Maximilian Eibl, Martin Gaedke. (Hrsg.): INFORMATIK 2017,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 2287

Auswirkung systeminduzierter Delays auf die menschliche Gedächtnisleistung in einem virtuellen agentenbasierten Trainingssetting

Maria Wirzberger [1], René Schmidt[2], Günter Daniel Rey[3], Wolfram Hardt[4]

Abstract: Verzögerungen in der Reaktion technischer Systeme können neben negativen Auswirkungen auf das systembezogene Nutzererleben auch Einbußen in der kognitiven Leistungsfähigkeit zur Folge haben. Die vorliegende Arbeit untersucht derartige Effekte anhand eines dialoggestützten Gedächtnistrainings mit einem virtuellen Agenten. Basierend auf der systemseitig kontrollierten Implementierung definierter Verzögerungszeiten zeigten sich in einem experimentellen Wizard-of-Oz-Setting neben ungünstigen Effekten längerer Verzögerungen auch vermittelnde Einflüsse nutzerseitiger Charakteristika auf die beobachtete Gedächtnisleistung. Ausgehend von diesen Befunden werden abschließend mögliche Optimierungen und Anknüpfungspunkte für weitere Forschungsarbeiten diskutiert.

Keywords: Systemdelay; Dialogsystem; Wizard-of-Oz; Virtueller Agent; Gedächtnistraining

1 Motivation

Bezogen auf die Mensch-zu-Mensch Interaktion wurde die Rolle verzögerter Reaktionen im Dialogablauf bereits untersucht, im Kontext technischer Systeme ist sie jedoch eher wenig beforscht. Bei der Interaktion mit Sprachdialogsystemen können solche Verzögerungen verringerte Akzeptanz und erhöhte Frustration zur Folge haben. So schätzten bei[SH10] die Teilnehmenden ein um 600ms schneller reagierendes Dialogsystem als höflicher, effizienter und transparenter ein. Im Gegensatz dazu wurde bei [Hi99] das Dialogsystem mit sehr kurzer Antwortzeit (M = 0.37s) schlechter bewertet. Motiviert durch die widersprüchliche Befundlage, adressierte das vorliegende Projekt Effekte verzögerter Systemantwort im Rahmen eines dialoggestützten Gedächtnistrainings mit einem virtuellen Agenten. Während bestehende Arbeiten lediglich die subjektive Systembewertung erfragten, stand hier der Zusammenhang zwischen systemseitiger Verzögerung und kognitiver Leistungsfähigkeit unter Berücksichtigung vermittelnder nutzerseitiger Einflussfaktoren im Vordergrund. Das resultierende Testsetting stützte sich inhaltlich auf die Vermittlung der Loci-Methode

[1] TU Chemnitz, Psychologie digitaler Lernmedien, maria.wirzberger@phil.tu-chemnitz.de
[2] TU Chemnitz, Technische Informatik, rene.schmidt@informatik.tu-chemnitz.de
[3] TU Chemnitz, Psychologie digitaler Lernmedien, guenter.daniel-rey@phil.tu-chemnitz.de
[4] TU Chemnitz, Technische Informatik, wolfram.hardt@informatik.tu-chemnitz.de

doi:10.18420/in2017_229

[DBC85], einer etablierten Gedächtnistechnik, die auf die Verbesserung der Merkleistung durch die räumliche Verortung zu merkender Inhalte zielt. Wie für die Entwicklung und prototypische Erprobung von Sprachdialogsystemen charakteristisch, folgte die durchgeführte Nutzerstudie methodisch der Wizard-of-Oz Technik [Ke84], welche systemintendierte Interaktionshandlungen versuchsleitergesteuert simuliert.

