

Exposé zur Feldstudie im Forschungsprojekt SPELL

Autor: Christian Elsenbast^{1 2}

Mitwirkende: Dr.-Ing. Melanie Reuter-Oppermann³, Simon Franke⁴, Dr. Eric Rietzke^{5 6}, Rolf van Lengen¹

Zweck: Das Exposé dient der Planung und Kommunikation einer Feldstudie innerhalb des Forschungsprojekts SPELL.

Inhalt

1 Forschungsgegenstand	2
1.1 Projekthintergrund	2
2.2 Leitstellenarbeitsplatz	2
2 Forschungsfragen	6
Abhängige Variable (AV) „Aufmerksamkeit“	Fehler! Textmarke nicht definiert.
Abhängige Variable (AV) „Stress“	6
3 Forschungsstand	7
3.1 Datenlage Notrufabfrage und Einsatzbearbeitung	7
3.2 Eye-Tracking	8
3.3 Hautleitwert	9
3.4 Herzfrequenz und Herzfrequenzvariabilität	9
4 Methodisches Vorgehen	11
4.1 Untersuchungsdesign	11
4.2 Stichprobenziehung	15
4.3 Datenerhebung	15
4.4 Datenaufbereitung und Datenanalyse	21
5.5 Ergebnispräsentation	22
Literaturverzeichnis	23

¹ Fraunhofer IESE, Digital Healthcare

² Universität Bremen, Fachgebiet Wirtschaftspsychologie und Personalwesen

³ TU Darmstadt, Fachgebiet Wirtschaftsinformatik

⁴ DRK-Landesverband Rheinland-Pfalz e.V., Informations- und Technologie-Center

⁵ DFKI – Deutsches Forschungszentrum für Künstliche Intelligenz

⁶ LiveReader GmbH

1 Forschungsgegenstand

1.1 Projekthintergrund

Behörden- und Organisationen mit Sicherheitsaufgaben (BOS), wie Rettungsdienst, Feuerwehr, Polizei, Katastrophenschutz und THW, sind tragende Säulen der staatlichen Daseinsfürsorge. Hilfeersuche, Notrufe und Einsätze werden in Leitstellen organisiert und koordiniert abgearbeitet, die teilweise mehrere BOS räumlich und funktionell zusammenfassen. Leitstellen sind jedoch nicht ausschließlich im Bereich der BOS zu finden. Die Überwachung komplexer Gebäude und Vernetzung mit sicherheitskritischen Datenpunkten (z. B. Brand- und Rauchmelder, Einbruchssicherung) sind hier beispielhaft zu nennen.

Krisensituationen müssen ebenfalls durch die Leitstellen und spezialisierten Lagezentren koordiniert werden. Solche Situationen haben tiefgreifende und nachhaltige Auswirkungen auf viele Wirtschafts- und Lebensbereiche und erfordern eine situationsspezifische, intersektorale und interorganisationale Zusammenarbeit von unterschiedlichsten Beteiligten innerhalb und außerhalb der BOS.

Eine technische Herausforderung ist, dass nicht nur die Daten innerhalb des eigenen Operabilitätsbereichs mit den Einsatzmitteln (z. B. Rettungswagen) ausgetauscht werden können, sondern dass verschiedene Daten mit Daten von Nachbarleitstellen ausgetauscht und abgeglichen werden müssen. Leitstellen arbeiten heute jedoch noch überwiegend isoliert im eigenen System und haben wenig Zugriff auf geteilte Daten (Bergmann, 2021). Dieses Problem wird im BMWI-geförderten Projekt SPELL aufgegriffen. Das Akronym SPELL steht für „Semantische Plattform zur intelligenten Entscheidungs- und Einsatzunterstützung in Leitstellen und beim Lagemanagement“. Mit Lösungen, die auf künstlicher Intelligenz (KI) basieren, soll eine umfassende Daten- und Wissensgrundlage zur Effizienzsteigerung von Leitstellen geschaffen werden. Mit prognostischen Werkzeugen, mathematischen Optimierungsverfahren und Simulationen, kognitiver Modellierung sowie Verfahren des maschinellen Lernens¹ (ML) lassen sich die Auslastung bestimmter Ressourcen auch unter Berücksichtigung komplexer Einflussfaktoren voraussagen und präventive Maßnahmen zuverlässig antizipierend im Voraus abschätzen. Dies soll zu einem übersichtlicheren Lagebild führen, auf dessen Basis schnellere und dem Einsatzziel folgend passendere Entscheidungen getroffen werden können. Bei den Entwicklungen und Implementierungen gilt es zu beachten, dass eine anthropozentrische Perspektive beibehalten wird. Vorweggreifend sei erwähnt, dass zu den Leitstellendisponent:innen kaum systematisierbare Erkenntnisse hinsichtlich soziodemografischer oder spezieller Merkmale, z. B. Technikbereitschaft (TB), bestehen.⁷

2.2 Leitstellenarbeitsplatz

Menschliche Fehler, wie zum Beispiel schwere Behandlungs- oder Diagnosefehler der Leitstellendisponent:innen (Bischof, 2018), sind ebenso ein relevanter Bestandteil der Betrachtung,

⁷ Dieses Kapitel deckt sich mit Kapitel 1 des Exposé zum Promotionsvorhaben des Autors

wenn es um dieses komplexe Arbeitsfeld „Leitstelle“ geht. Diese Komplexität entsteht durch unübersichtliche und eigendynamische Situationen, die unter Zeitdruck in einer technisch anspruchsvollen Umgebung bewältigt werden müssen. Diese Situationen reflektieren komplexe Realitäten, wie sie z. B. von Dörner (2008, S. 285) beschrieben werden. Die Situationen sind meist für die Leitstellendisponent:innen intransparent, da nicht alle benötigten Informationen zur Verfügung stehen und unklar ist, ob bestimmte Bedingungen für das Handeln gegeben sind oder nicht (Dörner, 2008, S. 285). In Entscheidungsprozessen müssen sowohl von der Leitstelle, also auch von den Einsatzkräften vor Ort, eine Vielzahl verschiedener Informationen berücksichtigt werden (Elsenbast, 2021), die unter Einwirkung verschiedener Stressoren bewertet werden müssen. Die Entscheidungsfindung ist also kritisch und kann weitreichende und irreversible Folgen für Menschen, Umwelt und Sachwerte haben. Dynamische Lagen zeichnen sich dadurch aus, dass viele Informationen auf die Handelnden einströmen, die verarbeitet, bewertet und verstanden werden müssen, wobei oftmals Informationen fehlen und sich die Lage dynamisch verändert (Hofinger, 2016, S. 104). Es müssen von den Leitstellendisponent:innen und Einsatzkräften unsichere Entscheidungen bei begrenzten zeitlichen, materiellen und kognitiven Ressourcen getroffen werden (Hofinger, 2003, S. 13). Wie Hofinger ausführt, kommt es dabei leicht zu Überforderungen und Menschen antworten zur Vereinfachung und zum Selbstschutz mit einer Vielzahl von Strategien, die das eigentliche inhaltliche Handlungsziel implizit unterpriorisieren. Hofinger schreibt dazu (2016, S. 104–105):

„Einsätze im Rettungsdienst können vorausgeplant werden. Die nötigen Entscheidungen werden als Handlungsanweisungen festgehalten. Je nach Anwendung können dies Standardprozeduren sein [...], Algorithmen oder Leitlinien sowie Handbücher etc. All diesen Formen ist gemein, dass sie die Personen, [...], durch inhaltliche Vorgaben beim (Aus-)Suchen möglicher Herangehensweisen entlasten.“

Dieser Aussage lässt sich entgegenhalten, dass es kaum rettungsdienstliche Einsätze gibt, die tatsächlich vorausgeplant werden können, wie es die Autorin beschreibt. Es lassen sich lediglich durch ein Maximieren der Abstraktionshöhe Einsätze zu (Einsatz-)Stichwörtern gruppieren, beispielsweise „Akuter Thoraxschmerz“. Bei jedem Notruf obliegt es initial den Disponent:innen der Leitstellen das richtige Stichwort auszuwählen und die dazu passende Checkliste zu suchen und zu öffnen. Dies ist gegenwärtig ein Prozess, der aktiv durch die Disponent:innen anzustrengen ist und parallel zu einer Notrufabfrage („Notrufgespräch“) erfolgen muss. Hier entsteht bei der Allokation von Aufmerksamkeit bereits ein Zielkonflikt, der kognitive Kosten mit sich bringt. Hier kann angenommen werden, dass sich verschiedene Systeme der Notrufabfrage, die sich in den an die Disponent:innen gestellten Aufgaben unterscheiden, auch hinsichtlich der kognitiven Auslastung unterscheiden. Als Beispiele aus der Praxis sind Unterschiede zwischen „Standardisierter Notrufabfrage“, einer freien Abfrage und der „Strukturierten medizinischen Ersteinschätzung für Deutschland (SmED)“ zu erwarten (Neubert, 2022, S. 10). Letzteres wird vor allem im Bereich der kassenärztlichen Notdienste verwendet.

Auch bei einfach zu führenden Notrufabfragen kann angenommen werden, dass die Arbeitsbelastung vor allem auf Grund der mentalen und zeitlichen Anforderungen nicht zu vernachlässigen ist. Bei Notrufabfragen, bei denen Anrufende schwer zu verstehen sind, in einer fremden Sprache reden, Hintergrundgeräusche laut, emotionalisierend wirken oder sich erst ein Überblick über eine unübersichtliche Lage erarbeitet werden muss, ist anzunehmen, dass die mentale Arbeitsbelastung (Task Load) zusätzlich signifikant erhöht wird. Hofinger (2003, S. 11) erwähnt jedoch auch, dass insbesondere bei der sprachlichen Verarbeitung nur wenige kognitive Prozesse in kurzer Zeit abgebildet werden können. Hier scheint es wichtig, dass dies auch für die Prozesse „Notrufdialog“ und „manuelles Auswählen von Standard Operating Procedures“ berücksichtigt wird. Ein intelligentes System muss automatisiert und kontextsensitiv die Informationen (im Sinne einer „hirnleistungsgerechten“ Visualisierung) bereitstellen, die zum gefragten Zeitpunkt benötigt werden und nicht-benötigte oder redundante Informationen und Distraktoren unterdrücken⁸.

