Evaluating Complementary Interfaces as a Design-Informing Activity

Doctoral thesis

for obtaining the academic degree

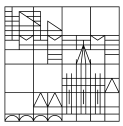
Doctor of Natural Sciences (Dr. rer. nat.)

submitted by

Johannes Zagermann

at the

Universität Konstanz



Faculty of Sciences

Department of Computer and Information Science

Konstanz, 2024

Date of the oral examination: 18.06.2024

1. Reviewer: Prof. Dr. Harald Reiterer

2. Reviewer: Prof. Dr. Albrecht Schmidt

3. Reviewer: Prof. Dr. Daniel Keim

Abstract

An increasing number of interactive devices is finding their way into our everyday lives. People own several devices and find creative ways to combine them. Research has found interest in these Ubiquitous Computing environments and pushed the boundaries of technological possibilities by sophistically blending devices together – following a "the more the merrier" approach – a common theme in Human-Computer Interaction research that is often better at proposing new technologies than at validating them – leading to a fragmented research landscape and a need for evaluation.

This dissertation addresses this need for evaluation by studying Complementary Interfaces as representatives for Ubiquitous Computing environments using controlled lab-based experiments that serve as design-informing activities – guided by the general research question of how to design and evaluate Complementary Interfaces.

Building on the theoretical foundations of Mark Weiser's vision of the computer for the 21st century and related research on Cross-Device Interaction, Hybrid User Interfaces, and Multimodal Interaction, the concept of Complementary Interfaces is defined as meaningful combinations of interfaces that support users in their current task at hand. As part of this dissertation, three use cases of Complementary Interfaces are empirically studied using experimental methods. The selection of these use cases allows us to investigate the influences of variants of output modalities, input techniques, and their interaction – by focusing on different knowledge activities as well as combinations and dependencies of components. Specifically, the presented Complementary Interfaces allow us to study the role of a shared interactive surface on collaborative activities, the influences of a set of meaningful combinations of input and output modalities on spatial memory, and the effects of interaction techniques on the utilization of components.

The experimental designs for each Complementary Interface purposefully integrate typical components for experiments in Human-Computer Interaction to frame the empirical research, which allows the use of experiments as design-informing activities for future research and design of Complementary Interfaces.

As prior research has identified the influences of handling multiple devices and modalities on cognitive workload, a spotlight is placed on current and past measurements of cognitive workload – a potential dependent variable in experimental settings for Complementary Interfaces.

This dissertation closes with an extension of the concept of Complementary Interfaces towards new dimensions by showcasing how the presented ideas for the design

and evaluation can be applied to meaningful combinations of heterogenous devices and modalities – eventually opening up research opportunities for highly dynamic collaborative Complementary Interfaces – emphasizing the need to study the utility, relevance, and role of each piece of the puzzle to eventually create a *symbiosis of interfaces*.

Kurzfassung

Immer mehr interaktive Geräte halten Einzug in unseren Alltag. Dabei nutzen wir Smartphones, Tablets und Co. einzeln oder auch auf kreative Art und Weise kombiniert. Diese Kombinationen sind auch Gegenstand aktueller Forschungsarbeiten im Bereich Mensch-Computer-Interaktion. Hier werden zum Beispiel durch ausgeklügelte Techniken mehrere Geräte nahtlos miteinander verbunden, um dabei die Grenzen des technischen Möglichen auszuloten – getreu dem Motto "je mehr, desto besser". Dabei wird allerdings oft die Evaluation dieser interaktiven Systeme vernachlässigt, was zu einer eher fragmentierten Forschungslandschaft führt, wodurch die Notwendigkeit an empirischen Evaluationen steigt.

Diese Dissertation befasst sich mit dieser Notwendigkeit an Evaluationen, indem sie Complementary Interfaces, stellvertretend für Ubiquitous Computing Umgebungen, durch kontrollierte laborbasierte Experimente untersucht, um dadurch Einsichten für das weiterführende Design und die Evaluation dieser interaktiven Systeme zu erzielen. Dabei stellt die allgemeine Forschungsfrage, wie Complementary Interfaces designt und evaluiert werden können, im Vordergrund.

Aufbauend auf den theoretischen Grundlagen von Mark Weisers Vision des Computers für das 21. Jahrhundert und verwandten Arbeiten zu Cross-Device Interaction, Hybrid User Interfaces und Multimodal Interaction wird das Konzept zu Complementary Interfaces definiert: Sinngebende Kombinationen von Eingabe- und Ausgabeschnittstellen, die dazu beitragen, dass Benutzer in ihrer Arbeit unterstützt werden. Im Rahmen dieser Dissertation werden dafür drei Anwendungsfälle von Complementary Interfaces empirisch durch Experimente untersucht. Die Auswahl dieser Anwendungsfälle ermöglicht es, die Einflüsse von unterschiedlichen Varianten von a) Ausgabemodalitäten, b) Eingabetechniken sowie c) deren Interaktion zu untersuchen. Dabei werden verschiedene Tätigkeiten im Bereich Wissensarbeit, Kombinationen von interaktiven Komponenten und deren Abhängigkeiten in Betracht gezogen. Im Einzelnen erlauben die vorgestellten Complementary Interfaces Untersuchungen zur Rolle von geteilten interaktiven Oberflächen für kollaborative Aktivitäten, den Einflüssen einer Reihe von sinngebenden Kombinationen von Eingabe- und Ausgabemodalitäten auf das räumliche Erinnerungsvermögen, sowie die Auswirkungen von Interaktionstechniken auf die Nutzung interaktiver Komponenten. Die einzelnen Benutzerstudien setzen gezielt verschiedene typische Komponenten für Experimente in der Mensch-Computer-Interaktionen ein. Dabei fungieren die jeweiligen Experimente als sinngebende Methoden für weiterführende Arbeiten für Design und Evaluation von Complementary Interfaces.

