

# UX-Evaluation in der Erinnerungspflege bei Demenz

Stephan Huber<sup>1</sup>, Alexander Bejan<sup>2</sup>, Beate Radzey<sup>3</sup>, Renate Berner<sup>3</sup>, Patrizia Murko<sup>4</sup>, Jörn Hurtienne<sup>1</sup>

Lehrstuhl für Psychologische Ergonomie, Universität Würzburg<sup>1</sup>

Hochschule Furtwangen<sup>2</sup>

Demenz Support Stuttgart<sup>3</sup>

Fürstlich Fürstenbergisches Altenpflegeheim Hüfingen<sup>4</sup>

Kontakt: [stephan.huber@uni-wuerzburg.de](mailto:stephan.huber@uni-wuerzburg.de)

## Zusammenfassung

Gängige Methoden zur Messung der User Experience während der Nutzung sind aufgrund kognitiver Einschränkungen nicht im Kontext Demenz anwendbar. In diesem Beitrag berichten wir von Feldtests einer Smartwatch-App, die es Evaluatoren erlaubt, beobachtete Emotionen zeitnah zu dokumentieren. Explorative Feldstudien zeigen, dass Proxemo generell für Evaluationen in der Erinnerungspflege geeignet ist und von Evaluatoren gegenüber Notizen auf Papier bevorzugt wird. Die Auswahl der Emotionen und der Einsatz für Gruppensitzungen sind noch optimierbar.

## 1 Einleitung

Zur Messung der User Experience (UX) steht Designern und Forschern eine breite Auswahl an Methoden zur Verfügung. In den meisten UX-Methoden sind die Nutzer aufgefordert, zusätzlich zur primären Aufgabe, ihre Wahrnehmungen während oder nach der Nutzung zu reflektieren und die Nutzungs-Erfahrung zu berichten (z.B. lautes Denken, retrospektives Interview, Fragebogen). Reflexion und Kommunikationsvermögen können aber durch Sprachverlust und den Abbau kognitiver Fähigkeiten stark eingeschränkt sein, weshalb existierende UX-Methoden z.B. bei älteren Menschen mit fortgeschrittener Demenz unbrauchbar sind. Da Nutzer mit Demenz als Zielgruppe der Mensch-Computer-Interaktion zunehmend an Bedeutung gewinnen (Morrissey et al., 2017), werden andere als die bisherigen Methoden zur Evaluation und Gestaltung der UX benötigt. Einen Ansatz dazu bieten Beobachtungsmethoden.

Zur Evaluation des Wohlbefindens von Menschen mit Demenz sind Quality-of-Life Methoden wie z.B. Lawton et al. (1999) sehr verbreitet. Davon inspiriert konzipierten Huber et al. (2017)

eine Smartwatch-App zur zeitnahen Dokumentation von Emotionen durch Beobachter, die Nutzer unauffällig bedienen können und zwischen Interaktionen die Hände frei haben. Dieses Evaluationswerkzeug – *Proxemo* – wurde als Konzept präsentiert und in einem frühen Entwicklungsstadium formativ von Experten evaluiert. Im vorliegenden Beitrag explorieren wir, (1) wie *Proxemo* in realen Evaluationssituationen eingesetzt wird. Außerdem interessierte uns, (2) inwiefern sich die erhobenen Daten als Unterstützung von Videoanalysen eignen, bei denen der Effekt interaktiver Systeme auf Emotionen und den Abruf autobiographischer Erinnerungen im Fokus steht.

## 2 Methode

Das zu evaluierende System war eine interaktive Monitorwand mit den Maßen von ca. 2,5 m x 1,5 m in einem ländlichen Demenzpflegeheim in Süddeutschland. Damit konnten Bewohner (a) in Gruppensitzungen multimediale Präsentationen zu Lebensthemen erleben oder (b) in Einzelsitzungen mit einem Haustier-Avatar ein virtuelles Haus erkunden. Beide Formate wurden von Betreuungskräften begleitet (Bejan et al., 2018). Wir erhoben Daten von drei Evaluationsteams, wovon zwei mit *Proxemo* arbeiteten.