Bei einem Ausblick in die Zukunft ist davon auszugehen, dass die Tätigkeiten der BOS-Leitstellen weiterhin von Multifaktorialität und oftmals auch von Komplexität geprägt sein werden. Es ist ebenso davon auszugehen, dass systemseitige Transformationsprozesse auch weiterhin zu steigenden Ansprüchen an die Leitstellendisponent:innen führen. So ist bereits jetzt absehbar, dass sich Leitstellen der BOS immer mehr von Notrufannahmestellen zu Gatekeepern transformieren (Neubert, 2022) und dabei auch digitale Datenlieferanten, wie z. B. Apps oder Notrufsysteme aus Fahrzeugen (eCall) berücksichtigt werden müssen. Während aktuell die Entscheidung über den Zugang zur medizinischen Notfallversorgung bei den Patient:innen liegt und eine Gatekeeper-Funktion der Leitstellen nicht etabliert ist, ergibt sich hier ein primärer Ansatz für die Reformierung der Notfallstrukturen, indem gemeinsame Notfallleitstellen (GNL) betrieben werden. Diese integrieren den kassenärztlichen Notdienst („Hausarzt-Notdienst“) und tragen so zu einer Diversifizierung extraklinischer Ressourcen bei. Durch das Einbinden dieser Ressourcen ist der Behandlungspfad für Patient:innen, die keiner rettungsdienstlichen Versorgung und keiner unmittelbaren Hospitalisierung bedürfen, nun auch bei den BOS-Leitstellen integriert, wovon eine deutliche Entlastung der Rettungsmitteln zu erwarten ist. Die Zusammenlegung kann einerseits durch eine räumliche Zusammenlegung, aber auch durch Vernetzung beider Einheiten, bei struktureller und räumlicher Trennung realisiert werden. Es ist fraglich, ob eine räumliche und strukturelle Zusammenlegung in allen Bundesländern mittelfristig realistisch ist (Neubert, 2022, S. 9), so dass die Notwendigkeit einer sicheren und effizienten Vernetzung entsteht. In puncto Vernetzung stellen sich nicht nur Fragen, die auf die Architektur des digitalen Ökosystems abzielen, sondern auch auf ein personen-zentriertes Design der Systemerweiterung. Erinnernd an den ersten Teil dieses Kapitels muss diese Erweiterung so gestaltet sein, dass ein intelligentes System automatisiert und kontextsensitiv die Informationen

⁸ Dieses Kapitel deckt sich mit Kapitel 4 des Exposé zum Promotionsvorhaben des Autors

hirnleistungsgerecht bereitstellt, die zum gefragten Zeitpunkt benötigt werden und redundante Informationen und Distraktoren unterdrückt.

2 Forschungsfragen

2.1 Abhängige Variable Aufmerksamkeit

Forschungsfrage 1a: Wie unterscheiden sich standardisierte und frei geführte Notrufabfrage hinsichtlich der durch Eye-Tracking quantifizierten Aufmerksamkeitsverteilung der Leitstellendisponent:innen in Bezug auf visualisierte Inhalte des Einsatzleitsystems?

Forschungsfrage 1b: Wie unterscheiden sich die Phase der Notrufabfrage von anschließenden Phasen der Disposition hinsichtlich der durch Eye-Tracking quantifizierten Aufmerksamkeitsverteilung der Leitstellendisponent:innen in Bezug auf visualisierte Inhalte des Einsatzleitsystems?

2.2 Abhängige Variable (AV) „Stress“

Forschungsfrage 2a: In welcher Ausprägung deuten Herzfrequenzvariabilität (SDNN, RMSSD), elektrodermalen Aktivität (Hautleitwert) und Pupillometrie auf erhöhten Mental Workload während der Notrufabfrage und während der weiteren Einsatzsachbearbeitung bei Leitstellendisponent:innen hin?

Forschungsfrage 2b: Wie unterscheidet sich “Design SPELL” von der Standardisierten Notrufabfrage (SNA) und einer freien Abfrage hinsichtlich der Feststellbarkeit sympathikotoner Parameter bei Leitstellendisponent:innen während der der Notrufabfrage?

Forschungsfrage 2c: In welcher Beziehung steht die Ausprägung physiologischer Stressparameter mit der empfundenen informatorisch-mentalener Arbeitsbelastung?

In darauffolgenden Studien können weitere unabhängige Variablen berücksichtigt werden, z. B. die Auswirkungen von Früh-, Spät- und Nachtdienst sowie von unterschiedlichen Einsatzleitsystemen. Ebenso kann die weitere Integration von SPELL evaluiert werden, wenn der anwendungsfähige Reifegrad des digitalen Ökosystems erreicht ist.

4 Forschungsstand

4.1 Datenlage Notrufabfrage und Einsatzbearbeitung

Verfügbare Daten zu deutschen BOS-Leitstellen und deren Arbeit sind im Vergleich zu anderen hochentwickelten Rettungssystemen noch rar. Werden wissenschaftliche Arbeiten aus anderen Ländern herangezogen, ist zu berücksichtigen, dass sich das deutsche System, bestehend aus Rettungsdienst, qualifiziertem Krankentransport und (zukünftig vielleicht integriertem) kassenärztlichen Bereitschaftsdienst, stark von anderen Systemen unterscheidet. Dies gilt auch in der Notrufabfrage und Einsatzbearbeitung.

Als Beispiel kann im Vergleich Bonn-Birmingham auf die unterschiedliche Ausbildung des Leitstellenpersonals (Rettungsassistent versus 360-Stunden-Schulung) verwiesen werden, aber auch auf die unterschiedliche Effektivität des Systems: „Unsere Analyse zeigt, dass Birmingham hinsichtlich der Logistik das effizientere System betreibt als Bonn. Eine vorbildliche Leitstellenorganisation in Birmingham mit automatisierter Fahrzeug- und Notfalllokalisierung sowie Computer-unterstützter Disposition führt zu einer größeren Effizienz hinsichtlich der Fahrzeugauslastung, die 2-fach höher ist als in Bonn.“ (Fischer et al., 2004, S. 641) Weiterhin stellen die Autoren fest, dass moderne und effizient geplante Organisations- und Dispatch-Strukturen in Birmingham die aufzubringenden Kosten reduzieren und weisen so auch auf eine ökonomische Komponente hin.

Auch in puncto standardisierte Notrufabfrage (SNA) bestehen bereits angeführte Unterschiede, was im Vergleich zwischen USA, Großbritannien und Deutschland deutlich wird. Die standardisierte Notrufabfrage stützt sich auf verschiedene Softwaresysteme, die über eine Abfrage entlang eines Entscheidungsbaumes eine Priorisierung ermöglichen sollen, wobei die sog. Überbeschickung genauso vermieden werden soll wie eine Unterbeschickung. Die unterschiedlichen Notrufabfragesysteme sollen eine einheitliche Entscheidungsbasis herstellen und so zur Reduzierung von „Bias“ und „Noise“ beitragen. Eine Versorgung ohne Priorisierung, die unabhängig vom Meldebild nur durch die Eingangszeit des Notrufs gesteuert wird, ist als unökonomisch, ineffektiv und ethisch problembehaftet zu beschreiben, was die sog. Sheffield-Studie deutlich aufzeigt (Nicholl, 1996). Hinsichtlich der Ausgestaltung der SNA kann beobachtet werden, dass die USA und Großbritannien häufig auf das sog. AMPDS (Advanced Medical Priority Dispatch) zugreifen, während dies in Deutschland kaum verbreitet ist und in manchen europäischen Ländern gar kritisch diskutiert wird. So berichten Trimmel et al. (2006) von einem signifikanten Anstieg der Gesamteinsatzzahlen der Notarztthubschrauber um 78% bei einer gleichzeitigen Zunahme der Fehleinsätze um 186% nach Einführung von AMPDS Österreich. Zurückkommend auf die Notrufabfrage in Deutschland muss angemerkt werden, dass es eine ganze Reihe unterschiedlicher Softwaresysteme für die SNA am Markt verfügbar ist. Eine Übersicht gibt Mayr (2020) und fasst zusammen, dass alle Systeme sich darin gleichen, „dass sie Fragen und Antworten in Form von Struktur, Algorithmen oder Protokollen vorgeben, die [...] zu einer einheitlichen Informationsbasis führen (sollen).“ (Mayr, 2020) Luiz et al. (2019) heben diesbezüglich hervor, dass die

SNA und die telefonische Anleitung von Anrufer:innen zur Wiederbelebung durch Leitstellenmitarbeiter:innen (T-CPR) zentrale Forderungen von Fachgesellschaften (Fischer et al., 2016; Monsieurs et al., 2015), des Bundesverbands der Ärztlichen Leiter Rettungsdienst (2008) und des Fachverbands Leitstellen e. V. (Fachverband Leitstellen e.V., 2014) sind. Trotz des hohen Stellenwerts der SNA muss konstatiert werden, dass im deutschsprachigen Raum noch kein flächendeckender Einsatz der SNA beobachtet werden kann (Luiz et al., 2019; Mayr, 2020) und die SNA bis dato nur in sechs deutschen Bundesländern gesetzlich gefordert wird (Trautmann et al., 2022, S. 5).

Während im internationalen Kontext mannigfaltige Daten zu einsatztaktischen Parametern zu finden sind, bestehen kaum belastbare empirische Daten zu den Leitstellendisponent:innen der BOS, deren Wahrnehmung und biopsychosozialer Aspekte. Wahlgreen et al. (2021) untersuchten in ihrer Fragebogenstudie (n=61) Disponent:innen einer 911-Leitstelle in Snohomish County (USA) hinsichtlich des wahrgenommenen Stresslevels und befragten die Teilnehmenden hinsichtlich der Ursachen. Bei berufsbezogenen Stressoren gaben 23,1% „call volume/work load“ als Ursache an, 12,8% „stressful calls/callers“ und 12,8% gaben „policy and procedures“ an. Zu ähnlichen Ergebnissen kommen Aumiller (2008) und Hart (1995) bei Disponent:innen im Polizei- und Vollzugsdienst. Zu differenzierteren Ergebnissen kommen Herbig und Müller (2016), im Rahmen einer anlassbedingten Gefährdungsbeurteilung einer Integrierten Leitstelle. So betonen sie die hohen mental-informatorischen Anforderungen. U. a. stellten sie fest, dass etwa alle 8,77 Sekunden eine andere Informationsquelle betrachtet wurde, was in 410,7 Blickrichtungswechsel pro Stunde (Range = 314 – 492) kumuliert und im Rahmen der zitierten Arbeit als hohe informatorisch-mentale Belastung beschrieben wird.