Da frühere Forschungsarbeiten gezeigt haben, dass die gleichzeitige Nutzung von mehreren Geräten und Modalitäten zu einer erhöhten kognitiven Arbeitsbelastung führen können, wird ein Schwerpunkt auf aktuelle und frühere Messmethoden der kognitiven Arbeitsbelastung gelegt – eine mögliche abhängige Variable für Studiendesigns zur Untersuchung von Complementary Interfaces.

Die Dissertation bietet zudem einen Ausblick auf mögliche Erweiterungen des Konzepts der Complementary Interfaces, indem gezeigt wird, wie die vorgestellten Ideen für Design und Evaluation auf sinngebende Kombinationen heterogener Geräte und Modalitäten angewandt werden können, was schließlich Forschungsmöglichkeiten für kollaborative Complementary Interfaces eröffnet. Dabei wird Notwendigkeit hervorgehoben, den Nutzen, die Relevanz und die Rolle einzelner Puzzleteile von Complementary Interfaces weiter zu untersuchen.

Acknowledgments

It feels a bit overwhelming to recognize and name everyone who has contributed to shaping this work, whether directly or indirectly. As a member of the Human-Computer Interaction Group at the University of Konstanz for quite some years, I collaborated with countless inspiring individuals in diverse contexts. I thank all of them, including anyone I may have unintentionally overlooked.

I am incredibly grateful to my advisor, **Harald Reiterer**, for inspiring, guiding, and trusting me. Harald, you gave me the freedom and opportunities I needed and created an atmosphere that allowed me to grow in so many ways. Thank you! I want to thank **Albrecht Schmidt**. Thank you, Albrecht, for supporting and valuing me and my work. Thank you, **Daniel Keim**, for taking the time to serve as an additional reviewer and **Oliver Deussen**, for being the chairperson of the examination committee.

Even before I started this doctoral research journey, I felt immense support: Thank you, **Florian Geyer** for allowing me to contribute to your research project already as a student researcher, and thank you, **Roman Rädle** for motivating me throughout my master studies, early PhD projects, and beyond! A big thank you to **Ulrike Pfeil**! Thank you, Uli, for your positive attitude, for all the countless "*let's have a cup of coffee and then we will solve this problem*" moments, and, of course, our joint adventures at conferences.

Thank you, **Daniel Klinkhammer**, for being such a warm-hearted office buddy! Thank you, **Simon Butscher**, for your calm pragmatism, and thank you, **Jens Müller**, for being the "dude" and all the joy we had as the "J and J" team.

Over the last few years, I had multiple times the chance to be the "sparring partner" to my dear colleagues **Daniel Immanuel Fink**, **Jonathan Wieland**, and **Sebastian Hubenschmid** – which not only lead to an impressive set of publications but also to a lot of fun and good times. Thank you, Daniel, for being the perfect office mate and fellow coffee person! Thank you, Jonathan, for all our conversations about cameras, lenses, and photos! Thank you, Sebastian, for explaining all the technical details I might not remember and your organizational superpowers!

I want to thank **Anke Reinschlüssel** and **Tiare Feuchtner** for your fresh perspectives, your positivity, and your incredible “let’s get this done” attitude.

Thank you, **Thomas Kosch**, **Jakob Karolus**, and **Paweł Woźniak** for pushing the envelope of spreadsheets together and all our conversations about the good, the bad, and the ugly cognitive workload.

I am thankful for all former and current members of the HCI Group Konstanz, including **Maximilian Dürr**, **Yunlong Wang**, **Abdelrahman Zaky**, **Matthias Albrecht**, **Uta Wagner**, **Marianne Kirn-Reutter**, as well as **Ingrid Henschler** but also former and current student researchers such as **Moritz Skowronski**, **Lisa-Maria Reutlinger**, **Philipp von Bauer** and last but not least **Stephan Huber** – a former student researcher who quickly became a good friend, neighbor, and eventually my best man.

To my family and family-in-law, thank you for your love, patience, and keen interest in me and my work. Thank you, **Clemens and Johanna Zagermann** for providing me with opportunities beyond imagination and sending home-baked gingerbread across Europe. Thank you to my sisters **Yvonne and Sonja Zagermann** (and family) for your love and creativity. Thank you, **Daniel and Rita Ammann**, with **Rahel**, **Petra**, and **Eva** for your curiosity and support.

