

## **Ein Ansatz zur Initiierung proaktiver Dialoge von Sprachassistenten im Fahrzeug**

Lesley-Ann MATHIS

*Fraunhofer Institut für Arbeitswirtschaft und Organisation,  
Nobelstraße 12, D-70569 Stuttgart*

**Kurzfassung:** Die Vision zukünftiger Sprachassistenten im Fahrzeug vermittelt das Bild eines digitalen Begleiters, der mit Nutzenden proaktiv interagiert. Um die Nutzerakzeptanz für proaktives Verhalten von Sprachassistenten sicherzustellen, müssen verschiedene Einflussfaktoren für die Wahl des Zeitpunkts und Inhalts der Ansprache betrachtet werden. Ziel dieses Dissertationsprojekts ist es, einen Ansatz zu entwickeln, der proaktives Verhalten von Sprachassistenten im automatisierten Fahrzeug zu einem geeigneten Zeitpunkt mit relevanten Inhalten auslöst. Dabei wird ein hybrider Ansatz verfolgt, der sowohl wissensbasierte Strukturen für das Entscheidungsverhalten verwendet als auch auf die Generierung natürlicher Sprache durch ein großes Sprachmodell zurückgreift. Um die Akzeptanz der Ansprachen aus Nutzersicht sowie die Anwendbarkeit des Ansatzes in der Praxis sicherzustellen, ist eine Validierung mit Nutzenden und einem Entwicklungsteam geplant.

**Schlüsselwörter:** Sprachassistenten, Proaktivität, automatisiertes Fahren, Sprachmodelle, Fahrzeuginteraktion, nutzerzentriert

### **1. Motivation**

Automobilhersteller präsentieren zukünftige Sprachassistenten im Fahrzeug als digitale Begleiter, deren Verhalten über das einer funktionalen Schnittstelle hinausgeht. Toyota stellte im *Concept-i* beispielsweise die Assistentin „Yui“ vor, die Vorschläge basierend auf dem aktuellen Zustand oder der Situation der Nutzenden macht, um das Fahrerlebnis zu verbessern (Lugano 2017). Ein wesentliches Merkmal von Yui, das sie von aktuellen Sprachassistenten im Fahrzeug differenziert, ist ihr proaktives Verhalten. Proaktivität gilt als charakteristisches Merkmal menschlichen Verhaltens und wird in der Arbeits- und Organisationspsychologie als vorausschauendes und selbstinitiiertes Verhalten definiert, mit dem Ziel die Situation zu verbessern (Nerdinger et al. 2014). Nutzende erwarten von zukünftigen Sprachassistenten im Fahrzeug, dass sie proaktiv handeln und Wissen über die Nutzenden und den Kontext besitzen (Schmidt & Braunger 2018). Die Gestaltung proaktiven Dialogverhaltens bringt jedoch verschiedene Herausforderungen mit sich, da für die Akzeptanz der Interaktion eine geeignete Situation und ein relevantes Thema für die Ansprache von Nutzenden gewählt werden muss (Nothdurft et al. 2015).

Für sprachbasierte Interaktion spielen große Sprachmodelle (engl. *Large Language Models*, LLM) eine zunehmend wichtige Rolle. Ihre Stärke liegt z. B. in der natürlichen Sprachgenerierung, die auf Basis der empfangenen Systemeingabe erfolgt. Diese Eingaben, auch Prompts genannt, werden als Schlüssel für den optimalen Einsatz von

LLM angesehen (Teubner et al. 2023). Der Einsatz von Sprachmodellen für sprachbasierte Interaktion ist Gegenstand aktueller Forschung und wird für Assistenten im Fahrzeug von Automobilherstellern wie z. B. Mercedes-Benz durch die Integration von ChatGPT bereits umgesetzt (Mercedes-Benz Group 2023). Davon verspricht sich der Hersteller, dass Nutzende über die bisherigen Funktionen hinaus Dialoge mit größerer Themenvielfalt führen können und so eine zunehmend intuitive Interaktion mit dem Assistenten im Fahrzeug möglich wird. Bisherige Sprachassistenten handeln jedoch reaktiv und reagieren ausschließlich auf Eingaben der Nutzenden. Die Umsetzung von proaktivem Dialogverhalten unter Einsatz von LLM bietet das Potenzial, positive Nutzungserlebnisse für Passagierinnen und Passagiere durch Sprachassistenten im Fahrzeug zu schaffen.