4.2 Eye-Tracking

Eye Tracking (ET) ist eine experimentelle Methode zur Aufzeichnung von Augenbewegungen und der Position des Blicks über Zeit und Aufgaben hinweg. Es ist eine gängige Methode zur Beobachtung der Verteilung der visuellen Aufmerksamkeit (Carter & Luke, 2020, S. 49). Moderne Video-Eye-Tracker verfolgen die Augen durch Erkennung der Pupille. Das bedeutet, dass viele ET auch effektiv für die Pupillometrie eingesetzt werden können. Die Pupillometrie ist eine Technik, bei der Veränderungen im Durchmesser der Pupille aufgezeichnet werden. Die Größe der Pupille ändert sich als Reaktion auf Änderungen der Leuchtdichte, eine Änderung die eine Latenzzeit von etwa 200 ms hat (Ellis, 1981, nach Carter & Luke, 2020). Die Pupillen weiten sich auch als Reaktion auf interne kognitive/affektive Prozesse wie Aufmerksamkeitsverschiebungen, Motivation, geistige Anstrengung und kognitive Belastung. „Klassische“ ET-Studien liegen im Kontext der BOS-Leitstellen nicht vor. Herbig und Müller (2016) haben „in Anlehnung an „Eye-Tracking“-Verfahren [...] die Blickrichtungswechsel zwischen den Monitoren siebenmal für je eine Stunde zu unterschiedlichen Zeiten und an unterschiedlichen ELP erfasst“. Das technische Setting wird nicht weiter beschrieben.

4.3 Hautleitwert

Der Hautleitwert ist ein Parameter aus der Gruppe der elektrodermalen Aktivität, der leitfähigkeits- und Potenzialveränderungen der Haut reflektiert (Döring et al., 2015, S. 517). Die Höhe der Hautfähigkeit ist abhängig vom Füllzustand der sekretorischen Ausführungsgänge der Haut mit Schweiß. Je höher dieser gestiegen ist, desto höher ist die elektrische Hautleitfähigkeit, so Döring et al. Ein Anstieg des Hautleitwerts ist erst nach kurzer Latenz zu erwarten.

Messungen der elektrodermalen Aktivität, wie z. B. der Hautleitwert, sind für die Überwachung der Aktivität des autonomen Nervensystems sehr gut geeignet, da diese Aktivität vom sympathischen Zweig des autonomen Nervensystems bestimmt wird, der bei Stress dominant ist (Jacobs et al., 1994, S. 1171). Außerdem hat sich gezeigt, dass die elektrodermale Reaktion, ebenso wie Indizes der kardialen Leistungsfähigkeit wie Blutdruck und Herzfrequenz, bei psychischem Stress im Labor und unter anderen bedrohlichen, neuartigen oder herausfordernden Bedingungen zuverlässig ansteigt. So Jacobs et al.

Die Erhebung des Hautleitwerts konnte bereits verschiedenen Studien im Kontext der Mensch-Computer-Interaktion Reliabilität beweisen (Tronstad et al., 2022). So berichten Liapis et al. (2015) in ihrem Quasi-Experiment über die Messung des Hautleitwerts bei der Lösung verschiedener Zwischenfälle am PC, wie eine unerwartet fehlende Datei, Hardware Probleme, geringe Netzwerkgeschwindigkeit etc. Healey und Picard (2005) stellen ein Methode zur Erfassung und Analyse physiologischer Daten während realer Fahraufgaben vor, um das relative Stressniveau eines Fahrers zu bestimmen. Dabei stützen sie sich vor allem auf den Hautleitwert.

4.4 Herzfrequenz und Herzfrequenzvariabilität

Die Herzrätigkeit ist das Ergebnis einer komplexen Verschränkung aus neuronaler und kardialer Tätigkeit, die von weiteren Regelkreisläufen beeinflusst wurden.

„Während die Herzschlagfrequenz Information über den Beanspruchungsgrad des Herz-Kreislauf-Systems liefert, macht die Herzfrequenzvariabilität (Heart Rate Variability, HRV) darüber hinaus Aussagen über die Qualität der Herz-Kreislaufregulation. [...] Hauptwirkfaktor ist dabei das Zusammenspiel von Sympathikus und Parasympathikus. Diese beiden Nerven steuern die Aktivität des Herzens.“ (Hottenrott 2007 nach Böckelmann, 2012, S. 276).

Die Analyse der Herzfrequenzvariabilität (HRV) ist eine geeignete Methode, um nicht-invasiv die Leistung des zentralen Nervensystems zu untersuchen (Malik, 1996, S. 1060). Die HRV-Analyse ist mittlerweile die am häufigsten verwendete Messung des autonomen Herz-Kreislauf-Regulierungssystems (Borowicz & Banach, 2014, S. 547). Je höher die HRV ist, desto besser kann sich das Herz auf Stressfaktoren einstellen. Kurzfristige Änderungen der HRV sind als Beanspruchungsantwort auf einwirkende Belastungsfaktoren zu verstehen (Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V., 2006, S. 2). Die Herzfrequenzvariabilität ist dabei durch

eine Reihe endogener Faktoren wie Alter und Geschlecht und durch exogene wie physische und psychische Belastung, Körperlage, Tageszeit, Medikamente beeinflusst (Böckelmann, 2012, S. 277). Deshalb muss bei Messwiederholungen die Körperlage, die Uhrzeit (circadiane Rhythmus), der zeitliche Abstand zur letzten Nahrungsaufnahme usw. möglichst gleich sein.

Die AWMF-Leitlinie nennt „Beanspruchungsanalysen an Arbeitsplätzen mit vorrangig psychischen Belastungen im Rahmen von arbeits- bzw. psychophysiologischen Untersuchungen“ explizit als Fragestellungen für die Herzrhythmusanalyse (Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V., 2006, S. 1). Im Kontext der Beanspruchungsanalysen an Arbeitsplätzen bietet sich die HRV-Messungen an, da sie die vegetative/autonome Balance des Organismus widerspiegelt (Parameter der Aktivierung), was sie für psychophysiologische Untersuchungen an Arbeitsplätzen mit psychischer Belastung qualifiziert, so vorzitierte Leitlinie.

Mittels einer EKG-Aufzeichnung über mehrere Herzzyklen hinweg kann u. a. der Abstand von einer Herzkammererregung (Kammerdepolarisation) zur nächsten gemessen werden. Dies ist das sog. Normal-to-Normal-Intervall (NN), in der deutschsprachigen Literatur auch als RR-Intervall bekannt. Diese Bezeichnung rührt von der sog. R-Zacke her, die mit der höchsten Amplitude der Kammerdepolarisation den markantesten Punkt im EKG darstellt. Die summative Betrachtung unterschiedlicher NN-Intervalle wird als Herzfrequenzvariabilität bezeichnet (Malik, 1996, S. 1044). Zur statistisch-diagnostischen Auswertung der HRV können (neben Herzfrequenz) verschiedene Parameter herangezogen werden. Die am weitest verbreiteten sind nach Malik der Tabelle 1 zu entnehmen. Diese beschreiben auch jene Parameter (zzgl. der Herzfrequenz), die im Lauf der Studie erhoben werden.

Tabelle 1: HRV-Parameter

Parameterbezeichnung	Abkürzung	Kurzbeschreibung
Standard deviation of the NN intervals	SDNN	Quadratwurzel der Herzfrequenzvarianz. Da die Varianz mathematisch gleich der Gesamtleistung der Spektralanalyse ist, spiegelt die SDNN alle zyklischen Komponenten wider, die für die Variabilität im Aufzeichnungszeitraum verantwortlich sind. Indikator für Gesamtvariabilität.
Square root of the mean squared differences of successive NN intervals	RMSSD	Quadratwurzel des quadratischen Mittelwertes der Summe aller Differenzen zwischen benachbarten NN-Abständen. Indikator für parasympathischen Einfluss.
Proportion of NN50	pNN50	Prozentsatz der Paare benachbarter NN-Abstände, die sich jeweils >50 ms voneinander unterscheiden. Indikator für spontane Änderung der Herzfrequenz.

5 Methodisches Vorgehen

5.1 Untersuchungsdesign

5.1.1 Mixed-Methods-Design-Research

Die angestrebte Forschungsarbeit bedient sich quantitativer und qualitativer Ansätze, die im Gesamtkontext als Mixed-Methods-Design-Research zu verstehen sind. In Bezug auf Mixed-Methods-Design ist die Komplementaritäts-These hervorzuheben. Dieser zu Folge sind quantitatives und qualitatives Paradigma wissenschaftstheoretisch grundlegend verschieden, wobei von einer wechselseitigen Noninferiorität auszugehen ist (Döring et al., 2015, S. 74). Der Komplementaritäts-These folgend weisen beide Paradigmen Schwächen und Limitation auf. Diesen Einschränkungen soll durch ein Mixed-Methods-Design, das jeweils vollständige qualitative und quantitative Teilstudien aufweist (ibidem), begegnet werden.