Betreuer und Bewohner saßen auf einem Sofa gegenüber der interaktiven Wand. Eine weitwinklige Kamera und je ein Evaluator befanden sich jeweils im rechten Winkel dazu, sodass Bildschirminhalt und Gesichtsausdrücke erfasst werden konnten. Evaluatoren waren mit einem Klemmbrett für Notizen und einer Smartwatch Samsung Gear S2 ausgerüstet, auf der die *Proxemo*-App lief. Durch Berührung eines Emojis auf der Bedienoberfläche von *Proxemo* (siehe Abbildung 1) wird die assoziierte Emotion mit Zeitstempel dokumentiert.

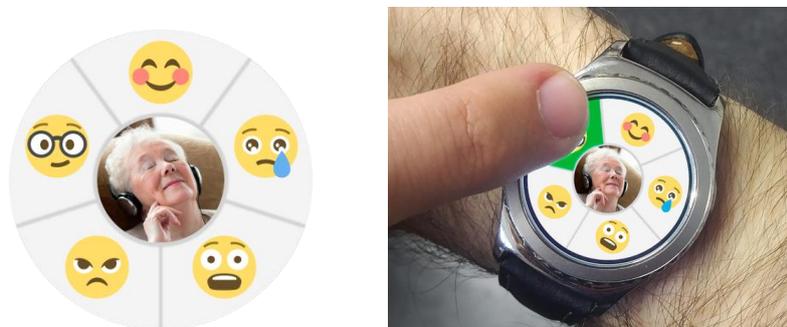


Abbildung 1. Die App *Proxemo* als Screenshot (links) und in Verwendung (rechts). Die Emojis repräsentieren oben beginnend im Uhrzeigersinn die Emotionen Freude, Trauer, Angst, Wut und Aufmerksamkeit (Lawton et al., 1999).

### 2.1 Ablauf in der Erinnerungssitzung

Insgesamt vier unerfahrene Evaluatoren wurden kurz in die Funktionalität und Bedienung der App eingewiesen und gebeten, für eine hohe Datenqualität Emotionen zu dokumentieren, sobald diese erkannt wurden. Dabei wurde ihnen nicht vorgeschrieben, ob und wie die Uhr zu

tragen oder zu positionieren sei. Die insgesamt neun Sitzungen dauerten jeweils rund 30 Minuten, danach wurden die Evaluatoren zu ihren Erfahrungen mit Proxemo während der letzten Sitzung interviewt und füllten die Fragebögen RAW TLX zu kognitiver Beanspruchung (Byers, 1989) und QUESI zu intuitiver Bedienbarkeit aus (Hurtienne & Naumann, 2010).

## 2.2 Ablauf der Auswertung

Wir synchronisierten die in Proxemo gesetzten Zeitstempel mit den Videodateien mithilfe des Video-Annotationstools ELAN (Max Planck Institute for Psycholinguistics, The Language Archive, Nimwegen, Niederlande) und stellten die kombinierte Datei den zwei Evaluations-Teams der multimedialen Präsentation und des virtuellen Hauses zur Verfügung. Sie konnten diese für ihre Videoanalysen verwenden und uns danach Rückmeldung geben.

Ein weiteres Evaluationsteam analysierte Videos weiterer Sitzungen mit der Monitorwand, bei denen Proxemo nicht eingesetzt wurde. Diese Evaluatoren extrahierten ebenfalls im Video zu sehende Emotionen, ohne durch Proxemo auf die Emotionen nach Lawton et al. (1999) beschränkt zu sein, jedoch auch ohne die vor Ort gesetzten Zeitstempel als Hilfestellung.

