMO₄, RES₃;
- Zu umfangreiche Beschreibung (durchschnittlicher Wert $> 3,5$) betraf die folgenden Leitlinien: GEN₂, CH₃, SYM₁, SYM₁₀, TECH₅, INT₁, PROC₂, PROC₆;
- Die durchschnittliche Bewertung für die Verständlichkeit war 4,57 +- 0,69 (5 – sehr zutreffend);
- Wenig verständlich ($< 4,5$): GEN₁, GEN₂, TECH₃, INT₁, PROC₂, PROC₅, PROC₆, PROC₇, EMO₂, RES₃, RES₅, REP₂;
- Nur eine Leitlinie wurde hinsichtlich ihrer Verständlichkeit durchschnittlich unter 4 bewertet: PROC₂ (3,97+-1,05);
- Die durchschnittliche Bewertung für die Anwendbarkeit war 3,99 +-0,98 (5-Punkte-Likert-Skala, 5 – sehr zutreffend, 3 – neutral)
- Vergleichsweise geringe Anwendbarkeit (durchschnittlich unter 4) wurde für folgende Leitlinien angegeben: GEN₁, GEN₃, CH₁, CH₄, TECH₄, TECH₅, INT₁, PROC₁, PROC₂,

PROC₃, PROC₄, PROC₅, PROC₆, PROC₇, EMO₁, EMO₂, EMO₃, EMO₄, EMO₅, RES₂, RES₅, REP₄;

- Sehr geringe Anwendbarkeit (unter 3,5): PROC₂, PROC₆, EMO₂ – diese Leitlinien wurden auch als wenig verständlich bewertet.

Wir konzentrierten uns auf jene Leitlinien, die im Fragebogen signifikant schlechter bewertet wurden und verbesserten diese in Version 1.1 der Leitlinien.

5.2. Fokusgruppe

Nachdem sie sich mit den Leitlinien und dem Fragebogen vertraut gemacht hatten, wurden die Teilnehmer*innen in Fokusgruppen eingeteilt. Jede Gruppe bestand aus 6 bis 8 Teilnehmer*innen; insgesamt gab es 7 Gruppen. Jede der Gruppen wurde darum gebeten, die folgenden Fragen zu beantworten:

- Wählen Sie die 5 Leitlinien aus, die für Sie am wenigsten verständlich sind (mit Begründung)
- Wählen Sie die 5 Leitlinien aus, die Ihrer Meinung nach am schwierigsten anzuwenden sind (mit Begründung)
- Geben Sie 3 Leitlinien an, die Ihrer Meinung nach am wertvollsten für Therapeut*innen sind
- Geben Sie 3 Leitlinien an, die Ihrer Meinung nach am wertvollsten für Produktentwickler*innen sind
- Geben Sie 3 Leitlinien an, die Ihrer Meinung nach am wertvollsten für Forscher*innen sind
- Gibt es Ihrer Meinung nach Leitlinien, die nicht notwendig sind und daher weggelassen werden könnten?
- Gibt es Ihrer Meinung nach Leitlinien, die noch ergänzt werden könnten?

Dies ist die Zusammenfassung der Ergebnisse der Fokusgruppen (die Zahlen in Klammer geben an, von wie vielen Fokusgruppen dieser Aspekt angegeben wurde):

- Viele Leitlinien (2), Seitenzahlen ergänzen (1);
- Wiederholungen (3), Widersprüche (3)
- Die am wenigsten verständlichen Leitlinien: PROC₂ (4 mal genannt), PROC₅ (3 mal genannt), GEN₃ (3 mal genannt), je einmal genannt: RES₃, SYM₁₀, TECH₃, PROC₇, SYM₁, INT₁, REP₂, INT₅
- Die am schwierigsten anzuwendenden Leitlinien: alle EMO Leitlinien (3), PROC₂ (2), PROC₅ (2), RES₂ (2), REP₄ (2), je einmal genannt: CH₄, PROC₃, PROC₆, PROC₇, TECH₄, SYM₃, SYM₆, SYM₉, REP₃, RES₃;
- Keine der Gruppen schlug vor Leitlinien wegzulassen; es wurde aber vorgeschlagen einige zusammenzufügen: CH₁ + CH₂ + CH₃ + SYM₂, SYM₄ + SYM₅, SYM₇ + SYM₈, merge SYM₆ + SYM₇ + SYM₈, INT₂ + TECH₅, GEN₃ + TECH₁.