2 Methodik

2.1 Experimentelle Nutzerstudie

Stichprobe: Die Prüfung des Systems erfolgte im Rahmen einer experimentellen Nutzerstudie mit $N = 35$ Versuchspersonen ($M = 24.0$ Jahre, $SD = 4.5$, Range: 19-38, $n = 19$ weiblich). Die Teilnehmenden verfügten mehrheitlich über deutsche Sprachkenntnisse auf Muttersprachniveau ($n = 29$), die übrigen nutzten die Sprache seit $M = 7.7$ Jahren ($SD = 4.5$) aktiv. In Bezug auf die vorherige Teilnahme an Gedächtnistrainings bestanden überwiegend keine Vorerfahrungen ($n = 33$) bzw. diese beschränkten sich im Falle der Teilnahme auf maximal zwei Gelegenheiten.

Studiendesign: Die Untersuchung des Einflusses systeminduzierter Delays auf die menschliche Gedächtnisleistung erfolgte über eine kontinuierliche experimentelle Variation der Delaylänge als unabhängiger Variable aus einem gleichverteilten Intervall zwischen 3000 und 8000 ms in Schritten von jeweils 500 ms. Als Delay wird dabei im Folgenden die Systemantwortzeit definiert als diejenige Zeitspanne, welche das System vom Ende der Aussage einer Versuchsperson bis zum Starten der daraufhin generierten Systemantwort benötigt. Die abhängige Variable der Gedächtnisleistung wurde ermittelt anhand des Anteils korrekt wiedergegebener Begriffe - ohne Berücksichtigung der Reihenfolge - in drei Testdurchgängen mit fünf, sieben und neun deutschen Substantiven [Sc11]. Zusätzlich wurde der Anteil korrekt wiedergegebener Begriffe während der drei Trainingsdurchgänge mit fünf, sieben und neun deutschen Substantiven aufgezeichnet. Ein Durchgang mit sieben Substantiven vor den Trainingsdurchgängen diente als Baseline zur Bestimmung der individuellen Merkfähigkeit.

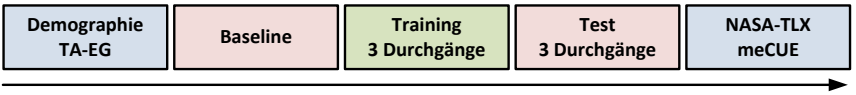


Abb. 1: Schematischer Ablauf der Versuchstermine. Blau = Fragebögen; Rot = Wortlisten ohne Agent; Grün = Wortlisten mit Agent.

Material & Ablauf: Jeder Versuchstermin begann mit der schriftlichen Aufklärung der Versuchsperson zum Inhalt und Ablauf der Datenerhebung und der Anonymität der Datenspeicherung und -analyse. Anschließend wurden zunächst demographische Kerndaten sowie die subjektiv eingeschätzte Technikaffinität mit Hilfe des TA-EG[5] [Ka09] erfasst.

[5] Subskalen: Begeisterung, Kompetenz, wahrgenommene positive Technikfolgen und wahrgenommene negative Technikfolgen

Wie in Abbildung 1 dargestellt, folgte darauf die Erhebung der individuellen Baseline der Merkfähigkeit. Dazu wurden nach entsprechender Anweisung sieben Substantive für jeweils 5000 ms mit einem Abstand von 200 ms schriftlich auf dem Bildschirm präsentiert und nach einem Punktsignal in der Bildschirmmitte sprachlich wiedergegeben. Abschließend erhielt die Versuchsperson eine neutrale Rückmeldung über ihre Leistung. Während der darauf folgenden, als Sprachdialog angelegten Trainingssequenz, wurde ein virtueller Agent mit einer Wohnzimmerszenerie genutzt, um die Versuchsperson über drei Durchgänge hinweg schrittweise in die Anwendung der Loci-Methode einzuführen. Als Agent wurde dabei der in Abbildung 2 rechts dargestellte Fuchs verwendet, welcher in der kommerziellen Mobilanwendung FaceRig[6] enthalten ist. Mit dieser lassen sich Videos anhand kamerabasierter Gesichtserkennung erstellen. Die Entscheidung für diesen spezifischen Agenten gründet in der technisch ausgereifteren Animationsfähigkeit im Vergleich zu alternativen, in der Anwendung verfügbaren Gestalten. Die ersten beiden Durchgänge basierten auf dem in Abbildung 2 links gezeigten gegenständlichen Bild eines Wohnzimmers[7], in welchem gezielt relevante Gegenstände hervorgehoben waren. Im ersten Durchgang erfolgte eine konkrete Anleitung des Agenten zum Einsatz der Gegenstände, z.B. "*Legen Sie das eingeblendete Wort an der Lampe ab*.", während die Zuordnung der zu merkenden Begriffe zu den Gegenständen im zweiten Durchgang durch die Versuchsperson selbst gesteuert wurde. Darauf aufbauend nutzte der dritte Durchgang eine nur noch schemenhaft erkennbare Abbildung eines Wohnzimmers mit der Instruktion, sich das eigene oder ein gut bekanntes Wohnzimmer vorzustellen und die zu merkenden Begriffe nacheinander an den dort vorhandenen Gegenständen abzulegen. Ziel dieses Transfers war die Ver-