Tabelle 2: Kurzbeschreibung Mixed-Methods-Design

Forschungsfrage		Design / Paradigma und Kurzbeschreibung
1a	Wie unterscheiden sich standardisierte und frei geführte Notrufabfrage hinsichtlich der durch Eye-Tracking quantifizierten Aufmerksamkeitsverteilung der Leitstellendisponent:innen in Bezug auf visualisierte Inhalte des Einsatzleitsystems?	Quantitativ, Auswertung von Eye-Tracking-Daten
1b	Wie unterscheiden sich die Phase der Notrufabfrage von anschließenden Phasen der Disposition hinsichtlich der durch Eye-Tracking quantifizierten Aufmerksamkeitsverteilung der Leitstellendisponent:innen in Bezug auf visualisierte Inhalte des Einsatzleitsystems?	
2a	In welcher Ausprägung deuten Herzfrequenzvariabilität (SDNN, RMSSD), elektrodermalen Aktivität (Hautleitwert) und Pupillometrie auf erhöhten Mental Workload während der Notrufabfrage und während der weiteren Einsatzsachbearbeitung bei Leitstellendisponent:innen hin?	Quantitativ, Auswertung Hautleitwert und Herzfrequenzvariabilität
2b	Wie unterscheidet sich "Design SPELL" von der Standardisierten Notrufabfrage (SNA) und einer freien Abfrage hinsichtlich der Feststellbarkeit sympathikotoner Parameter bei Leitstellendisponent:innen während der der Notrufabfrage?	
2c	In welcher Beziehung steht die Ausprägung physiologischer Stressparameter mit der empfundenen informatorisch-mentalen Arbeitsbelastung?	Quantitativ durch Fragebogen (NASA-TLX) und ggf. qualitativ durch Kurzinterview, beides im Selbstberichtsverfahren (Adressat:innen sind Versuchspersonen aus Forschungsfragen 2a und 2b)

4.1.2 Within-Subjects-Design versus Between-Subjects-Design

Bei Within-Subjects-Designs (WSD) durchläuft dieselbe Versuchsperson nacheinander alle experimentellen Bedingungen. Between-Subjects-Designs (BSD) sehen dies nicht vor. Bei BSD wird jede Versuchsperson nur einer Stufe der unabhängigen Variable (UV) zugeordnet. Welches Design nun ins jeweilige Forschungsdesign integriert wird, hängt von statistischen, methodologischen und theoretischen Aspekten ab.

WSD gewähren eine Reihe höchst forschungsrelevanter Vorteile. So ist nicht von einer Konfundierung der UV mit personengebundenen Störvariablen, beispielsweise durch interindividuelle Differenzen, auszugehen (Sedlmeier & Renkewitz, 2018, S. 157), woraus eine höhere Sensitivität und Teststärke resultiert. Da in allen Bedingungen dieselben Versuchspersonen untersucht werden, sind alle personengebundenen Störvariablen bereits perfekt parallelisiert. In der Konsequenz können mit WSD auch kleine Effekte der UV entdeckt werden. Bedenkt man die Funktion statistischer Tests, die beispielsweise bei Teilerhebungen für die Auswertung von Mittelwert- oder Varianzunterschieden zwischen Stichproben verwendet werden, ist festzuhalten, dass der Effekt der UV an der Fehlervarianz relativiert wird (Sedlmeier & Renkewitz, 2018, S. 160). Das erklärt, warum statistische Tests, die im WSD angewendet werden, grundsätzlich leichter einen Populationseffekt entdecken. Verdeutlicht wird der Zusammenhang zwischen Stichprobengröße und Teststärke in Abb. 1. Zu sehen ist ein t-Tests für unabhängige Stichproben (rechts) und für abhängige Stichproben (links). Die zugrunde gelegten Parameter sind im jeweiligen Bild (Überschrift, zweite Reihe) ablesbar. Die A-priori-Teststärkenanalyse wurde mir G*Power (Version 3.1.9.7) berechnet.

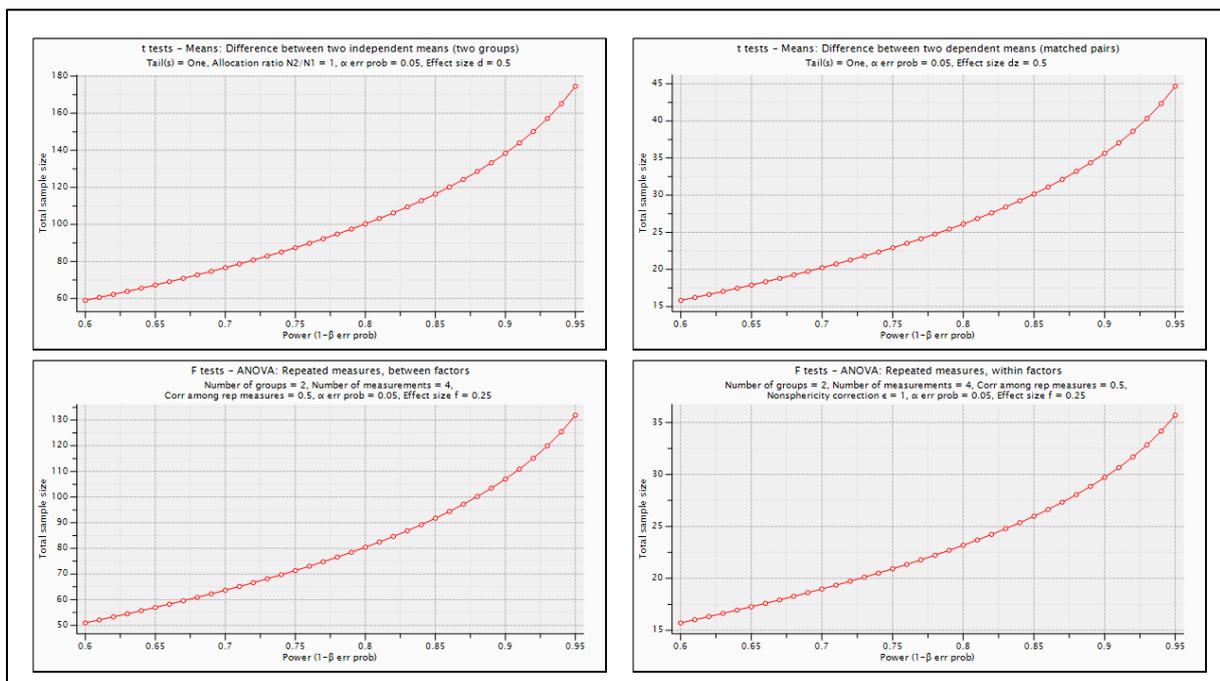


Abbildung 1: Stichprobengröße und Testpower (Queller: GPower)

Außerdem ermöglichen es WSD für jede einzelne Versuchsperson zu bestimmen, wie die UV auf sein/ihr Verhalten wirkt. WSD bieten außerdem einen Kontext und Vergleichsstandard, auf dessen Grundlage Versuchspersonen ein Urteil abgeben können. Das Fehlen solcher Landmarken kann bei BSD zu schwerwiegenden Verzerrungen führen, wie Birnbaum (1999) in seiner experimentellen Studie mit dem Titel „How to show that $9 > 221$ “ aufzeigt.

Durch die Stärken des WSD stellt sich also die Eingangsfrage, ob beim Beantworten der jeweiligen Forschungsfrage Aspekte vorliegen, die gegen dessen Anwendung sprechen. So ist die Initialfrage: Kann die entsprechende UV innerhalb einer Person sinnvoll untersucht werden? Weiterhin ist zu beachten, dass sog. Demand Characteristics auftreten können, die durch Vermutungen der Versuchspersonen über die Hypothesen im Experiment und eine darauffolgende Anpassung des Verhaltens erklärbar sind (Sedlmeier & Renkewitz, 2018, S. 161). Außerdem sind mögliche Positionseffekte und Carry-Over-Effekte zu bedenken. Während Positionseffekte kontrolliert werden können, beispielsweise durch verschiedene Balanciermethoden oder den Einsatz eines lateinischen Quadrats, ist das bei Carry-Over-Effekten kaum möglich. Die hohe Wahrscheinlichkeit von Carry-Over-Effekten stellt deshalb einen Grund dar, auf BSD auszuweichen.

Aus diesen Erwägungen heraus wird für jeden Forschungsfrage analysiert, ob ein WSD oder BSD zur Beantwortung eingesetzt wird (s. Tabelle 2).

Tabelle 3: Within-Subjects versus Between-Subjects-Design

Forschungsfrage		WSD oder BSD	Begründung
1a	Wie unterscheiden sich standardisierte und frei geführte Notrufabfrage hinsichtlich der durch Eye-Tracking quantifizierten Aufmerksamkeitsverteilung der Leitstellendisponent:innen in Bezug auf visualisierte Inhalte des Einsatzleitsystems?	Messwiederholung Within-Between- Interaktion	Die Leitstellen arbeiten entweder mit einer freien Abfrage oder mit einer SNA. So sind die Leitstellendisponent:innen, die eine freie Abfrage durchführen, nicht mit der SNA vertraut und vice versa. Ein WSD ohne Between-Interaktion würde sich nur mit einer intensiven Familiarisierung eignen, wobei in diesem Fall die Akzeptanz der Forschung durch die Disponent:innen kritisch hinterfragt werden muss.
1b	Wie unterscheiden sich die Phase der Notrufabfrage von anschließenden Phasen der Disposition hinsichtlich der durch Eye-Tracking quantifizierten Aufmerksamkeitsverteilung der Leitstellendisponent:innen in Bezug auf visualisierte Inhalte des Einsatzleitsystems?		Hier steht zusätzlich die Einsatzbearbeitung außerhalb der Notrufabfrage im Fokus. Berücksichtigt werden zunächst nur Leitstellen, die eine funktionale Trennung zwischen Calltaker und Dispatcher ⁹ aufweisen, um die Vergleichbarkeit zu gewährleisten.
2a	In welcher Ausprägung deuten Herzfrequenz-variabilität (SDNN, RMSSD), elektrodermalen Aktivität (Hautleitwert) und Pupillometrie auf erhöhten Mental Workload während der Notrufabfrage und während der weiteren Einsatzsachbearbeitung bei Leitstellendisponent:innen hin?		Weiterhin bestehen die Gruppen „SNA“ und „frei Abfrage“, die im BSD verglichen werden. Es entstehen im Lauf der Einsatzbearbeitung intraindividuelle Messzeitpunkte, die als a.) Messwiederholung im WSD verstanden werden können, aber b.) auch die unabhängige Variable (UV) „SPELL“ als Simulatoranteil (s. 4.4) berücksichtigen. Da SPELL derzeit noch nicht im Realbetrieb getestet werden kann, ist ein BSD mit der UV SPELL nicht zielführend, weil angenommen werden kann, dass Unterschiede zwischen Realbetrieb und Simulatorbetrieb hinsichtlich auftretender Stressoren bestehen.
2b	Wie unterscheidet sich „Design SPELL“ von der Standardisierten Notrufabfrage (SNA) und einer freien Abfrage hinsichtlich der Feststellbarkeit sympathikotoner Parameter bei Leitstellendisponent:innen während der der Notrufabfrage?		
2c	In welcher Beziehung steht die Ausprägung physiologischer Stressparameter mit der empfundenen informatorisch-mentalenen Arbeitsbelastung?		