Die Fokusgruppen lieferten einige nützliche qualitative Informationen – sie deckten nicht nur auf wo Änderungsbedarf besteht, sondern machten auch Vorschläge, wie die Beschreibungen verbessert werden könnten. Wir verbesserten die Leitlinien in Version 1.1 basierend auf den Anmerkungen der Fokusgruppen.

5.3. Expert*innenevaluation mittels AGREE

Anschließend wurden die Leitlinien von 3 Expert*innen evaluiert, die dafür AGREE (The Appraisal of Guidelines for Research and Evaluation) verwendeten. Dies ist ein Instrument, mit dem praxisrelevante Leitlinien und ihre Beschreibungen evaluiert werden können. Hier wurde die AGREE II Version verwendet – sie besteht aus 23 Kriterien, die in 6 qualitative Bereiche eingeteilt werden; zusätzlich gibt es noch 2 allgemeine Kriterien [39]. Die Bereiche sind: Umfang und Zweck, Einbeziehung der Zielgruppe, Entwicklungspräzision, Klarheit der Darstellung, Anwendbarkeit, und redaktionelle Unabhängigkeit. Dies ist die Zusammenfassung der Ergebnisse der Expert*innenevaluation:

- Das allgemeine Kriterium „Qualität dieser Leitlinie“ schnitt mit 6,12 ab (es wurde eine 7-Punkte-Likert-Skala verwendet);
- Das allgemeine Kriterium „Die Anwendung dieser Leitlinie wird empfohlen“ schnitt mit 2,9 ab (es wurde eine 3-Punkte-Skala verwendet);
- Von 23 Kriterien wurden 20 mit mehr als 6 Punkten bewertet (7-Punkte-Skala);
- Die folgenden Kriterien wurden mit weniger als 6 Punkten bewertet:
 - Eine Methode die Leitlinie zu aktualisieren (4,95),
 - Die Leitlinie beschreibt Anwendbarkeit (5,73),
 - Die Leitlinie gibt Evaluationskriterien an (5,02);
- Qualitative Anmerkungen waren unter anderem folgende:
 - In der Danksagung sollte eine Erklärung ergänzt werden, die besagt, dass die Förderorganisation keinen Einfluss auf den Inhalt der Leitlinien hatte;
 - Es sollten Referenzen ergänzt werden, um die Leitlinien fundierter aufzustellen;
 - Die RES Leitlinien sind sehr allgemein und beziehen sich nicht nur auf roboterunterstützte Autismustherapiestudien;
 - Die EMO Leitlinien sind naturgemäß eher deskriptiv und beinhalten daher keine Handlungsempfehlungen.

In Version 1.2 verbesserten wir die Leitlinien basierend auf den Expert*innenevaluationen mittels AGREE.

5.4. Änderungen – Leitlinien 1.0 und 1.2

Wir verbesserten die Leitlinien basierend auf den Ergebnissen des oben beschriebenen Evaluationsprozesses. Die Änderungen betreffen hauptsächlich folgende Aspekte:

- Zusammenfassung von Leitlinien (49 Leitlinien in Version 1.0, 46 in Version 1.1 and 44 in Version 1.2);

- Vermeidung von Wiederholungen und Widersprüchen zwischen den Leitlinien; Hinzufügen von Seitenzahlen;
- Ergänzung einer Erklärung hinsichtlich der Unabhängigkeit des Inhalts der Leitlinien vom Fördergeber;
- Es wurden Informationen über die Entwicklung der Leitlinien eingefügt (Abschnitt 3);
- Es wurde ein Abschnitt über Evaluation und Kontrolle der Leitlinien eingefügt (Abschnitt 5);
- Ergänzung von Informationen betreffend Anwendbarkeit der Leitlinien (Abschnitt 6);
- Verbesserung der Beschreibungen von beinahe allen Leitlinien (besonderes Augenmerk wurde auf jene Leitlinien gelegt, die von den Fokusgruppen angegeben wurden und auf jene, die in der Fragebogenevaluation und/oder der Evaluation mittels AGREE-Instrument weniger gut abschnitten).