## 2. Stand der Technik

Das vorliegende Dissertationsvorhaben greift auf bisherige Forschung im Bereich proaktive Sprachassistenten sowie dem jungen Forschungsfeld des sog. Promptings, dem Generieren von Prompts für LLM, zurück.

### 2.1 Proaktive Sprachassistenten

Verschiedene Studien zu Sprachassistenten untersuchten, wann geeignete Zeitpunkte für proaktive Ansprachen von Nutzenden sind. So schlagen Kim et al. (2018) ein Modell zur Vorhersage geeigneter Zeitpunkte beim manuellen Fahren vor, das vor allem Fahrzeug- und Umgebungsdaten einbezieht. Eine Studie von Semmens et al. (2019) zeigte, dass Situationen mit konstanter Fahrt geradeaus oder bei Stillstand für proaktive Ansprachen im Gegensatz zu Kurven sowie Spur- oder Geschwindigkeitswechseln bevorzugt werden. Für das automatisierte Fahren ist jedoch nicht alleine die Fahrsituation entscheidend für die Akzeptanz einer proaktiven Ansprache, sondern auch der aktuelle Gefühlszustand der Nutzenden, die verbleibende Fahrzeit sowie das Ausführen fahrfremder Tätigkeiten (Mathis et al. 2023a). Im häuslichen Umfeld werden Ansprachen durch einen proaktiven Smart-Home Lautsprecher vor allem kurz nach dem Ankommen, beim Wechsel von Tätigkeiten und bei Aktivitäten wie Relaxen, Nutzung des Internets oder sozialer Medien oder dem Schauen von Videos akzeptiert (Cha et al. 2020).

Der Zeitpunkt und Inhalt einer proaktiven Ansprache sind eng miteinander verknüpft, da die Relevanz einer Ansprache je nach Zeitpunkt variieren kann (Nothdurft et al. 2015). Bisherige Arbeiten zu relevanten Inhalten während der Fahrt konzentrierten sich auf exemplarische Anwendungsfälle, wie z. B. proaktive Vorschläge für die Parkplatz- oder Tankstellensuche (Schmidt et al. 2019). Für das automatisierte Fahren ergeben sich neue Anwendungsfälle für proaktive Ansprachen, u. a. beim Ausführen von fahrfremden Tätigkeiten. So sind Angebote und Vorschläge zur Steigerung des Wohlbefindens oder des Nutzungserlebnisses während der Fahrt denkbar (Mathis et al. 2023b), wie Vorschläge zur Minderung von Reisekrankheit bei kurvigen Streckenabschnitten, Unterstützung bei fahrfremden Tätigkeiten wie Arbeitstätigkeiten oder das Anleiten einer Achtsamkeitsübung für das mentale Wohlbefinden.

## 2.2 Prompting von LLM für die Mensch-Technik Interaktion

Der Einsatz von LLM eröffnet neue Möglichkeiten für die Sprachinteraktion im Fahrzeug, ohne vorab aufwendige Datenerhebungen oder das Training von Modellen durchzuführen. Um einen passenden und nützlichen Output des Sprachmodells zu erhalten, ist eine standardisierte Gestaltung des Prompts notwendig. Wie Spasić & Janković (2023) für ein Beispiel im Bildungssektor darlegen, eignet sich beispielsweise eine Struktur aus einer Rollendefinition für das LLM, einer Instruktion sowie einem Startwort um den gewünschten Output zu erhalten.

Ein Ansatz zur Nutzung von LLM für eine dialogbasierte Interaktion auf dem Smartphone zeigt auf, dass große Sprachmodelle mit entsprechenden Strategien zur Promptgenerierung z. B. für das Beantworten von Nutzerfragen erfolgreich eingesetzt werden können (Wang et al. 2023). Allerdings bringt das bisher begrenzte Wissen über die Steuerbarkeit und Kontrollierbarkeit von LLM Herausforderungen mit sich, die für einen zuverlässigen Einsatz für die Interaktion mit Nutzenden weiter erforscht werden sollten (Wang et al. 2023).