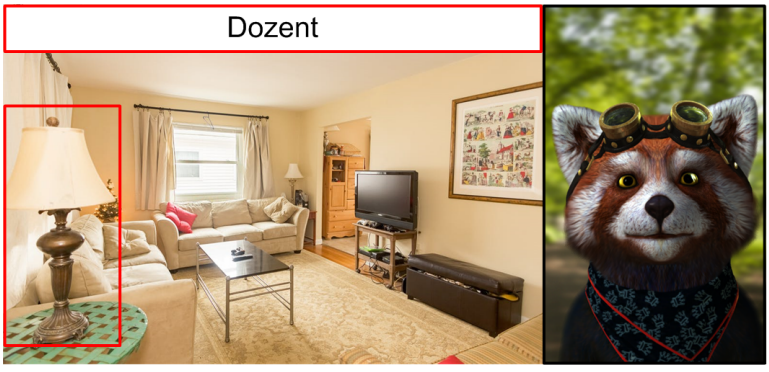


Abb. 2: Schematische Darstellung des Trainingsszenarios mit Wohnzimmer und Agent im ersten und zweiten Durchgang.

knüpfung der Gedächtnistechnik mit der eigenen Alltagswelt. Die zu merkenden Wörter wurden jeweils über dem Wohnzimmerbild eingeblendet und von der Versuchsperson im Anschluss verbal wiedergegeben. Der Agent erschien rechts davon während bzw. in der Bildschirmmitte zwischen den Trainingssequenzen. Der Trainingsaufbau folgte damit dem

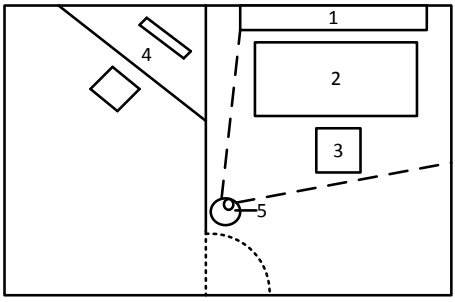
[6] Verfügbar unter https://facerig.com/; Bildverwendung mit freundlicher Genehmigung des Unternehmens

[7] Wohnzimmerbild lizenzfrei verfügbar unter https://static.pexels.com/photos/77931/pexels-photo-77931.jpeg

Prinzip der abnehmenden instruktionalen Unterstützung (Guidance fading; [SAK11]) um eine stabilere Implementierung der Gedächtnistechnik zu ermöglichen. Charakteristisch für die Wizard-of-Oz Methodik, erfolgte sämtliche sprachliche Interaktion des Systems mit der Versuchsperson durch die Versuchsleitung, welche per Knopfdruck aus einem Set vordefinierter Antwortpfade die jeweils passende Dialogoption auswählte. Während der auf die Trainingsdurchgänge folgenden Testdurchgänge, gestaltete sich die Aufgabe der Versuchpersonen ähnlich zum dritten Durchgang des Trainings, allerdings ohne Präsenz des Agenten. Abschließend wurden die subjektiv eingeschätzte mentale Beanspruchung anhand des NASA-TLX[8] [HS88] erhoben, bei welchem in Anlehnung an [Pf91] auf den Einsatz der Paarvergleiche verzichtet wurde. Zusätzlich wurde die individuelle Bewertung nutzungsbezogener und emotionaler Qualitäten des Systems anhand des meCUE[9] [MR] erfasst. Den finalen Schritt jedes Versuchstermins bildete die Aufklärung der Versuchsperson über das spezifische Setting sowie die durch die Versuchsleitung gesteuerte Systeminteraktion.