6. Anwendbarkeit

Die Leitlinien wurden für drei Zielgruppen entwickelt (siehe Abschnitt 2.2): Therapeut*innen von Kindern mit ASS, Produktentwickler*innen und Forscher*innen. Es ist uns bewusst, dass diese verschiedenen Zielgruppen unterschiedliche Anwendungsinteressen haben – Leitlinien, die für eine Zielgruppe wertvoll sind, können für eine andere gegebenenfalls nicht relevant sein. Daher baten wir Fokusgruppen jene Leitlinien anzugeben, die für die jeweilige Subzielgruppe wertvoll sind.

Folgende Leitlinien sind für **Therapeut*innen** am wertvollsten:

- GEN₂ und GEN₃ am Beginn der Leitlinien;
- Alle SYM-Leitlinien, um zu verstehen, wie Ausdrücke von Emotionen erfasst werden;
- Alle INT-Leitlinien, die beschreiben, wie die Planung und Durchführung einer Interaktion erfolgen kann;
- Alle TECH-Leitlinien, sofern technischer Support fehlt.

Folgende Leitlinien sind für **Produktentwickler*innen** am wertvollsten:

- GEN₁ um sich dessen bewusst zu werden, dass der Fokus auf die Bedürfnisse des Kindes gelegt werden sollte und GEN₃ um ethische Anforderungen wie Privatsphäre zu gewährleisten;
- Alle PROC-Leitlinien – wie erfolgt die Datenverarbeitung;
- Alle EMO-Leitlinien, die die Besonderheiten der emotionalen Ausdrücke von Kindern mit ASS beschreiben;
- TECH₂ bis TECH₅ – technische Anforderungen an die zu entwickelnden Produkte;
- CH₁ und CH₂, INT₂.

Folgende Leitlinien sind für **Forscher*innen** am wertvollsten:

- Alle RES-Leitlinien – diese konzentrieren sich auf Studien zur Interaktion von Robotern und Kindern mit ASS;

- Alle REP-Leitlinien, die sich damit beschäftigen, wie Studien in wissenschaftlichen Artikeln beschrieben werden sollten;
- Ausgewählte CH- und EMO-Leitlinien – um die Einschränkungen zu verstehen, die die Arbeit mit Kindern mit ASS mit sich bringt und um die Datenqualität zu evaluieren;
- PROC₅ und PROC₇.

Wir sind uns dessen bewusst, dass einige der Leitlinien nicht so einfach anzuwenden sind und das liegt daran, dass dieses Gebiet einige Herausforderungen mit sich bringt, die sich kaum ändern lassen. In der Fragebogenevaluation wurden die GEN-, PROC-, EMO- und RES-Leitlinien sowie ausgewählte TECH- und CH-Leitlinien hinsichtlich der Anwendbarkeitskriterien weniger gut bewertet.

Der GEN-Abschnitt beinhaltet zum Einstieg eher allgemeine Empfehlungen. Die folgenden Herausforderungen erschweren die Umsetzung einiger der Leitlinien:

- Es ist schwierig Emotionen genau zu beschreiben und zu erkennen und die Herausforderung ist noch größer, wenn automatische Emotionserkennungstechnologien für die Klassifikation der Emotionen verwendet werden sollen;
- Die Arbeit mit Kindern mit ASS ist herausfordernd – dies liegt an ihren Schwierigkeiten, aber auch an ihrem Verhalten in stressigen Situationen und ihren individuellen Vorlieben;
- Kinder mit ASS drücken Emotionen anders aus als sich typisch entwickelnde Kinder im gleichen Alter;
- Therapie für Kinder mit ASS sowie entsprechende Studien müssen sich am Kind orientieren, was gut für das Kind ist, aber schwierig für die Forschung, da Daten gegebenenfalls nicht verfügbar oder von geringer Qualität sein können und/oder die Wiederholbarkeit der Studie eingeschränkt ist.