## 3. Lösungsansatz

Bisherige Arbeiten zu proaktiven Sprachassistenten im Fahrzeug haben gemeinsam, dass einzelne Aspekte der Proaktivität untersucht oder prototypisch umgesetzt wurden. Im Rahmen dieser Dissertation wird ein Ansatz entwickelt, der eine proaktive Interaktion mit den Nutzenden zu einem geeigneten Zeitpunkt mit relevantem Inhalt während der automatisierten Fahrt initiiert. Dazu liefert der Ansatz als Ergebnis einen Prompt, den ein LLM zu einer proaktiven Ansprache verarbeiten und an die Nutzenden ausgeben kann.

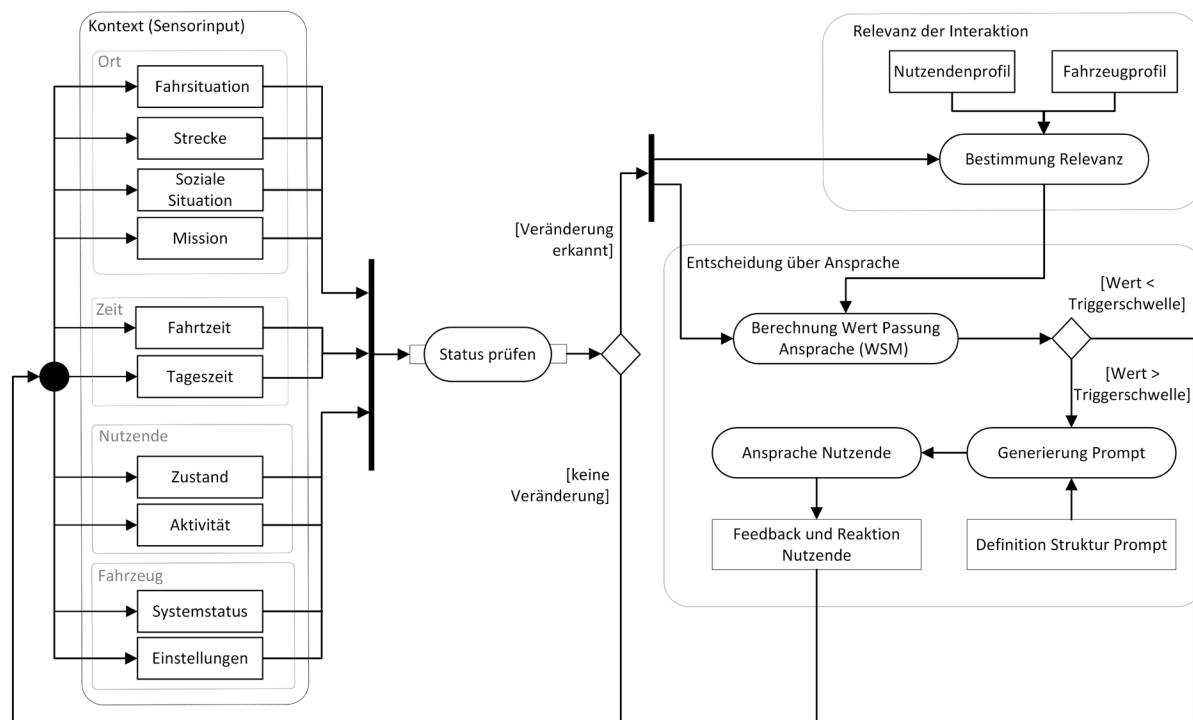
### 3.1 Konzeptionelles Modell

Die aktuelle Version des konzeptionellen Modells des Ansatzes ist in Abb. 1 dargestellt und wird im Folgenden näher erläutert. Basierend auf bisheriger Forschung zu proaktiver Sprachassistenten und kontextabhängigen Diensten in Fahrzeugen wurde eine Taxonomie von Einflussfaktoren für die Wahl eines geeigneten Ansprachezeitpunktes im Fahrzeug entwickelt. Diese Kontextfaktoren werden in die Kategorien Ort, Zeit, Nutzende und Fahrzeug geclustert. Die Faktoren erhalten individuelle Gewichtungen, ob sie einen förderlichen oder hinderlichen Einfluss auf einen geeigneten Ansprachezeitpunkt nehmen.