⁹ Calltaker sind primär für den Prozess der Notrufabfrage verantwortlich, Dispatcher für Alarmierung und weitere Einsatzmitteldisposition. Das Vorgehen der Leitstellen ist hier nicht einheitlich. So ist einerseits eine funktionale und personelle Trennung von Calltaker und Dispatcher zu beobachten und andererseits, dass beide Funktionen zusammengelegt ausgeübt werden.

5.2 Stichprobenziehung

Durchgeführt wird eine Teilerhebung, daher wird mit Stichproben gearbeitet. Bei der Auswahl der Leitstellendisponent:innen sind die Altersgruppen 18-34, 35-54 und > 55 zu berücksichtigen. Eine disproportionale Ziehung ist als ultima ratio zulässig.

Ausschlusskriterien:

- Akute und einschränkende Erkrankung (z.B. grippaler Infekt) oder einschränkender Gesundheitszustand (ausgeprägte und als leistungsmindernd wahrgenommene Müdigkeit)
- Fortgeschrittener Diabetes Mellitus (Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V., 2006, S. 5)
- Herz-Kreislauf-Erkrankungen:
 - Z.n. Herztransplantation und/oder klinisch relevante myokardiale Dysfunktion (Malik, 1996, S. 1054)
 - persistierendes Vorhofflimmern (Koshy et al., 2018)
- Neurologische und psychiatrische Erkrankungen:
 - Epilepsie (Borowicz & Banach, 2014)
 - Psychotische Störung mit oder ohne medikamentöse Therapie (Alvares et al., 2016),
 - Schizophrenie bei medikamentöser Therapie (Iwamoto et al., 2012)
- Medikamenten- und Substanzeinnahme:
 - Antidepressiva und/oder Neuroleptika, außer Risperidon (Alvares et al., 2016; Silke et al., 2002)
 - Valproat und/oder Phenobarbital (Borowicz & Banach, 2014, S. 548)
 - Opioide (Garland & Howard, 2021) und/oder langwirksame Benzodiazepine (z.B. Diazepam, Flunitrazepam, Ketazolam) innerhalb der letzten 72 Stunden
 - Kurzwirksame Benzodiazepine (z.B. Lorazepam) innerhalb der letzten 24 Stunden
 - Rauschmittelkonsum inkl. Cannabinoide in den letzten 72 Stunden
- Ophthalmologische Einschränkungen:
 - Brillenträger:innen können grundsätzlich teilnehmen, aber die Kompatibilität der Brille mit dem Eyetracker ist von der Bauart (Gesamtbreite, -höhe, Form von Gläser und Steg etc.) abhängig. Daher sind Nicht-Brillenträger:innen zu bevorzugen.

5.3 Datenerhebung

Es sind drei Gruppen zu berücksichtigen: 1. SNA (Leitstelle Ludwigshafen), 2. Freie Abfrage (Leitstelle Mainz?), 3. SPELL (Leitstellen Ludwigshafen und Mainz). Die Datenerhebung der Forschungsfragen kann kombiniert werden, d.h. die Messung von Pupillometrie, Herzfrequenzvariabilität, Hautleitwert und Aufmerksamkeitsverteilung kann zeitlich synchron bei jeweils einer Versuchsperson stattfinden. Es sind mindestens zwei Messwiederholungen unter möglichst gleichen Bedingungen zu realisieren. Das

impliziert, dass nicht nur die gleichen Personen benötigt werden, sondern diese auch zu möglichst gleichen Uhrzeiten und Schichtzeiten gemessen werden müssen. Die Messwiederholungen müssen an unterschiedlichen Tagen stattfinden. Direkt aufeinanderfolgende Tage sind möglich, aber nicht notwendig. Hier ist eine enge Abstimmung zwischen Forschungspartner:innen (IESE & TUDA) und Anwendungspartner:innen (DRK ITC) notwendig.

Im Weiteren befindet sich die Beschreibung des Ablaufs, analog zur tatsächlichen Reihenfolge im Experiment, wie er auch in Abbildung 2 dargestellt wird. Zu beachten ist, dass SPELL derzeit in Form eines Demonstrators zur Notrufabfrage eingesetzt werden kann, jedoch kein Einsatzleitsystem darstellt. Aus diesem Grund ergibt sich der Unterschied in Phase 2.

Ebenso ist zu berücksichtigen, dass das Experiment zwar als Feldversuch zu verstehen ist, jedoch Phase 2 in zwei Anteile, nämlich Realbetrieb (2a) und Simulatoranteil (2b), getrennt ist. Die ist darin begründet, dass SPELL zum jetzigen noch nicht als Notrufabfragetool im Realbetrieb eingesetzt werden kann. Dies bedingt die Testung im Simulationsbetrieb (Leitstellensimulation). Dabei unterscheidet sich der Leitstellenarbeitsplatz nur dadurch, dass der Einsatzleitreechner nicht an den Realbetrieb der Leitstelle gekoppelt ist, sondern abgetrennt als „Übungsdummy“ fungiert. Da keine vergleichenden Daten der physiologischen Parameter (HRV, Hautleitwert, Pupillometrie) zwischen Simulations- und Realbetrieb vorliegen, ist unklar welche Auswirkungen dieser Unterschied tatsächlich hat. Aus diesem Grund werden zunächst physiologische Daten im Realbetrieb erhoben (jeweils mit SNA und freier Abfrage), danach im Simulationsbetrieb mit SPELL.

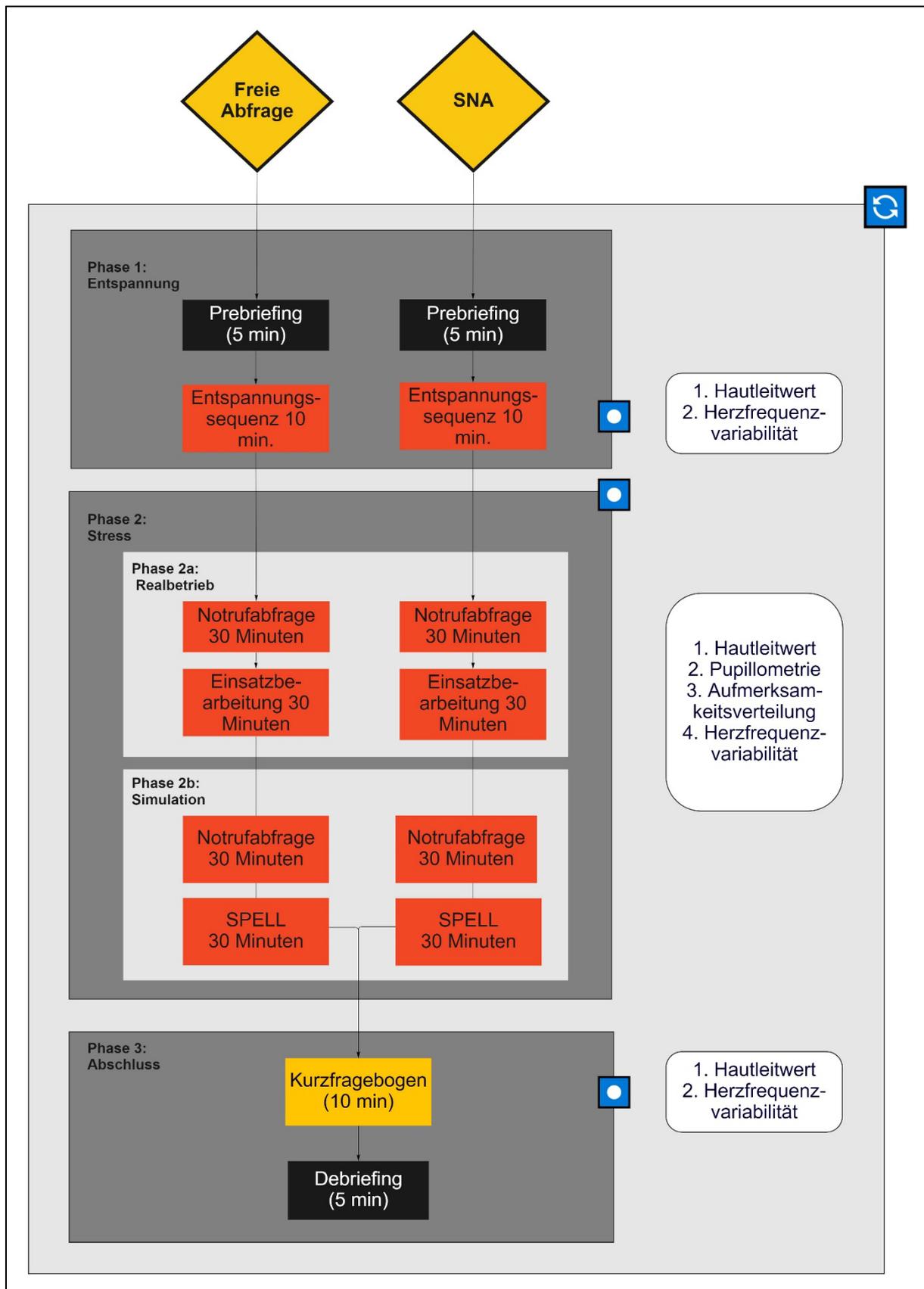


Abbildung 2: Versuchsablauf¹⁰

¹⁰ Symbole: „Weißer Punkt auf blauem Grund“: In dieser Phase oder in diesem Schritt erfolgt die Messung der physiologischen Parameter (s. weißer Kasten), „Wiederholungszeichen“: Messwiederholung

5.3.1 Prebriefing

Die Versuchspersonen werden mittels standardisierter Informationen über das Forschungsvorhaben informiert. Zur den Informationen gehören auch Angaben zur Datenerhebung, -verwendung, -aufbewahrung, -löschung, die Möglichkeit zum Widerruf und zum Abbruch der Teilnahme im laufenden Versuch. Zustimmung und Teilnahme sind freiwillig. Eine fehlende Zustimmung führt zum sofortigen Ausschluss. Im Prebriefing werden die bereits vorher kommuniziert und zur Selektion verwendeten Ausschlusskriterien abgefragt (s. 4.3).