Aufgrund der oben erwähnten Herausforderungen können einige Anwendbarkeitsprobleme nicht eliminiert werden; dazu zählen die Grundwahrheit des kindlichen Emotionszustands zu erfahren, einen Raum ohne Hintergrund- bzw. Störgeräusche zu finden, oder einen frei verfügbaren Datensatz zu generieren.

7. Zukünftige Arbeiten

Dieses Dokument wurde im Rahmen des internationalen EMBOA-Projekts entwickelt, das vom Erasmus Plus Programm der Europäischen Union gefördert wurde. Das Dokument wird kostenlos unter der Creative Commons-Lizenz CC BY vertrieben. Es ist auf der Projekt-Homepage <http://emboa.eu/> auf Englisch, Polnisch, Mazedonisch, Deutsch und Türkisch verfügbar. Das Dokument darf frei distribuiert werden.

Auch wenn das Projekt im Jahr 2022 endet, planen wir weitere Forschungsarbeiten zu diesem Thema durchzuführen und möglicherweise auch eine Erweiterung dieser Leitlinien zu erstellen. Einige Ideen für die Erweiterung der Leitlinien wurden im Zuge des

Evaluationsprozesses oder im Rahmen von Beobachtungen des Projektteams generiert, wurden aber bislang noch nicht aufgegriffen, da sie außerhalb der EMBOA-Projektziele lagen. Dazu zählen:

- Entwicklung von Leitlinien, die in Richtung zukünftiger automatischer Interaktion gehen sollen mit dem Ziel die affektive Schleife zu schließen, die in den Leitlinien erwähnt wird;
- Definition/Entwicklung von zukünftigen Technologien, anstatt Leitlinien zu erstellen, wie mit den vorhandenen Technologien gearbeitet werden sollte;
- Erstellung von Empfehlungen für die Zusammenarbeit zwischen den Zielgruppen – Forscher*innen, Therapeut*innen und Produktentwickler*innen;
- Einteilung der Leitlinien in folgende Kategorien: „Planungsphase“, „vor der Einheit mit dem Kind“, „während der Einheit mit dem Kind“, „Nachbearbeitungsphase“.

Wenn Sie mit den Leitlinien arbeiten möchten, wenn Sie Verbesserungs- oder Erweiterungsvorschläge haben oder wenn Sie die Leitlinien in ihre eigene Muttersprache übersetzen möchten, kontaktieren Sie bitte sehr gerne Agnieszka Landowska, nailie@pg.edu.pl.