Für die Auswahl von Inhalten für die Ansprachen wird davon ausgegangen, dass eine proaktive Ansprache nur dann relevant ist, wenn dadurch ein Mehrwert für Nutzende entsteht. Ausgewählt werden Inhalte, wenn bestimmte auslösende Ereignisse im gegebenen Kontext zutreffen, die durch Informationen aus dem Nutzenden- oder Fahrzeugprofil bedingt werden (Yorke-Smith et al. 2012). Beispielsweise ist eine Ansprache zur Vermeidung von Reisekrankheit dann relevant, wenn die Nutzerin bzw. der Nutzer anfällig für Reisekrankheit ist, gerade ein Buch liest und ein kurviger Streckenabschnitt auf der automatisierten Fahrt bevorsteht.

Die Entscheidung, ob eine proaktive Ansprache ausgelöst werden soll, wird anhand einer Triggerschwelle bestimmt. Dafür werden die Gewichtungen der Kontextfaktoren

und eine Gewichtung für die Relevanz des Inhalts gemäß eines Weighted-Sum-Modells addiert (siehe Abb. 1). Liegt ein geeigneter Zeitpunkt und relevanter Inhalt vor, wird ein Prompt generiert und eine proaktive Ansprache mittels eines LLM formuliert. Im Rahmen des Ansatzes wird ein standardisiertes Muster zur Gestaltung des Prompts definiert, das von einem Sprachmodell zur Generierung der proaktiven Ansprache genutzt wird. Darüber hinaus ist ein Feedback Modul im Ansatz integriert, um zukünftige Interaktionen entsprechend dem Feedback der Nutzenden anzupassen.



**Abbildung 1:** Konzeptionelles Modell des Ansatzes zur Initiierung proaktiver Dialoge in UML dargestellt (Arbeitsversion)

### 3.2 Umsetzung

Im Rahmen der Dissertation wird eine Referenzimplementierung entwickelt, um das proaktive Dialogverhalten interaktiv zu modellieren. Ziel der Implementierung ist es, die Wahl eines geeigneten Zeitpunkts und relevanter Inhalte auf der Grundlage eines simulierten Datensatzes (z. B. Fahrzeug- und Umgebungsdaten aus einer Fahrsimulationssoftware) zu demonstrieren. Die Gewichtungen der Kontextfaktoren werden basierend auf bisheriger Forschung und eigener Vorstudien abgeleitet.

Der resultierende Prompt wird mit einem ausgewählten Sprachmodell getestet, um eine Ansprache mit beispielhaften Anwendungsfällen zu initiieren. Als große Sprachmodelle kommen für die Umsetzung GPT 4 vom Unternehmen OpenAI sowie Chat Bison von Google infrage, die beide im Rahmen einer Vorstudie zur Generierung proaktiver Ansprachen aus Sicht der Nutzenden sehr gut bewertet wurden. Für die Implementierung des Ansatzes wurde das Node-RED Framework gewählt (<https://node-red.org/>). Node-RED ist ein Programmierwerkzeug, das sich besonders für die Modellierung von eventgesteuerten Softwarearchitekturen eignet und so die Initiierung von

proaktivem Verhalten durch bestimmte Ereignisse, wie z. B. die Änderung des Nutzerzustandes, geeignet darstellen kann.

## 4. Validierung

Eine Validierung des Ansatzes unter Einbeziehung der Nutzenden ist geplant, um die Akzeptanz des modellierten Dialogverhaltens zu untersuchen sowie die Anwendbarkeit des Ansatzes sicherzustellen. Die oben beschriebene Referenzimplementierung wird für die Evaluation eingesetzt, an der Nutzende und Experten beteiligt sind.

### 4.1 Studie mit Nutzenden

Im Rahmen einer Studie in einem High-Fidelity Fahrsimulator wird die Referenzimplementierung angewendet, um zu prüfen, dass mittels des Ansatzes proaktive Ansprachen zu geeigneten Zeitpunkten mit relevanten Inhalten initiiert werden. Dafür wird eine automatisierte Fahrt mit einem bestimmten Fahrtzweck simuliert, bei der proaktive Ansprachen mittels der Referenzimplementierung ausgelöst und durch eine Text-to-Speech Funktion wiedergegeben werden.

