

Kognitive Grundlagen der multimodalen Mensch-Maschine-Interaktion

Anja B. Naumann, Ina Wechsung & Robert Schleicher

Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany,
Email: anja.naumann@telekom.de

Einleitung

Multimodalen Benutzungsschnittstellen werden gegenüber unimodalen Schnittstellen Vorteile hinsichtlich der Robustheit, der Natürlichkeit und der Flexibilität zugeschrieben. Begründet wird diese Überlegenheit durch die Tatsache, dass zwischenmenschliche Interaktion typischerweise auch nicht auf eine Modalität beschränkt ist, sondern sich multipler Sinnesmodalitäten bedient [10]. Jedoch zeigte sich [4], dass multimodale Schnittstellen diesen positiven Erwartungen nicht immer gerecht werden und unimodalen Schnittstellen nicht zwangsläufig überlegen sind. So kann redundante Informationsdarstellung über mehrere Wahrnehmungskanäle zwar zur Fehlervermeidung beitragen [7], andererseits wird jedoch vermutet, dass dem Benutzer zusätzliche kognitive Ressourcen, z. B. bei der Entscheidung für eine der angebotenen Modalitäten, abverlangt werden, was zur Überforderung des Nutzers führen kann [9,12]. Erklärungsansätze für diese uneindeutige Befundlage liefert die psychologische Forschung.

Speicherung, Enkodierung und Repräsentation multimodaler Informationen

Im Mittelpunkt kognitionspsychologischer Betrachtungen von Multimodalität stehen die Wahrnehmung und die Verarbeitung von Informationen. Wahrnehmung im weiteren Sinne bezeichnet dabei den Prozess der Aufnahme und Interpretation von Information durch den Menschen [13]. Während des Wahrnehmungsprozesses werden Sinnesempfindungen in mentale Repräsentationen umgewandelt. Die Aufnahme erfolgt über die menschlichen Sinnesorgane, die mit verschiedenen Sinneskanälen, den Sinnesmodalitäten assoziiert sind. Nachdem die sensorischen Umgebungsvariablen von den Sinnesorganen aufgenommen wurden, gelangen sie, so der aktuelle Stand der Forschung, zunächst in das Arbeitsgedächtnis [2]. Mit dem Arbeitsgedächtnismodell von Baddeley & Hitch [2] wird dabei die Annahme vertreten, dass bei der Nutzung unterschiedlicher Modalitäten zum Teil auch auf unterschiedliche Ressourcen zurückgegriffen wird. Dabei umfasst das Arbeitsgedächtnis drei Komponenten: die zentrale Exekutive, den räumlich-visuellen Notizblock und die artikulatorische/phonologische Schleife (s. Abbildung 1). Die phonologische Schleife sowie der räumlich-visuelle Notizblock übernehmen die kurzfristige Speicherung von verbalen bzw. visuellen Informationen. Die zentrale Exekutive ist dabei für die Kontrolle des Arbeitsgedächtnisses zuständig: Denkprozesse und Handlungen werden überwacht und gegebenenfalls korrigiert.



Abbildung 1: Aufbau des Arbeitsgedächtnis nach Baddeley & Hitch [2]

Auch hinsichtlich der Langzeitspeicherung erfolgt die Umwandlung der Informationen in mentale Repräsentation zu einem großen Teil, jedoch nicht ausschließlich modalitätsspezifisch [5]. Dies bedeutet, dass verbale und visuelle Informationen vermutlich auf unterschiedliche Art und Weise verarbeitet werden [1]. Die Theorie der dualen Kodierung von Paivio [5] postuliert dabei zwei weitestgehend unabhängige kognitive Systeme. Das verbale System und das imaginale, nicht-verbale System. Im verbalen System werden sprachliche Informationen verarbeitet, im imaginalen System dagegen visuell-räumliche. Befunde aus der Hirnforschung bestätigen diese Annahme [6]: Es wurde gezeigt, dass je nachdem welcher Art (verbal vs. räumlich) die zu verarbeitende Information ist, unterschiedliche Hirnareale aktiv sind. Dennoch sind diese zwei verschiedenen Systeme miteinander vernetzt. Dies erklärt, warum multimodale Darbietungen, bspw. verbal-auditive Informationen zusammen mit bildlich-visuellen Informationen, unimodalen Darbietungen überlegen sein können. Die doppelte Kodierung führt dabei zu einer Erhöhung der Behaltensleistung. Widersprechen sich jedoch visuelle und auditive Informationen, kann diese Vernetzung auch negative Folgen haben. Ein Beispiel ist der McGurk-Effekt, bei dem bimodale Wahrnehmung zu einer Wahrnehmungstäuschung führt [3]. Ebenso zeigt das Phänomen der „Ton-Bild-Schere“, wie inkohärente multimodale Informationen zur Überforderung des Rezipienten führen können [11]. Demnach ist Multimodalität nicht per se von Vorteil, sondern nur unter der Voraussetzung von Kohärenz (semantischer Zusammenhang) und Kontiguität (zeitlich oder räumliches Zusammentreffen) der verschiedenen Informationen [8].

Die Theorie multipler Ressourcen

Eine Theorie, die viele dieser aktuellen Befunde integriert, bietet Wickens [12] an: In dem Modell der multiplen Ressourcen werden drei Verarbeitungsphasen (Enkodierung, Kognition/zentrale Verarbeitung, Reaktion), zwei Antwortcodes (manuell vs. vokal) zwei Wahrnehmungsmodalitäten (visuell vs. auditorisch), und zwei Eingabecodes (räumlich vs. verbal) unterschieden. Zusätzlich werden zwei visuelle Kanäle (fokal vs. ambient) differenziert. Greifen die Aufgaben auf dieselben

Ressourcen zu, folgt eine schlechtere Performanz. Bei unterschiedlichen Ressourcen sollte es nicht zu Leistungseinbußen kommen. Am Beispiel würde dies bedeuten, dass die Sehleistung nicht durch gleichzeitiges Sprechen verschlechtert wird. Abbildung 2 veranschaulicht dieses Modell. Prinzipiell gilt: Je weiter die Ressourcen von einander entfernt sind, umso besser ist die Performanz bei gleichzeitigem Zugriff.