2.2 Technisches Setup

Die Umsetzung des beschriebenen Szenarios ging mit spezifischen Voraussetzungen an die Lokalität sowie die softwaretechnische Umsetzung einher. Für die Lokalität wurden zwei benachbarte Räume gewählt, welche durch eine Tür miteinander verbunden waren (vgl. Abbildung 3). Hardwareseitig wurde als Visualisierungsmedium des agentenbasierten Trainingssystems ein Samsung Full HD TV gewählt mit einer Auflösung von 1080p, welcher mit einem Standard Windows 7 PC über HDMI verbunden wurde. Zur Erfassung der auditiven Signale der Versuchspersonen wurde eine Logitech QuickCam Pro 5000 mit 480p Auflösung genutzt, deren Audiosignale direkt auf ein Headset weitergeleitet wurden. Das Kamerabild wurde der Versuchsleitung zur visuellen Kontrolle des Versuchs bereit gestellt. Dies stellte sicher, dass die Versuchsleitung sowohl die Versuchsperson als auch den aktuellen Zustand des Agenten überwachen konnte, und unterstützte auf diese Weise die korrekte Reaktion auf die jeweiligen Aussagen.



1) Samsung Full HD TV
2) Tisch
3) Stuhl für Probanden
4) Arbeitsplatz des Versuchsleiters
5) Logitech QuickCam Pro 5000

Abb. 3: Schematische Darstellung des Versuchsaufbaus.

[8] Subskalen: Geistige Anforderungen, Körperliche Anforderungen, Zeitliche Anforderungen, Aufgabenausführung, Anstrengung und Frustration

[9] Subskalen: Nützlichkeit, Benutzbarkeit, visuelle Ästhetik, Status, Bindung, positive Emotionen, negative Emotionen, Nutzungsintention und Produktloyalität

Der Hauptfokus des Systems bestand darin, die intendierten Delays der Systemantwortzeiten durch eine speziell entwickelte Softwarelösung sicherzustellen. Grundidee der Implementierung stellte dabei die Repräsentation des Agenten in Form von Videos dar, welche mit beliebig langer Verzögerung aneinander gereiht werden konnten. Die Auswahl des nachfolgenden Videos wurde durch die Versuchsleitung über visuelle Schaltflächen am Bildschirm gesteuert. Zur exakten Einhaltung der Verzögerung diente ein Amplituden basierter Voice-Activation-Detector (VAD), welcher die verbalen Reaktionen der Versuchsperson überwachte und die Grundlage für die Zeitmessung bildete. Die genaue Beschreibung und Umsetzung des Softwareverhaltens basierte auf einem Automatengraphen, welcher in Abbildung 4 dargestellt ist. Dieser startete in einem definierten Startzustand, aus welchem heraus ein Zustandswechsel nur durch die Versuchsleitung erfolgen konnte, anhand der Aktivierung einer Schaltfläche im Bildschirmmenü. Mit diesem initialen Zustandswechsel wurde auch die Länge des Delays festgelegt, die anschließend für alle Systeminteraktionen verwendet wurde. Im nachfolgenden Zustand wartete das System auf die Ausgabe des VAD und wechselte, basierend auf der Detektion einer sprachlichen Interaktion der Versuchsperson, in den Zwischenzustand "WAIT FOR STOP" . Aus diesem Zwischenzustand heraus wurde nach Beendigung der Aussage der Versuchsperson in den darauf folgenden Wartezustand gewechselt und der Start der Delaymessung angestoßen. Während des Warte-