## 3 Ergebnisse

### 3.1 Anwendung von Proxemo

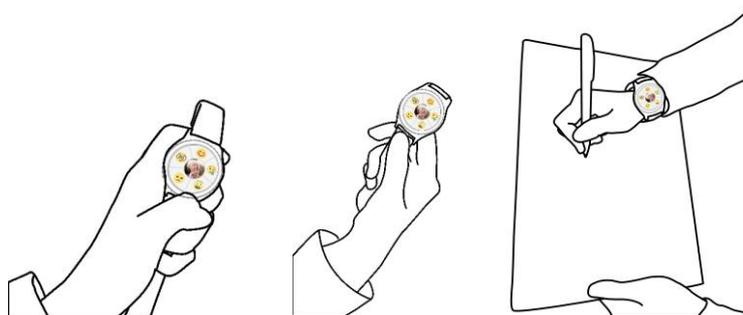


Abbildung 2. Evaluatoren trugen die Uhr am Arm, wenn sie Notizen machten und hielten sie sonst in der Hand.

Nur Evaluatoren, die Notizen machten oder dies vorhatten, befestigten die Smartwatch an ihrem Handgelenk. Evaluatoren, die keine Notizen machten, hielten die Uhr meist in der Hand und bedienten sie mit dem Daumen (siehe Abbildung 2). Die Uhr war dabei so geneigt, dass Evaluatoren das Display sehr gut einsehen konnten, Bewohner jedoch nicht. Manche versteckten die ganze Uhr in der Hand oder hinter überschlagenen Beinen, was aber nicht nötig gewesen wäre, da die Aufmerksamkeit der Bewohner auf der Monitorwand lag. Andere interagierten mit dem Arm, an dem die Uhr befestigt war, sehr natürlich. Beispielsweise gestikulierten sie oder richteten sich die Kleidung. Zur Dokumentation einer beobachteten Emotion wanderte

der Blick der Evaluatoren für etwa 1-2 Sekunden zur Uhr. Dies schätzten Evaluatoren als nicht ablenkend und sogar effizienter gegenüber Papier ein, da das Display der Uhr „alle wichtigen Emotionen enthält“ (E2) und „man einfach nur kurz den richtigen Knopf drücken muss“ (E1). Zeigten Bewohner generell wenige Emotionen, war es „gut möglich, zusätzlich noch Notizen zu machen“ (E1). Andere fanden, dass „die Uhr allein besser [ist], da paralleles Schreiben zur Nutzung der Uhr zu viele Ressourcen benötigt“ (E3) oder Schreiben „in der Situation nicht erforderlich [war]“ (E4). In den Gruppensitzungen fiel zusätzlich auf, dass die Evaluatoren die beobachteten Emotionen aller drei teilnehmenden Bewohner dokumentierten. Eine Evaluatormin, die der Aufforderung einer Bewohnerin gefolgt war und sich neben sie setzte, hatte Schwierigkeiten ihre Aufmerksamkeit von der Monitorwand vor ihr auf die Bewohner neben sich zu lenken. Zusätzlich musste sie sich etwas vorbeugen, um die Gesichtsausdrücke anderer Teilnehmer zu erkennen. Die Gesamtwerte aller Evaluatoren auf der QUESI Skala ( $MW = 4,35$ ,  $SD = 0,34$ ) und dem RAW TLX ( $M = 2,19$ ,  $SD = 1,57$ ) weisen darauf hin, dass die Proxemo-App als intuitiv bedienbar und nicht sehr belastend wahrgenommen wurde.

### 3.2 Verwertbarkeit der erhobenen Daten

Beide Evaluatorenteams haben die mit Proxemo generierten Daten bei der Videoauswertung verwendet. Dabei annotierte je ein Teammitglied, welches die Smartwatch mit Proxemo selbst nicht verwendet hatte, das gesamte Videomaterial (je etwa 2h). Ziel der Auswertung war, qualitativ aussagen zu können, welcher Auslöser welche Emotion hervorrief um bestimmte Inhalte, Interaktionen oder Moderationen zu optimieren.