Referenzen

1. Bartl-Pokorny, K. D., Uluer, P., Barkana, D. E., Baird, A., Kose, H., Zorcec, T., Robins, B., Schuller, B., Landowska, A., & Pykała, M. (2021). Robot-Based Intervention for Children With Autism Spectrum Disorder: A Systematic Literature Review. *IEEE Access*, 9, 165433-165450. <https://doi.org/10.1109/access.2021.3132785>
2. Landowska, A.; Karpus, A.; Zawadzka, T.; Robins, B.; Erol Barkana, D.; Kose, H.; Zorcec, T.; Cummins, N. Automatic Emotion Recognition in Children with Autism: A Systematic Literature Review. *Sensors* 2022, 22, 1649. <https://doi.org/10.3390/s22041649>
3. Karpus A., Landowska A., Miler J., Pykała M.: Systematic Literature Review – methods and hints, ETI Faculty Technical report, Gdansk University of Technology, 1/2020,
4. Bartl-Pokorny K.D., Pykała M., Erol Barkana D., Baird A., Köse H., Zorcec T., Robins B., Schuller B.W., Landowska A. Systematic Literature Review - Robot-Based Intervention for Children with Autism Spectrum Disorder, ETI Faculty Technical report, Gdansk University of Technology, 2/2020
5. Landowska A, Wróbel M., EMBOA project evaluation report, ETI Faculty Technical report, Gdansk University of Technology, X/2022
6. Abdullah, S.M.S.A., Ameen, S.Y.A., Sadeeq, M.A. and Zeebaree, S., 2021. Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, 2(02), pp.52-58.
7. Landowska, A., Zawadzka, T., & Zawadzki, M. (2021). Mining inconsistent emotion recognition results with the multidimensional model. *IEEE Access*.
8. Yang, K., Wang, C., Sarsenbayeva, Z., Tag, B., Dingler, T., Wadley, G., & Goncalves, J. (2021). Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. *The Visual Computer*, 37(6), 1447-1466.
9. Milling, M., Baird, A., Bartl-Pokorny, K. D., Liu, S., Alcorn A. M., Shen, J., Tavassoli, T., Ainger, E., Pellicano E., Pantic, M., Cummins, N., Schuller, B. W. (2022). Evaluating the Impact of Voice Activity Detection on Speech Emotion Recognition for Autistic Children. *Frontiers in Computer Science*, 4, 837269.
10. Milling, M., Bartl-Pokorny, K. D., Schuller, B. W. (2022). Investigating Automatic Speech Emotion Recognition for Children with Autism Spectrum Disorder in Interactive Intervention Sessions with the Social Robot Kaspar. medRxiv. <https://doi.org/10.1101/2022.02.24.22271443>
11. Schuller, B. (2018). What affective computing reveals about autistic Children’s facial expressions of joy or fear. *Computer*, 51(06), 7-8.
12. Landowska A.: (2019) Uncertainty in Emotion recognition, *Journal of Information, Communication and Ethics in Society*, Vol. 17 No. 3, pp. 273-291, Emerald Publishing, DOI 10.1108/JICES-03-2019-0034
13. [Landowska, Miler, 2016] Landowska, A. and Miler, J. (2016), “Limitations of emotion recognition in software user experience evaluation context”, *Federated Conference on Computer Science and Information Systems, (FedCSIS), IEEE*, pp. 1631 - 1640.
14. Belteki Z, Lumbreras R, Fico K, Haman E, Junge C. The Vocabulary of Infants with an Elevated Likelihood and Diagnosis of Autism Spectrum Disorder: A Systematic Review and Meta-Analysis of Infant Language Studies Using the CDI and MSEL. *Int J Environ Res Public Health*. 2022 Jan 27;19(3):1469. doi: 10.3390/ijerph19031469.
15. MacFarlane H, Salem AC, Chen L, Asgari M, Fombonne E. Combining voice and language features improves automated autism detection. *Autism Res*. 2022 Apr 23. doi: 10.1002/aur.2733.