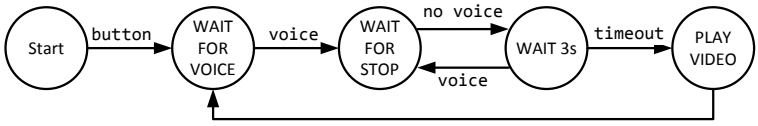


Abb. 4: Interner Ablauf der Softwareimplementierung zur Sicherstellung der Delaylänge.

zustands stand der Versuchsleitung die minimale Verzögerungslänge von drei Sekunden zur Verfügung, um die nachfolgende Aussage des Agenten aus den Möglichkeiten des logisch folgenden Dialogverlaufs auszuwählen. Erfolgte die Auswahl nicht innerhalb der verfügbaren Zeitspanne, wurde die standardisierte Antwort "*Ich habe Sie leider nicht verstanden.*" ausgegeben. Auf diese Weise konnte die Einhaltung der korrekten Systemantwortzeit auch im Falle fehlender Reaktionen der Versuchsleitung sichergestellt werden. Für den Fall einer Sprechpause der Versuchsperson wurde ein Wechsel in den vorhergehenden Zustand "WAIT FOR STOP" ermöglicht, um den unbeabsichtigten Start einer Delaymessung zu unterbrechen. Nachdem die Versuchsleitung eine Aussage gewählt hatte, musste vor dem Abspielen des Folgevideos zunächst die verbleibende Verzögerungszeit verstreichen. Nach erfolgreicher Wiedergabe des Videos wechselte das System wieder in den Zustand "WAIT FOR VOICE" und startete den beschriebenen Prozessablauf erneut.

3 Ergebnisse

Die inferenzstatistische Analyse der experimentell erhobenen Daten erfolgte anhand eines regressionsanalytischen Pfadmodells. Aufgrund technischer Probleme und daraus resultierender fehlender Werte mussten dabei fünf Datensätze ausgeschlossen werden. Wie aus

Abbildung 5 ersichtlich, umfasste das Pfadmodell neben dem Zusammenhang zwischen der Testleistung als abhängiger und der Länge des Delays als unabhängiger Variable eine Reihe nutzerbezogener Faktoren, welche diesen Zusammenhang potentiell beeinflussen. Konkret waren dies die individuelle Merkfähigkeit (Baseline), das Ausmaß an Technikaffinität, das Alter sowie die vorhandenen Sprachfertigkeiten (Deutsch als Muttersprache vs. Fremdsprache). Die genannten Variablen wurden als direkte Einflüssen auf die Testleistung wie auch als Moderatoreffekte auf den Delay in das Modell aufgenommen. Zusätzlich wurde der Einfluss des Delays während des Trainings und dessen Auswirkung auf die Testleistung durch einen Mediationseffekt berückichtigt.

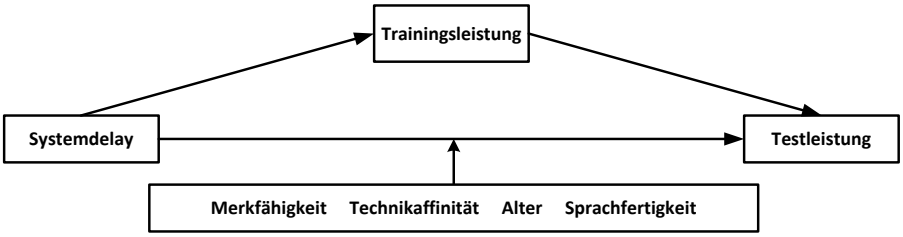


Abb. 5: Schematisches Modell des postulierten Zusammenhangs zwischen den erhobenen Variablen.

In Tabelle 1 sind die direkten Einflüsse des Delays sowie der nutzerbezogenen Faktoren auf die Testleistung in Form standardisierter β- Gewichte zusammengefasst. Dabei zeigt sich, dass die Testleistung mit steigendem Delay signifikant sinkt, während sich eine höhere Leistung während der Trainingsphase sowie eine erhöhte Technikaffinität hinsichtlich Begeisterung, wahrgenommener positiver und negativer Technikfolgen signifikant positiv auf die Testleistung auswirken.