5.3.2 Kurzfilm

Zu Beginn wird ein spannender Kurzfilm gezeigt, um eine Baseline-Messung von Herzfrequenz, Herzfrequenzvariabilität und Hautleitwert zu ermöglichen.

5.3.3 Eyetracking

Es werden vor allem Third- und Fourth Order Data (Sharafi et al., 2020) analysiert. Diese sind: Fixation count, Fixation duration und Percentage of fixations or fixation rate. Über Gazepaths und Heatplots können die Blickpfade analysiert werden. So entsteht ein exaktes Mapping, das Aufschluss über die Informationsverarbeitung, die Priorisierung und die Reihenfolge gibt und erlaubt, dass in zukünftigen Entwicklungen grafische Elemente hervorgehoben werden oder in den Hintergrund treten, um Informationen intelligent und kontextsensitiv zu visualisieren und um kognitive Ressourcen zu schonen. Als technische Parameter sind hauptsächlich Precision, Accuracy und ein Drift der Systeme zu berücksichtigen (Sharafi et al., 2020). Die Accuracy ist der Unterschied zwischen den wahren und gemessenen Blickdaten. Beispielsweise ist bei einer Accuracy von 0.5 Grad von einer 0.5 mm Streuung auf 50 cm um das tatsächliche foveale Fixationszentrum auszugehen. Die Precision gibt an, wie gut ein ET die gleichen Ergebnisse für zwei aufeinander folgende Blicke an derselben Stelle reproduzieren kann. Unter Drift versteht man die allmähliche Abnahme der Genauigkeit der Erfassung von Blickbewegungsdaten im Laufe der Zeit. Drift wird durch die Verschlechterung der Kalibrierung verursacht, die aufgrund der Physiologie des Auges, z. B. Veränderungen der Feuchtigkeit, und anderer Faktoren, wie einem simplen Verrutschen der Tracking-Brille, auftritt .

Bei der Umsetzung ist das Setting der Leitstellen zu berücksichtigen. Meist sind mehrere Bildschirme nebeneinandergeschaltet, teilweise auch zweireihig übereinander. In einigen Leitstellen wird dies noch durch weitere größere Bildschirme an einer großen Wand ergänzt. Somit beschränkt sich das Eyetracking aus technischer Sicht Eye-Tracking-Systeme. Als ET wird ein Pupils Labs Core verwendet (Pupil Labs, 2022), die mit einer World-Camera (60 Hz) und zwei Eye-Cameras (200 Hz) ausgestattet ist. Die Stromversorgung und Datenübertragung erfolgt per USB-C-Kabel. Ein Echtzeit-Algorithmus nutzt die Cornealreflektion der Pupille (Dark-Pupil-Technology), um deren Form und Position zu erkennen (Velisar & Shanidze, 2021, S. 1). Der ET verwendet einen modellbasierten Ansatz für die

Schätzung des Blicks, der die im Bild erfassten Pupillenparameter an ein optisches 3D-Modell anpasst und die Ausrichtung der optischen Achse des Auges in Bezug zu den Augenkameras schätzt, wie Valisar & Shanidze beschreiben. Zu Beginn des ET ist eine Kalibration der Eye-Kameras mit der Pupils-Core-Software durchzuführen. Aufgrund des zu erwartenden Drifts, sind notwendig Nachkalibrationen während der ET-Sequenz zu erwarten. Deshalb müssen die Eye-Kameras durch die dazugehörige Software hinsichtlich korrektem Tracking der Pupille beobachtet werden. Trotz robuster Methoden zur Fehlerkontrolle, wie sie Morimoto et al. (Morimoto et al., 1999) beschreiben, sind mögliche Störfaktoren nicht komplett auszuschließen, so dass angenommen werden muss, dass nicht alle Datensätze durchgehend reliable Daten aufweisen und im schlimmsten Fall komplett ausgeschlossen werden müssen. Dies wirkt sich auch auf die Planung der Stichprobengröße aus („Sicherheitsreserve“).

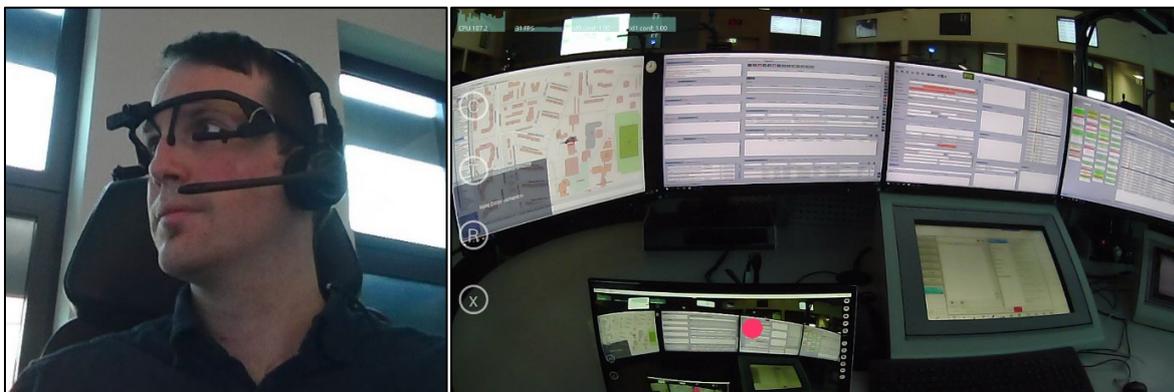


Abbildung 3 (rechts): World-Kamera - Aufzeichnung aus Disponent:innen-Perspektive (eigene Abbildung)

Abbildung 4 (links): Angelegte Pupil Core mit leitstellenüblichen Headset (eigene Abbildung)

5.3.4 Herzfrequenz und Herzfrequenzvariabilität

Hauptsächlich bestehen drei Möglichkeiten zur kontinuierlichen Erhebung der HRV: 1. Oberflächen-EKG nach Einthoven (Drei-Kanal-EKG), 2. Ableitung mit Brustgurt als Ein-Kanal-EKG, 3. Photoplethysmographie (z.B. in Wearables). Die Ableitung nach Einthoven ist damit verbunden, dass mindesten drei Elektroden (rechte Arm, linker Arm, links Bein) mit einem zentralen Geräte per Kabel verbunden werden müssen, was für die Leitstellen-Situation mangels Bewegungsfreiheit und erwartbarer Bewegungsartefakte kontraproduktiv ist. Bei der Ableitung mittels Brustgurt sind Akzeptanzprobleme durch den Anlageort zu erwarten und hygienische Aspekte zu berücksichtigen. Beispielsweise wird eine Desinfektion erforderlich, wodurch Textilanteile des Gurtes nass werden. Entweder müssen mehrere Geräte vorgehalten werden oder es muss den Teilnehmenden zugemutet werden, einen nassen Gurt unter der Kleidung anzulegen. Die Ableitung über ein Wearable scheint aus Akzeptanzperspektive am wenigstens problematisch zu sein. Hernando et al. (2018) berichten über die Verwendung der Apple Watch zur Analyse der HRV, in einem vergleichbaren Szenario. Sie schließen, dass die Apple Watch sich durch die Erfüllung aller Qualitätskriterien für eine HRV Messung qualifiziert

hat. Im Vergleich zwischen Brustgurt (Polar H7) und Apple Watch¹¹ konnten die Autor:innen keinen signifikanten Unterschied feststellen. Zu ähnlichen Ergebnissen kamen Weaver et al. (2019) im Vergleich der Apple Watches (Series 3 und 4) mit dem Brustgurt Polar H10. Lediglich bei Sportübungen mit hoher Intensität konnten mehr bewegungsbedingte Artefakte bei den Apple Watches festgestellt werden. Auch die Messung von Stresslevels ist bereits dokumentiert, beispielsweise von Park et al. (2022). Sie evaluierten die HRV von Pflegefachkräften mittel Apple Watch (4, 6 und/oder 6) während der Herz-Lungen-Wiederbelebung. Zur Analyse der Herzfrequenz heben Koshy et al. (2018) im Vergleich zwischen dem Wearable Fit Bit und der Apple Watch die Genauigkeit der Apple Watch hervor, decken jedoch Schwäche bei Vorhofflimmern auf.

5.3.5 Hautleitwert

Die Hautleitfähigkeit wird mittels Gelelektroden, die palmar an der nicht-dominanten Hand aufgeklebt werden (s. Abb. 4 und 5), abgeleitet. Im Vorversuch müssen ggf. Variationen der Elektrodenpositionierung unternommen werden, falls diese Position stören ist oder zu relevanten Artefakten führt. Ein Vergleich verschiedener Gelarten für Hautleitfähigkeitsmessungen deutet darauf hin, dass feste Hydrogelelektroden im Laufe der Zeit stabilere Messungen liefern, während nasse Gele eine stärkere Drift in der Messung verursachen können und anfälliger für Druckartefakte sind (Tronstad et al., 2022, S. 12).

Als Messgerät wird das Eda4Move genutzt (Movisense GmbH, o.J.). Die Stromversorgung wird über einen Akku hergestellt. Die Akkulaufzeit beträgt nach Herstellerangaben 4 Tage. Das Gerät wurde bereits überzeugend in diversen wissenschaftlichen Arbeiten eingesetzt (Blum & Fried, 2021; Borovac et al., 2021; Rother et al., 2021; van Halem et al., 2020).