Ein Großteil der mit Proxemo dokumentierten Emotionen entfiel auf die Kategorien *Freude* und *Aufmerksamkeit*. Die Zuordnung der Emotionen zu im Video erkennbaren Bewohnern war fast immer eindeutig. „Meist gilt es für die Person, die gerade lacht oder mit der die Betreuerin spricht. Generell war auch ohne zusätzliche Notizen anhand der Videos meist klar, was [mit den Zeitstempeln] gemeint war“ (Auswerter A1).

Proxemo-Daten gaben bei der Videoannotation „zusätzliche Sicherheit [und] haben die Auswertung auf jeden Fall erleichtert“ (A2). Besonders relevant wurden Proxemo-Daten, „wenn die Gesichter der Bewohner auf den Videos zeitweise nicht zu sehen [waren], weil sie sich außerhalb des Aufnahmebereichs befanden, sich von der Kamera wendrehen, oder die Betreuerin zwischen Bewohner und Kamera stand. Trotz Ton konnte so anhand der Videodaten teilweise nicht auf Emotionen geschlossen werden, sodass diese ohne Proxemo-Daten unklar gewesen wären“ (A1).

Auch quantitative Aussagen lassen sich mit den Zeitstempeln treffen. So lässt sich anhand der „Tabellen mit Rohdaten [...] ohne das gesamte Transkript lesen oder gar erstellen zu müssen einen schnellen Überblick darüber geben, welche Sessions für welchen Bewohner gut funktioniert haben“ (A1). Auswerter A2 nutzte die Emotionsstempel mit einer Standarddauer von einer Sekunde als Ausgangsbasis und markierte im Video, wie lange die Emotion nach einem Auslöser tatsächlich beobachtbar war. Er fand es „schwierig zu sagen, ich habe 50 Mal *Freude* gehabt [sondern] besser zu sagen, ich hatte 12 Minuten lang *Freude*.“

Beide Evaluatorenteams annotierten die Videos in ihrer Gesamtheit und dokumentierten alle auftretenden Emotionen. Ihnen fiel auf, dass „Desinteresse“ (E3) beziehungsweise „abgelenkt, abwesend sein“ (E2) in Proxemo fehlte, obwohl es häufig zu beobachten war, dass manche Bewohner unaufmerksam waren, „eine Bewohnerin sogar mal 2,5 Minuten vor sich [hindöste]“ (A1). Eine überflüssige Emotion war laut E3 Wut. Zusätzlich sind „Wut und Trauer [...] nicht immer klar unterscheidbar“ (E4). Auf der positiven Seite sah A2 den Bedarf „vielleicht noch etwas zwischen Aufmerksamkeit und Freude“ einzufügen.

Das Evaluationsteam, das Videos von Sitzungen ohne den Einsatz von Proxemo ausgewertet hatte, ging anders vor. Sie codierten nicht jegliche Emotionsregung, sondern nur Emotionen, die im Zusammenhang mit meist autobiographisch geprägten, für den Menschen mit Demenz bedeutsamen Momenten auftraten. Über die in Proxemo vorkommenden Emotionen hinaus annotierten die Auswerter hier noch Stolz und Wehmut, welche beispielsweise auftraten, wenn Menschen über ihre Herkunft oder Heimat sprachen (Stolz) oder sich an einen sehr schönen Lebensabschnitt erinnerten, der aber klar vorbei war (Wehmut).

## 4 Diskussion

In dieser ersten, explorativen Feldstudie erwies sich Proxemo als gut geeignet, beobachtete Emotionen zeitnah und effizient zu dokumentieren. Voraussetzungen für eine hohe Datenqualität ist, wie auch bei anderen Bewertungen durch Stellvertreter, dass die Evaluatoren gut in der Methode geschult sind und sie die Situation mit allen Teilnehmern gut einsehen und einordnen können. Die resultierenden Daten unterstützen eine qualitative Videoanalyse und können vom Video gelöst schnell quantitative Einblicke geben.