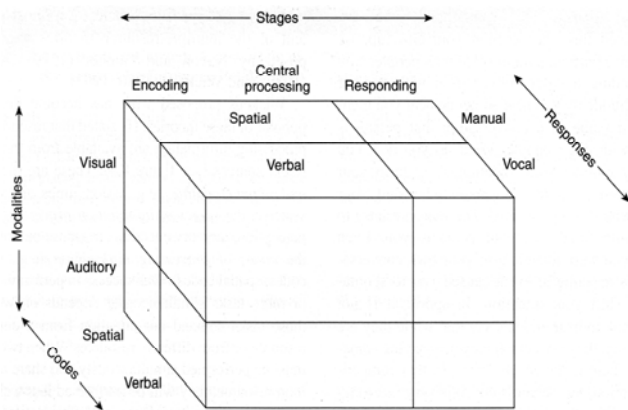


Abbildung 2: Modell der multiplen Ressourcen nach Wickens [12].

Zu Interferenzen kommt es demnach nur, wenn die Gesamtkapazität der Ressourcen überschritten wird oder zu stark auf eine Ressource zurückgegriffen wird. Dies kann jedoch auch durch unterschiedliche Modalitäten erfolgen: Zum Beispiel dann, wenn verbale Informationen simultan auditiv und visuell präsentiert werden. Sind die Informationen unterschiedlicher Art, treten Interferenzen auf, da verbale Informationen unabhängig von der Darbietung im Wesentlichen dieselben kognitiven Prozesse nach sich ziehen und demnach dieselben Ressourcen beanspruchen. So ist es bspw. kaum möglich, ein Hörbuch zu hören und gleichzeitig einen anderen Text zu lesen. Bei simultaner Präsentation identischer Information entsteht ein Synchronisierungsproblem: Die Informationsaufnahme erfolgt beim Lesen schneller als beim Hören [8].

Gestaltungsempfehlungen für multimodale Schnittstellen

Angesichts der dargestellten Befunde bleibt zusammenfassend festzuhalten, dass multimodale Benutzungsschnittstellen ein großes Potential bergen, das sich jedoch nur dann ausschöpfen lässt, wenn bestimmte Voraussetzungen erfüllt werden. Prinzipiell lassen sich folgende Empfehlungen hinsichtlich des Einsatzes von Multimodalität ableiten:

- Kohärenz und Kontiguität der Informationen müssen gewährleistet sein.
- Die einzelnen Modalitäten sollten so gewählt werden, dass möglichst wenig gemeinsame Ressourcen beansprucht werden.
- Der Einsatz und die Kombination der verschiedenen Modalitäten muss sowohl auf die

Aufgabe als auch auf den Kontext abgestimmt werden.

Weiterhin ist anzumerken, dass Multimodalität sich auch dann nachteilig auswirken kann, wenn der Nutzer durch zu viele mögliche Handlungsalternativen überfordert wird.

Literatur

- [1.] Anderson, J. R.: Kognitive Psychologie. Spektrum Akademischer Verlag, Heidelberg, 1996.
- [2.] Baddeley, A. D., & Hitch, G. J.: Working memory. In G. Bower (Hrsg.): The psychology of learning and motivation. Academic Press, New York (1974), 47-89
- [3.] McGurk, H. & MacDonald, J.: Hearing lips and seeing voices. Nature Vol. 264 (1976), 746-748
- [4.] Oviatt, S. L.: Ten myths of multimodal interaction. Communications of the ACM, Vol. 42 (1999), 576-583.
- [5.] Paivio, A.: Imagery and verbal processes. Holt, Rinehart, and Winston, New York, 1971
- [6.] Roland, P. E., & Friberg, L.: Localization of cortical areas activated by thinking. Journal of Neurophysiology, 53 (1985), 1219-1243
- [7.] Sarter, N.B.: Multisensory Feedback in Support of Human-Automation Coordination In Highly Dynamic Event-Driven Worlds. Paper presented at the Third Automation Technology and Human Performance Conference, 1999
- [8.] Schnotz, W., Bannert, M. & Seufert, T.: Towards an integrative view of text and picture comprehension: Visualization effects on the construction of mental models. In J. Otero, A. Graesser & J. A. Leon (Hrsg.): The Psychology of Science Text Comprehension. Erlbaum, Mahwah (2002), 385-416
- [9.] Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoît, C., Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defée, I., Münch, S., Hartung, K. & Blauert, J.: A Taxonomy of Multimodal Interaction in the Human Information Processing System. A Report of the ESPRIT Project 8579 MIAMI. NICI, Nijmegen, 1995.
- [10.] Weidenmann, B.: Multicodierung und Multimodalität im Lernprozess. In L.J. Issing & P. Klimsa (Hrsg.): Information und Lernen mit Multimedia, Beltz, Weinheim, (2002), 44-62.
- [11.] Wember, B.: Wie informiert das Fernsehen? Ein Indizienbeweis. List, München, 1976
- [12.] Wickens, C.D.: Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3 (2002), 159-177.
- [13.] Zimbardo, P.: Psychologie. Springer, Berlin, 1995