Die Studie sieht vor, dass Teilnehmende die proaktiven Ansprachen während der Fahrt in-situ anhand eines Items bezüglich der Eignung des Zeitpunkts und der Relevanz der Ansprache bewerten. Aus Sicht der Nutzenden soll sichergestellt werden, dass die proaktiven Ansprachen als attraktiv und nützlich beurteilt werden und das Dialogverhalten des Assistenten als möglichst menschlich und natürlich wahrgenommen wird. Zusätzlich soll deshalb nach der Fahrt ein videogestütztes Retrospective Think-Aloud (Elbabour et al. 2017) durchgeführt werden. Zur Bewertung der Ansprachen wird dabei die modulare Erweiterung des User Experience Questionnaire (UEQ+) eingesetzt. Die Ansprachen werden retrospektiv anhand der UEQ+ Skalen *Attraktivität*, *Nützlichkeit* und *Antwortverhalten des Sprachassistenten* von den Teilnehmenden bewertet (Schrepp & Thomaschewski 2023).

### 4.2 Validierung mit Expertinnen und Experten

Darüber hinaus wird der Ansatz in Fokusgruppen mit Entwicklungsteams aus dem Bereich Dialogsysteme und Sprachassistentenentwicklung im Fahrzeug evaluiert. Fokusgruppen sind für Softwareapplikationen eine etablierte Methode, um qualitative Daten aus Sicht der Anwendenden zu erheben (Shull et al. 2008). Dieser Teil der Validierung konzentriert sich auf die praktische Anwendbarkeit des Ansatzes im Fahrzeug und die Interpretierbarkeit für Entwicklerinnen und Entwickler. Die Fokusgruppe sieht vor, dass nach einer Erklärung des Ansatzes und Erprobung anhand ausgewählter Beispiele mit der Referenzimplementierung eine strukturierte Diskussion geführt wird. Dabei soll aus Sicht des Entwicklungsteams reflektiert werden, inwiefern der Ansatz um neue Anwendungsfälle erweiterbar ist und die Ausgaben des LLM interpretierbar für zukünftige Anpassungen sind.