Neben diesen Befunden gibt es auch Einschränkungen. Proxemo wurde für die Evaluation von Einzelsitzungen konzipiert, sodass aktuell nicht mehrere Nutzer gleichzeitig unterschieden werden können. Dank dem sehr ruhigen Umfeld der beobachteten Sessions mit maximal drei Bewohnern konnten Evaluatoren während der Videoanalyse die zuvor dokumentierten Emotionen auch nachträglich noch einzelnen Bewohnern zuordnen. Sitzungen mit vielen aktiven Bewohnern können Evaluatoren potentiell überfordern (Perrin, 1997) und sind durch Proxemo derzeit nicht abbildbar.

Drei der vier Proxemo-Nutzer hatten außerhalb des laufenden Projekts noch keine Evaluationserfahrung. Den Novizen gab die Einfachheit des Systems eine Struktur für die Evaluation vor. Eine Studie mit Experten in der Evaluation von Erinnerungspflege im Demenzkontext ist jedoch geplant. Eine weitere methodische Einschränkung ist, dass Proxemo zwar auf den hinreichend untersuchten Emotionen von Lawton et al. (1999) aufbaut, diese aber für Zeiträume von bis zu 5 Minuten und nicht für präzise Zeitstempel ausgelegt sind. Es muss auch noch getestet werden, ob eine reliable Übereinstimmung zwischen Evaluatoren angenommen werden kann.

In einer Iteration des Designs von Proxemo sollte zur oben genannten, getrennten Dokumentation für mehrere Nutzer auch die Auswahl der Emotionen überdacht werden. Die durch Proxemo abbildbaren Emotionen wurden unverändert von Lawton et al. (1999) übernommen.

Möglicherweise ist diese Auswahl zu generisch für die Erinnerungspflege, in der gezielt versucht wird, Emotionen in Situationen mit autobiographischem Bezug hervorzurufen. Rückmeldungen der Evaluatoren vor Ort und nach der Videoanalyse legen nahe, die Emotionen *Wut* und *Angst* zu entfernen oder in einer Kategorie zu vereinen, da diese schwer zu unterscheiden sind und in den beobachteten Sitzungen eine geringe Prävalenz hatten. Stattdessen könnten *Stolz* und *Wehmut* in Erwägung gezogen werden.

Zuletzt dokumentiert Proxemo lediglich die beobachteten Emotionen während der Nutzung und bildet damit nur einen Teilbereich der User Experience ab. Vorfreude oder nachwirkende Emotionen, welche einen wichtigen Aspekt der UX ausmachen, müssen anderweitig erfasst werden.

## Danksagung

Wir danken allen Studenten, Pflege- und Betreuungskräften, die bei den Feldtests mitwirkten, unseren Projektpartnern, sowie dem Bundesministerium für Bildung und Forschung (Projekt InterMem, 16SV7322). Unser Dank gilt weiterhin Emojione (<http://emojione.com>) für die Bereitstellung der Emojis, und Jan Erik Fischer, der die Applikation programmiert hat.

## Literaturverzeichnis

- Bejan, A., Wieland, M., Murko, P., & Kunze, C. (2018). *A Virtual Environment Gesture Interaction System for People with Dementia*. In Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems, Hong Kong, China.
- Byers, J. C. (1989). Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? *Advances in Industrial Ergonomics and Safety I: Taylor and Francis*.
- Huber, S., Preßler, J., Tung, N. L., & Hurtienne, J. (2017). *Evaluating Interaction-Triggered Emotions in People with Dementia*. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO.
- Hurtienne, J., & Naumann, A. (2010). QUESI—A questionnaire for measuring the subjective consequences of intuitive use. *Interdisciplinary College*, 536.
- Lawton, M. P., Van Haitsma, K., & Klapper, J. (1999). Observed affect and quality of life in dementia: Further affirmations and problems. *Journal of Mental Health and Aging*, 5(1), 69-82.
- Morrissey, K., Lazar, A., Boger, J., & Toombs, A. (2017). *HCIxDementia Workshop: The Role of Technology and Design in Dementia*. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO.
- Perrin, T. (1997). The positive response schedule for severe dementia. In *Aging & mental health* (Vol. 1, pp. 185).