Tab. 1: Direkte Effekte nutzer- und systembezogener Prädiktoren auf die Testleistung

Delay	Baseline	Training	Alter	Sprache	Begeisterung	Kompetenz	Positiv	Negativ
-.202*	.129	.354***	.258	-.121	.260*	-.024	.377*	.355***

*p < .05 **p < .01 ***p < .001. Sprache binär kodiert mit 0 = Muttersprache, 1 = Fremdsprache

Zusätzlich zeigen sich in Tabelle 2 signifikante positive Einflüsse der individuellen Merkfähigkeit und des Alters sowie signifikante negative Einflüsse der Sprachfertigkeit und der subjektiven Kompetenz im Umgang mit Technik auf den Zusammenhang zwischen Delaylänge und Testleistung. Im Detail deutet dies darauf hin, dass sich höhere Ausmaße individueller Merkfähigkeit sowie ein höheres Lebensalter kompensatorisch auf eine erhöhte Delaylänge auswirken. Des Weiteren scheinen längere Delays insbesondere die Leistung von Personen ohne deutsche Muttersprache und mit höherer selbsteingeschätzter Technikkompetenz zu beeinträchtigen. Generell deutet die Höhe der standardisierten β-Gewichte an dieser Stelle darauf hin, dass sich die kombinierte Betrachtung der nutzerseitigen Einflussfaktoren und des Delays deutlich stärker auf die Testleistung auswirken, als dies für die Delaylänge allein der Fall ist. Ein signifikanter Einfluss des Delays auf die Trainingsleistung zeigte

Tab. 2: Effekte nutzerbezogener Prädiktoren auf den Zusammenhang zwischen Delaylänge und Testleistung

Baseline	Alter	Sprache	Begeisterung	Kompetenz	Positiv	Negativ
.520***	.551**	-.466***	.252	-.363**	-.002	-.180

*p < .05 **p < .01 ***p < .001. Sprache binär kodiert mit 0 = Muttersprache, 1 = Fremdsprache

sich nicht und auch der angenommene Mediationseffekt auf die Testleistung erreichte keine Signifikanz. Insgesamt erzielte das Pfadmodell eine Varianzaufklärung von 80%. Darüber hinaus wurde anhand bivariater Korrelationen geprüft, ob sich die Länge des Delays signifikant auf die wahrgenommene Beanspruchung durch das Aufgabensetting (Skalen des NASA-TLX) sowie die nutzungsbezogene und emotionale Bewertung des System (Skalen des meCUE) auswirkt. Hier konnten jedoch ebenfalls keine signifikanten Zusammenhänge festgestellt werden.

4 Diskussion

Das Ziel des vorliegenden Settings bestand darin, den Einfluss der Länge systeminduzierter Delays auf die Gedächtnisleistung im Rahmen eines dialoggestützten Gedächtnistrainings mit einem virtuellen Agenten zu untersuchen. Trotz der kleinen Stichprobe zeigten sich, neben einer signifikanten Leistungsbeeinträchtigung durch längere Delays, signifikante Einflüsse system- und nutzerimmanenter Faktoren auf die Wirkung der Delaylänge, welche vielfältige Anknüpfungspunkte für weitere Forschung bieten. So lässt sich ausgehend vom Faktor Alter ein potentieller Forschungsfokus auf ein breiteres Altersspektrum definieren, der u.a. durch den Einbezug einer Seniorenstichprobe umgesetzt werden könnte. Hier bestünde eine mögliche Fragestellung beispielsweise darin, ob sich die scheinbar kompensatorische Wirkung steigenden Alters bei verlängerten Systemdelays auch in höheren Lebensjahren fortsetzt. Einen ebenfalls interessanten Befund bildet die scheinbar negative Wirkung höherer selbstberichteter Technikkompetenz im Zuge verlängerter Delays. Diese könnte möglicherweise daraus resultieren, dass Personen mit höherer Technikkompetenz präziser einschätzen können, welchen Delay aktuelle Systeme tatsächlich benötigen. Fällt dieser Delay nun höher aus als erwartet, ist die daraus resultierende Erwartungsdiskrepanz und Frustration deutlich höher als im Falle unspezifischer Erwartungen.