16. Bal, E.; Harden, E.; Lamb, D.; Van Hecke, A.V.; Denver, J.W.; Porges, S.W. Emotion Recognition in Children with Autism Spectrum Disorders: Relations to Eye Gaze and Autonomic State. *J. Autism Dev. Disord.* 2010, 40, 358–370.
17. Fadhil, T.Z.; Mandeel, A.R. Live Monitoring System for Recognizing Varied Emotions of Autistic Children. In *Proceedings of the 2018 International Conference on Advanced Science and Engineering (ICOASE)*, Duhok, Iraq, 9–11 October 2018; pp. 151–155.
18. Marinoiu, E.; Zafir, M.; Olaru, V.; Sminchisescu, C. 3D Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children with Autism. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2158–2167.
19. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Affect Recognition in Robot Assisted Rehabilitation of Children with Autism Spectrum Disorder. In *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, Rome, Italy, 10–14 April 2007; pp. 1755–1760.
20. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Online Affect Detection and Robot Behavior Adaptation for Intervention of Children with Autism. *IEEE Trans. Robot.* 2008, 24, 883–896.
21. Sorensen, T.; Zane, E.; Feng, T.; Narayanan, S.; Grossman, R. Cross-Modal Coordination of Face-Directed Gaze and Emotional Speech Production in School-Aged Children and Adolescents with ASD. *Sci. Rep.* 2019, 9, 18301.
22. Grossard, C.; Dapogny, A.; Cohen, D.; Bernheim, S.; Juillet, E.; Hamel, F.; Hun, S.; Bourgeois, J.; Pellerin, H.; Serret, S.; et al. Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions: An experimental study using random forest classifiers. *Mol. Autism* 2020, 11, 5.
23. Rudovic, O.; Lee, J.; Dai, M.; Schuller, B.; Picard, R.W. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci. Robot.* 2018, 3, eaa06760.
24. Tang, T.Y. Helping Neuro-Typical Individuals to “Read” the Emotion of Children with Autism Spectrum Disorder: An Internet-of-Things Approach. In *Proceedings of the 15th International Conference on Interaction Design and Children, IDC’16*, Manchester, UK, 21–24 June 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 666–671.
25. Di Palma, S.; Tonacci, A.; Narzisi, A.; Domenici, C.; Pioggia, G.; Muratori, F.; Billeci, L. Monitoring of autonomic response to sociocognitive tasks during treatment in children with Autism Spectrum Disorders by wearable technologies: A feasibility study. *Comput. Biol. Med.* 2017, 85, 143–152.
26. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. *Int. J.-Hum.-Comput. Stud.* 2008, 66, 662–677.
27. Silva, V.; Soares, F.; Esteves, J. Mirroring and recognizing emotions through facial expressions for a RoboKind platform. In *Proceedings of the 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)*, Coimbra, Portugal, 16–18 February 2017; pp. 1–4.
28. Adolphs R, Mlodinow L, Barrett LF. What is an emotion? *Curr Biol.* 2019 Oct 21;29(20):R1060-R1064. doi: 10.1016/j.cub.2019.09.008. PMID: 31639344; PMCID: PMC7749626.
29. H. Gunes, B. Schuller, M. Pantic and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2011, pp. 827-834, doi: 10.1109/FG.2011.5771357.
30. Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions?. *Psychological review*, 97(3), 315.

31. Rudovic, O.; Lee, J.; Dai, M.; Schuller, B.; Picard, R.W. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci. Robot.* 2018, 3, eaa06760.
32. Gay, V.; Leijdekkers, P.; Wong, F. Using sensors and facial expression recognition to personalize emotion learning for autistic children. *Stud. Health Technol. Inform.* 2013, 189, 71–76.
33. Hirokawa, M.; Funahashi, A.; Itoh, Y.; Suzuki, K. Design of affective robot-assisted activity for children with autism spectrum disorders. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014*; pp. 365–370.
34. Javed, H.; Jeon, M.; Park, C.H. Adaptive Framework for Emotional Engagement in Child-Robot Interactions for Autism Interventions. In *Proceedings of the 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI, USA, 26–30 June 2018*; pp. 396–400.
35. Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
36. Akinloye, F.O.; Obe, O.; Boyinbode, O. Development of an affective-based e-healthcare system for autistic children. *Sci. Afr.* 2020, 9, e00514
37. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. *Int. J.-Hum.-Comput. Stud.* 2008, 66, 662–677.
38. Krupa, N.; Anantharam, K.; Sanker, M.; Datta, S.; Sagar, J.V. Recognition of emotions in autistic children using physiological signals. *Health Technol.* 2016, 6, 137–147
39. AGREE II. Brouwers MC, Kho ME, Browman GP, Burgers J, Cluzeau F, Feder G, Fervers B, Graham, ID, Grimshaw J, Hanna S, Littlejohns P, Makarski J, Zitzelsberger L on behalf of the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *Can Med Assoc J.* Dec 2010, 182: E839-842; doi:10.1503/cmaj.090449