## 5. Literatur

- Cha N, Kim A, Park CY, Kang S, Park M, Lee J-G, Lee S, Lee U (2020) Hello There! Is Now a Good Time to Talk? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4: 1–28.
- Elbabour F, Alhadreti O, Mayhew P (2017) Eye Tracking in Retrospective Think-Aloud Usability Testing: Is There Added Value? *Journal of Usability Studies* Vol. 12: 95–110.
- Kim A, Choi W, Park J, Kim K, Lee U (2018) Interrupting Drivers for Interactions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2: 1–28.
- Lugano G (2017) Virtual assistants and self-driving cars. In: *Proceedings of the 15th International Conference on ITS Telecommunications (ITST)* 1–5.
- Mathis L-A, Werner K, Schmidt GJ (2023a) Proaktive Sprachassistenten in automatisierten Fahrzeugen. In: Hölzle K, Kreimeyer M, Roth D, Maier T, Riedel O (Hrsg.) *Stuttgarter Symposium für Produktentwicklung SSP 2023* 488–499. Fraunhofer IAO, Stuttgart.
- Mathis L-A, Werner K, Widloirther H (2023b) Exploring use cases and user perception of a proactive voice assistant in automated vehicles. In: *Human Factors in Transportation*. AHFE International.
- Mercedes-Benz Group (2023) Mercedes-Benz takes in-car voice control to a new level with ChatGPT | Mercedes-Benz Group. <https://group.mercedes-benz.com/innovation/digitalisation/connectivity/car-voice-control-with-chatgpt.html> (02.01.2024).
- Nerdinger FW, Blickle G, Schaper N (2014) *Arbeits- und Organisationspsychologie*. Mit 51 Tabellen. 3., vollst. überarb. Aufl. 2014 Springer, Berlin.
- Nothdurft F, Ultes S, Minker W (2015) Finding Appropriate Interaction Strategies for Proactive Dialogue Systems – An Open Quest. In: K. Jokinen, M. Vels (Hrsg.) *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication* 73–80. Linköping University Electronic Press.
- Schmidt M, Braunger P (2018) Towards a speaking style-adaptive assistant for task-oriented applications. In: *Studientexte zur Sprachkommunikation* 143–150.
- Schmidt M, Stier D, Werner S, Minker W (2019) Exploration and assessment of proactive use cases for an in-car voice assistant. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung*: 148-155.
- Schrepp M, Thomaschewski J (2023) Handbook for the modular extension of the User Experience Questionnaire. <https://ueqplus.ueq-research.org/> (30.12.2023).
- Semmens R, Martelaro N, Kaveti P, Stent S, Ju W (2019) Is Now A Good Time? In: Brewster S, Fitzpatrick G, Cox A, Kostakos V (Hrsg.) *CHI 2019. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, May 4-9, 2019, Glasgow, Scotland UK 1–12. ACM, New York, NY.
- Shull F, Singer J, Sjøberg DIK (2008) *Guide to advanced empirical software engineering*. Springer, London.
- Spasić AJ, Janković DS (2023) Using ChatGPT Standard Prompt Engineering Techniques in Lesson Preparation: Role, Instructions and Seed-Word Prompts. In: *2023 58th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)* 47–50. IEEE.
- Teubner T, Flath CM, Weinhardt C, van der Aalst W, Hinz O (2023) Welcome to the Era of ChatGPT et al. *Bus Inf Syst Eng* 65: 95–101.
- Wang B, Li G, Li Y (2023) Enabling Conversational Interaction with Mobile UI using Large Language Models. In: Schmidt A, Väänänen K, Goyal T, Kristensson PO, Peters A, Mueller S, Williamson JR, Wilson ML (Hrsg.) *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* 1–17. ACM, New York, NY, USA.
- Yorke-Smith N, Saadati S, Myers K, Morley DN (2012) The Design of a Proactive Personal Agent for Task Management. *Int. J. Artif. Intell. Tools* 21: 1–30.



Gesellschaft für Arbeitswissenschaft e.V.

## Arbeitswissenschaft in-the-loop

**Mensch-Technologie-Integration  
und ihre Auswirkung auf Mensch,  
Arbeit und Arbeitsgestaltung**

70. Kongress der  
Gesellschaft für Arbeitswissenschaft e.V.

Institut für Arbeitswissenschaft und  
Technologiemanagement IAT  
Universität Stuttgart

In Zusammenarbeit mit dem Fraunhofer-Institut für  
Arbeitswirtschaft und Organisation IAO

06. – 08. März 2024

---

## GfA-Press

---

**Bericht zum 70. Arbeitswissenschaftlichen Kongress vom 06. – 08. März 2024**

**Institut für Arbeitswissenschaft und Technologiemanagement (IAT), Universität Stuttgart**

**In Zusammenarbeit mit: Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO), Stuttgart**

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.

Sankt Augustin: GfA-Press, 2024

ISBN 978-3-936804-34-8

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle (s. u.) erhältlich.

Alle Rechte vorbehalten.

© **GfA-Press, Sankt Augustin, Schriftleitung: Prof. Dr. Rolf Ellegast**

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet:

- den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Kongressband oder Teile daraus in Print- und/oder Nonprint-Medien (Webseiten, Blog, Social Media) zu verbreiten.

Die Verantwortung für die Inhalte der Beiträge tragen alleine die jeweiligen Verfasser; die GfA haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

### **Geschäftsstelle der GfA**

Simone John, Tel.: +49 (0)30 1300-13003, Alte Heerstraße 111, D-53757 Sankt Augustin

[info@gesellschaft-fuer-arbeitswissenschaft.de](mailto:info@gesellschaft-fuer-arbeitswissenschaft.de) · [www.gesellschaft-fuer-arbeitswissenschaft.de](http://www.gesellschaft-fuer-arbeitswissenschaft.de)

### **Screen design und Umsetzung**

© 2024 fröse multimedia, Frank Fröse,

[office@internetkundenservice.de](mailto:office@internetkundenservice.de), [www.internetkundenservice.de](http://www.internetkundenservice.de)