Neben einer größeren Stichprobe, mit welcher sich die abgeleiteten Überlegungen valider prüfen ließen, sollte eine Folgeuntersuchung gezielt Konfundierungen im experimentellen Setting adressieren. Konkret betrifft dies die Notwendigkeit geteilter Aufmerksamkeit zwischen auditiver und visueller Modalität, welche durch die visuelle Präsentation der Substantive am Bildschirm parallel zur sprachlichen Interaktion mit dem Agenten entstand. Ein verbessertes Setting würde derartige Effekte durch die Reduktion auf eine rein sprachliche Interaktion ohne visuelle Präsentation der Substantive vermeiden. Zusätzlich bestand ein weiterer möglicher Einflussfaktor in der Präsenz des verschwommenen Wohnzimmerbildes während der Testdurchgänge, welches als zusätzlicher, nicht-intendierter

retrieval cue [TO68] gewirkt haben könnte. Abschließend betrachtet bildet die vorgestellte Studie einen vielversprechenden Ansatzpunkt für die Erschließung des zukunftsweisenden Forschungsfeldes nutzeradaptiver Sprachdialogsysteme.

Danksagung

Die vorgestellte Arbeit wurde im Rahmen des GRK 1780/1 mit freundlicher Unterstützung durch die Deutsche Forschungsgemeinschaft (DFG) realisiert.

Literaturverzeichnis

[DBC85] De Beni, Rossana; Cornoldi, Cesare: Effects of the mnemotechnique of loci in the memorization of concrete words. Acta Psychologica, 60:11–24, 1985.

[Hi99] Hirasawa, Jun-ichi; Nakano, Mikio; Kawabata, Takeshi; Aikawa, Kiyoaki: Effects of system barge-in responses on user impressions. In: Sixth European Conference on Speech Communication and Technology, EUROSPEECH. 1999.

[HS88] Hart, Sandra G.; Staveland, Lowell E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52:139–183, 1988.

[Ka09] Karrer, Katja; Glaser, Charlotte; Clemens, Caroline; Bruder, Carmen: Technikaffinität erfassen–der Fragebogen TA-EG. In (Lichtenstein, Antje; Stößel, Christian; Clemens, Caroline, Hrsg.): Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme. VDI Verlag, Düsseldorf, S. 196–201, 2009.

[Ke84] Kelley, John F.: An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Information Systems (TOIS), 2:26–41, 1984.

[MR] Minge, Michael; Riedel, Laura: meCUE-Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In (Boll, Susanne; Maaß, Susanne; Malaka, Rainer, Hrsg.): Mensch und Computer 2013: Interaktive Vielfalt. Oldenbourg Verlag, München, S. 89–98.

[Pf91] Pfendler, C: Vergleichende Bewertung der NASA-TLX-Skala und der ZEIS-Skala bei der Erfassung von Lernprozessen. Forschungsins. für Anthropotechnik, Forschungsges. für Angewandte Naturwissenschaften e.V., 1991.

[SAK11] Sweller, John; Ayres, Paul; Kalyuga, Slava: The guidance fading effect. In: Cognitive Load Theory, S. 171–182. Springer, New York, 2011.

[Sc11] Schrauf, Judith: Vom Konkreten im Abstrakten. Eine kognitionslinguistische Analyse zu Konkreta und Abstrakta. Dissertation, 2011.

[SH10] Skantze, Gabriel; Hjalmarsson, Anna: Towards incremental speech generation in dialogue systems. In: Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics, S. 1–8, 2010.

[TO68] Tulving, Endel; Osler, Shirley: Effectiveness of retrieval cues in memory for words. Journal of Experimental Psychology, 77:593–601, 1968.