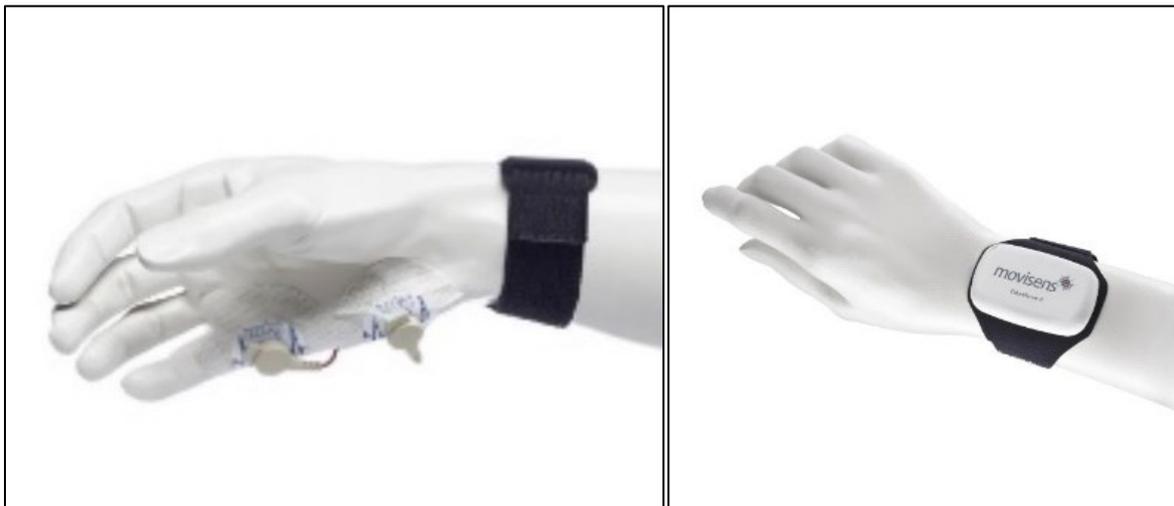


Abbildung 1 (links): Position der Elektroden (Quelle: Movisense GmbH, o.J.)

Abbildung 2 (rechts): Eda4Move-Einheit am Handgelenk (Quelle: Movisense GmbH, o.J.)

¹¹ Die Autor:innen nennen keine Modellbezeichnung der Apple Watch

Die Messung erfolgt in μ Siemens und kann über die Darstellung einer Kurve (Hautleitfähigkeitskurve) im Zeitverlauf visualisiert werden (Döring et al., 2015, S. 519). Ausgewertet wird vor allem die Frequenz signifikanter Peaks des Hautleitwerts. Einzelne Peaks die mit Daten der World-Kamera des Eyetrackers und Notizen aus den Beobachtungen kombiniert und so im Optimalfall auch gezielt Stressoren zugeordnet werden können. Alle Amplituden, die höher als der Mittelwert plus die Standardabweichung sind, werden als "signifikante Peaks" betrachtet und liefern somit Hinweise auf Stressspitzen.

5.3.6 Beobachtung

In Phase 2 findet zusätzlich eine Beobachtung der Teilnehmenden statt. Die umfasst:

- Messung des Zeitintervalls von Notrufannahme bis zur Alarmierung
- Dokumentation des tatsächlichen Taskloads
- Dokumentation von offensichtlich stressrelevanten Ereignissen (z.B. quantitative Belastungsspitzen, Besonderheiten, Zwischenfälle)

5.3.7 Fragebogen

Im Fragebogen wird der NASA-TLX (oder aquivalentes Instrument) mittels Online-Fragebogen an die Versuchspersonen adressiert. Der NASA Task Load Index ist ein Instrument zur subjektiven Bewertung der Arbeitsbelastung, mit dem Benutzer subjektive Bewertungen der Arbeitsbelastung von Bedienern durchführen können, die mit verschiedenen Mensch-Maschine-Schnittstellensystemen arbeiten (Hart, 2006). Ziel ist der Abgleich des Fragebogens als Selbstberichtsverfahren (subjektiv), mit den erhobenen physiologischen Parametern.

5.3.8 Debriefing

Das Debriefing soll den Versuchspersonen und Forschenden die Möglichkeit geben, offene Fragen zu klären und ein wertschätzenden Klima erhalten.

5.4 Datenaufbereitung und Datenanalyse

Hautleitwert: Aus dem Eda4-Move-Gerät können CSV-Dateien exportiert werden, die in RStudio aufbereitet und analysiert werden. Die Datenanalyse erfolgt mit RStudio.

Herzfrequenz und -variabilität: Die Datenanalyse erfolgt mit RStudio. Der Datenexport von Apple Watch zur RStudio wird derzeit erprobt und verschiedene Lösungswege gegeneinander abgewogen.

Eye-Tracking: Gazepath und Heatplots können mittels Pupil-Core-Software ausgewertet und visualisiert werden. Weitere Analysen der erhobenen ET-Daten finden mit RStudio statt.

Beobachtung: Die Beobachtung werden in einem standardisierten Protokoll pro Versuchsperson erfasst und einer qualitativen Inhaltsanalyse mit Maxqda 2022 unterzogen.

Statistische Auswertung: Die Stichproben (SNA versus frei Abfrage) können nach Prüfung der Voraussetzung am Datenmaterial, mittels t-Test auf signifikante Unterschiede analysiert werden. Dies wird für die Entspannungssequenz, die Notrufabfrage und Einsatzbearbeitung in Phase 2a sowie in Phase 2b für Notrufabfrage mit konventionellem System und mit SPELL geprüft.

Unterschiede zwischen den Messzeitpunkten können durch ANOVA (repeated measures, within-between-interaction) und Tukey HSD (1949) für verbundene Stichproben analysiert werden. Die Auswertung erfolgt mit RStudio.

5.5 Ergebnispräsentation

Die Ergebnisse werden dem SPELL-Konsortium intern präsentiert. Eine Verwertung der Ergebnisse in Form von Publikationen in peer-reviewed Journals ist vorgesehen.

Literaturverzeichnis

- Alvares, G. A., Quintana, D. S., Hickie, I. B. & Guastella, A. J. (2016). Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: a systematic review and meta-analysis. *Journal of psychiatry & neuroscience : JPN*, 41(2), 89–104. <https://doi.org/10.1503/jpn.140217>
- Aumiller, G., Corey, D., Allen, S., Brewster, J., Cuttler, M., Gupton, H. & Honig, A. (2008). Defining the Field of Police Psychology: Core Domains & Proficiencies. *Journal of Police and Criminal Psychology*, 23(1), 48. <https://doi.org/10.1007/s11896-008-9020-0>
- Bergmann, R. (2021). *Künstliche Intelligenz für das Notfallmanagement – Projekt SPELL erforscht KI-Plattform für Rettungsleitstellen*. CRISIS PREVENTION - Fachportal für Gefahrenabwehr, Innere Sicherheit und Katastrophenhilfe. <https://crisis-prevention.de/kommunikation-it/kuenstliche-intelligenz-fuer-das-notfallmanagement-projekt-spell-erforscht-ki-plattform-fuer-rettungsleitstellen.html>* [Zuletzt geprüft am 22.07.2021].
- Birnbaum, M. H. (1999). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods*, 4(3), 243–249. <https://doi.org/10.1037/1082-989X.4.3.243>
- Bischof, G. C. (2018). *Grober Behandlungsfehler durch Rettungsleitstelle, Beweislastumkehr*. <https://anwalt-bischof.de/grober-behandlungsfehler-durch-rettungs-leitstelle-beweislastumkehr/>* [Zuletzt geprüft am 05.10.2021].
- Blum, W. & Fried, P. (2021). *How does it feel to walk in Berlin? Designing an Urban Sensing Lab to explore walking emotions through EDA sensing* [Master thesis]. Aalborg University, Aalborg.
- Böckelmann, I. (2012). Analyse der Herzfrequenzvariabilität (HRV) – praktische Relevanz. *Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie*, 62, 275–279.
- Borovac, A., Stuldreher, I., Thammasan, N. & Brouwer, A.-M. (2021). *Validation of wearables for electrodermal activity(EdaMove) and heart rate*. 12th International Conference on Methods and Techniques in, Kraków, Poland. <https://doi.org/10.6084/m9.figshare.13013717>
- Borowicz, K. K. & Banach, M. (2014). Antiarrhythmic drugs and epilepsy. *Pharmacological reports : PR*, 66(4), 545–551. <https://doi.org/10.1016/j.pharep.2014.03.009>
- Bundesverband Ärztliche Leiter Rettungsdienst Deutschlands e.V. (2008). *Empfehlung des Bundesverbandes Ärztlicher Leiter Rettungsdienst Deutschlands e.V. zur Qualitätsverbesserung der Disposition und Beratung in Leitstellen*. <https://www.bv-aelrd.de/index.php/downloads/category/2-stellungnahmen?download=74:20080914-empf-bv-aelrd-qualitaetsverbesserung-der-disposition-und-beratung-in-leitstellen>* [Zuletzt geprüft am 05.07.2022].
- Carter, B. T. & Luke, S. G. (2020). Best practices in eye tracking research. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 155, 49–62. <https://doi.org/10.1016/j.ijpsycho.2020.05.010>
- Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V. (2006). *Herzrhythmusanalyse in der Arbeitsmedizin*. https://www.awmf.org/uploads/tx_szleitlinien/002-021_S1_Herzrhythmusanalyse_in_der_Arbeitsmedizin_08-2005_08-2010.pdf* [Zuletzt geprüft am 08.07.2022].
- Döring, N., Bortz, J., Pöschl, S., Werner, C. S., Schermelleh-Engel, K., Gerhard, C. & Gäde, J. C. (2015). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer-Lehrbuch. Springer.
- Dörner, D. (2008). Umgang mit Komplexität. In A. von Gleich & S. Gößling-Reisemann (Hrsg.), *Industrial Ecology* (S. 284–302). Vieweg+Teubner. https://doi.org/10.1007/978-3-8351-9225-6_24

- Eisenbast, C. (2021). *Künstliche Intelligenz für Leitstellen und Lagezentren – das Projekt SPELL - Teil 1*. Fraunhofer IESE. https://www.iese.fraunhofer.de/blog/spell/* [Zuletzt geprüft am 05.10.2021].
- Fachverband Leitstellen e.V. (2014). *Positionspapier zu den zukünftigen Schwerpunkten der Verbandsarbeit*. https://www.fvlst.de/wp-content/uploads/2021/02/positionspapier_fachverband_leitstellen.pdf* [Zuletzt geprüft am 05.07.2022].
- Fischer, M., Kehrberger, E., Marung, H., Moecke, H., Prückner, S., Trentzsch, H. & Urban, B. (2016). Eckpunktepapier 2016 zur notfallmedizinischen Versorgung der Bevölkerung in der Prähospitalphase und in der Klinik. *Notfall + Rettungsmedizin*, 19(5), 387–395. <https://doi.org/10.1007/s10049-016-0187-0>
- Fischer, M., Krep, H., Wierich, D., Heister, U., Hoeft, A., Edwards, S., Castrillo-Riesgo, L. G. & Krafft, T. (2004). Effektivitäts- und Effizienzvergleich der Rettungsdienstsysteme in Birmingham (UK) und Bonn (D). *Gesundheitsökonomie & Qualitätsmanagement*, 9(06), 369–381. <https://doi.org/10.1055/s-2004-813243>
- Garland, E. L. & Howard, M. O. (2021). Prescription opioid misusers exhibit blunted parasympathetic regulation during inhibitory control challenge. *Psychopharmacology*, 238(3), 765–774. <https://doi.org/10.1007/s00213-020-05729-z>
- Hart, P. M., Wearing, A. J. & Headey, B. (1995). Police stress and well-being: Integrating personality, coping and daily work experiences. *Journal of Occupational and Organizational Psychology*, 68(2), 133–156. <https://doi.org/10.1111/j.2044-8325.1995.tb00578.x>
- Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–908. <https://doi.org/10.1177/154193120605000909>
- Healey, J. A. & Picard, R. W. (2005). Detecting Stress during Real-World Driving Tasks Using Physiological Sensors. *Trans. Intell. Transport. Syst.*, 6(2), 156–166. <https://doi.org/10.1109/TITS.2005.848368>
- Herbig, B. & Müller, A. (2016). Hohe Belastungen in einer integrierten Rettungsleitstelle. *NeuroTransmitter*, 27(9), 12–18. <https://doi.org/10.1007/s15016-016-5636-y>
- Hernando, D., Roca, S., Sancho, J., Alesanco, Á. & Bailón, R. (2018). Validation of the Apple Watch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects. *Sensors (Basel, Switzerland)*, 18(8). <https://doi.org/10.3390/s18082619>
- Hofinger, G. (2016). Entscheidungsfindung. In A. Hackstein, V. Hagemann, F. von Kaufmann & H. Regener (Hrsg.), *Handbuch Simulation* (S. 104–106). S+K Verlagsgesellschaft Stumpf + Kossendey mbH.
- Iwamoto, Y., Kawanishi, C., Kishida, I., Furuno, T., Fujibayashi, M., Ishii, C., Ishii, N., Moritani, T., Taguri, M. & Hirayasu, Y. (2012). Dose-dependent effect of antipsychotic drugs on autonomic nervous system activity in schizophrenia. *BMC psychiatry*, 12, 199. <https://doi.org/10.1186/1471-244X-12-199>
- Jacobs, S. C., Friedman, R., Parker, J. D., Tofler, G. H., Jimenez, A. H., Muller, J. E., Benson, H. & Stone, P. H. (1994). Use of skin conductance changes during mental stress testing as an index of autonomic arousal in cardiovascular research. *American heart journal*, 128(6), 1170–1177.
- Koshy, A. N., Sajeev, J. K., Nerlekar, N., Brown, A. J., Rajakariar, K., Zureik, M., Wong, M. C., Roberts, L., Street, M., Cooke, J. & Teh, A. W. (2018). Smart watches for heart rate assessment in atrial arrhythmias. *International journal of cardiology*, 266, 124–127. <https://doi.org/10.1016/j.ijcard.2018.02.073>

- Liapis, A., Katsanos, C., Sotiropoulos, D., Xenos, M. & Karousos, N. (Hrsg.). (2015). *Lecture Notes in Computer Science. Recognizing Emotions in Human Computer Interaction: Studying Stress Using Skin Conductance*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-22701-6>
- Luiz, T., Marung, H., Pollach, G. & Hackstein, A [A.] (2019). Implementierungsgrad der strukturierten Notrufabfrage in deutschen Leitstellen und Auswirkungen ihrer Einführung [Degree of implementation of structured answering of emergency calls in German emergency dispatch centers and effects of the introduction in daily practice]. *Der Anaesthetist*, 68(5), 282–293. <https://doi.org/10.1007/s00101-019-0570-6>
- Malik, M. (1996). Heart Rate Variability - Standards of Measurement, Physiological Interpretation, and Clinical Use. *Circulation*, 93(5), 1043–1065.
- Mayr, B. (2020). Strukturierte bzw. standardisierte Notrufabfrage: Leisten die Systeme tatsächlich, was sie vorgeben zu leisten? [Structured or protocol-based dispatching programs]. *Notfall + Rettungsmedizin*, 23(7), 505–512. <https://doi.org/10.1007/s10049-020-00733-4>
- Monsieurs, K. G., Nolan, J. P., Bossaert, L. L., Greif, R., Maconochie, I. K., Nikolaou, N. I., Perkins, G. D., Soar, J., Truhlář, A., Wyllie, J. & Zideman, D. A. (2015). European Resuscitation Council Guidelines for Resuscitation 2015: Section 1. Executive summary. *Resuscitation*, 95, 1–80. <https://doi.org/10.1016/j.resuscitation.2015.07.038>
- Morimoto, C. H., Koons, D., Amit, A., Flickner, M. & Zhai, S. (1999). Keeping an eye for HCI. In *XII Brazilian Symposium on Computer Graphics and Image Processing (Cat. No. PRO0481)* (S. 171–176). IEEE Comput. Soc. <https://doi.org/10.1109/SIBGRA.1999.805722>
- Neubert, T. (2022). Von Notrufannahmestellen zum Gatekeeper: Die veränderte Rolle der Leitstellen. *BOS Leitstelle aktuell*, 12(2), 8–22.
- Nicholl, J. (1996). *The Safety and Reliability of Priority Dispatch Systems: Final Report to the Department of Health*.
- Park, H. J., Choi, D., Park, H. A. & Lee, C. A. (2022). Nurse evaluation of stress levels during CPR training with heart rate variability using smartwatches according to their personality: A prospective, observational study. *PloS one*, 17(6), e0268928. <https://doi.org/10.1371/journal.pone.0268928>
- Pupil Labs. (2022). *Pupil Labs Core*. https://pupil-labs.com/products/core/* [Zuletzt geprüft am 08.07.2022].
- Rother, A., Niemann, U., Hielscher, T., Völzke, H., Ittermann, T. & Spiliopoulou, M. (2021). Assessing the difficulty of annotating medical data in crowdworking with help of experiments. *PloS one*, 16(7), e0254764. <https://doi.org/10.1371/journal.pone.0254764>
- Sedlmeier, P. & Renkewitz, F. (2018). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler* (3. Aufl.). *Pearson Studium - Psychologie*. Pearson.
- Sharafi, Z., Sharif, B., Guéhéneuc, Y.-G., Begel, A., Bednarik, R. & Crosby, M. (2020). A practical guide on conducting eye tracking studies in software engineering. *Empirical Software Engineering*, 25(5), 3128–3174. <https://doi.org/10.1007/s10664-020-09829-4>
- Silke, B., Campbell, C. & King, D. J. (2002). The potential cardiotoxicity of antipsychotic drugs as assessed by heart rate variability. *Journal of psychopharmacology (Oxford, England)*, 16(4), 355–360. <https://doi.org/10.1177/026988110201600410>
- Trautmann, R., Reuter-Oppermann, M. & Christiansen, j. (2022). *PSAP-G-ONE - Eine explorativ-deskriptive Studie über Leitstellen der nichtpolizeilichen Gefahrenabwehr in der Bundesrepublik Deutschland*. Deutsche Gesellschaft für Rettungswissenschaften e.V. https://www.dgre.org/download/2067/* [Zuletzt geprüft am 05.07.2022].

- Trimmel, H., Wodak, A. & Voelckel, W. (2006). Hubschrauberdisposition mit dem Advanced-Medical-Priority-Dispatch-System – Erwartungen erfüllt? *Notfall + Rettungsmedizin*, 9(5), 437–445.
<https://doi.org/10.1007/s10049-006-0837-8>
- Tronstad, C., Amini, M., Bach, D. R. & Martinsen, Ø. G. (2022). Current trends and opportunities in the methodology of electrodermal activity measurement. *Physiological measurement*, 43(2).
<https://doi.org/10.1088/1361-6579/ac5007>
- Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, 5(2), 99.
<https://doi.org/10.2307/3001913>
- van Halem, S., van Roekel, E., Kroencke, L., Kuper, N. & Denissen, J. (2020). Moments that Matter? On the Complexity of using Triggers Based on Skin Conductance to Sample Arousing Events within an Experience Sampling Framework. *European Journal of Personality*, 34(5), 794–807.
<https://doi.org/10.1002/per.2252>
- Velisar, A. & Shanidze, N. (2021). Noise in the Machine: Sources of Physical and Computation Error in Eye Tracking with Pupil Core Wearable Eye Tracker. In A. Bulling, A. Huckauf, H. Gellersen, D. Weiskopf, M. Bace, T. Hirzle, F. Alt, T. Pfeiffer, R. Bednarik, K. Krejtz, T. Blascheck, M. Burch, P. Kiefer, M. Dodd & B. Sharif (Hrsg.), *ACM Symposium on Eye Tracking Research and Applications* (S. 1–3). ACM. <https://doi.org/10.1145/3450341.3458495>
- Wahlgren, K., Fraizer, A., Taigman, M., Gay, M., Ronald, Qi, W., Faudere, D., Jones, A. & Olola, C. (2021). Factors Contributing to Stress Levels of Emergency Dispatchers, 8, 2020.
- Weaver, L., Wooden, T. & Grazer, J. (2019). Validity of Apple Watch Heart Rate Sensor Compared to Polar H10 Heart Rate Monitor [Georgia College and State University]. *Journal of Student Research*. Vorab-Onlinepublikation. <https://doi.org/10.47611/jsr.vi.662>