Above all, I thank my lovely wife **Anja Zagermann-Ammann**. I am grateful for your understanding during approaching deadlines, your support and patience during all these odd working hours and work travels, and, of course, all your love that makes my life so wonderful. Thank you!

Publications

Parts of the publications listed below are presented in this thesis. Reused material is highlighted in the introduction and in the beginning of the corresponding chapters. For a better overview, publications that contributed to the core of this thesis are listed first. Other publications that are not directly part of this dissertation are listed second. They are categorized according to their type (e.g., conference publications) and presented chronologically, starting with the newest publication for each category.

Core Publications

Journal Publications

Thomas Kosch*, Jakob Karolus*, **Johannes Zagermann***, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. *ACM Computing Surveys* 55, 13s, Article 283 (December 2023), 39 pages. DOI: https://doi.org/10.1145/3582272 *contributed equally to this research. *see Chapter 6*

Johannes Zagermann*, Sebastian Hubenschmid*, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. 2022. Complementary Interfaces for Visual Computing. *it - Information Technology* 64, no. 4-5, 145-154. DOI: https://doi.org/10.1515/itit-2022-0031 *contributed equally to this research. *see Chapters 1,2, and 7*

Conference Publications

Johannes Zagermann, Ulrike Pfeil, Philipp von Bauer, Daniel Immanuel Fink, and Harald Reiterer. 2020. "It's in My Other Hand!" – Studying the Interplay of Interaction Techniques and Multi-Tablet Activities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, ACM, Honolulu, HI, USA, 1-13. DOI: https://doi.org/10.1145/3313831.3376540 *see Chapter 5*

Johannes Zagermann, Ulrike Pfeil, Daniel Immanuel Fink, Philipp von Bauer, and Harald Reiterer. 2017. Memory in Motion: The Influence of Gesture- and Touch-Based Input Modalities on Spatial Memory. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, ACM, Denver, Colorado, USA, 1899-1910. DOI: https://doi.org/10.1145/3025453.3026001 *see Chapter 4*

Johannes Zagermann, Ulrike Pfeil, Roman Rädle, Hans-Christian Jetter, Clemens Klokmose, and Harald Reiterer. 2016. When Tablets Meet Tabletops: The Effect of Tabletop Size on Around-the-Table Collaboration with Personal Tablets. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, ACM, San Jose, California, USA, 5470-5481. DOI: https://doi.org/10.1145/2858036.2858224

see Chapter 3

Workshop Papers

Johannes Zagermann, Sebastian Hubenschmid, Daniel Immanuel Fink, Jonathan Wieland, Harald Reiterer, and Tiare Feuchtner. 2023. Challenges and Opportunities for Collaborative Immersive Analytics with Hybrid User Interfaces. In: *"HybridUI: 1st Workshop on Hybrid User Interfaces: Complementary Interfaces for Mixed Reality Interaction" (IEEE ISMAR 2023 adjunct)*. DOI: 10.1109/ISMAR-Adjunct60411.2023.00044

see Chapter 7

Sebastian Hubenschmid*, **Johannes Zagermann***, Daniel Immanuel Fink*, Jonathan Wieland*, Tiare Feuchtner*, and Harald Reiterer*. 2021. Towards Asynchronous Hybrid User Interfaces for Cross-Reality Interaction. In: *ACM 2021 Workshop "Transitional Interfaces in Mixed and Cross-Reality: A new frontier?" (TI 2021) (ACM ISS '21 adjunct)*. *contributed equally to this research.

see Chapter 7

Other Publications

Journal Publications

Francesco Chiossi, **Johannes Zagermann**, Jakob Karolus, Nils Rodrigues, Priscilla Balestrucci, Daniel Weiskopf, Benedikt Ehinger, Tiare Feuchtner, Harald Reiterer, Lewis L. Chuang, Marc Ernst, Andreas Bulling, Sven Mayer, and Albrecht Schmidt. 2022. Adapting visualizations and interfaces to the user. *it - Information Technology* 64, 4-5 (2022), 133-143. DOI: https://doi.org/10.1515/itit-2022-0035

Fearn Bishop, **Johannes Zagermann**, Ulrike Pfeil, Gemma Sanderson, Harald Reiterer, and Uta Hinrichs. 2020. Construct-A-Vis: Exploring the Free-Form Visualization Processes of Children. In: *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2020), 451-460. DOI: https://doi.org/10.1109/TVCG.2019.2934804

Conference Publications

Daniel Immanuel Fink, Moritz Skowronski, **Johannes Zagermann**, Anke V. Reinschlüssel, Harald Reiterer, and Tiare Feuchtner. 2024. There Is More to Avatars Than Visuals: Investigating Combinations of Visual and Auditory User Representations for Remote Collaboration in Augmented Reality. Proc. ACM Hum.-Comput.

Interact. 8, ISS, Article 548 (December 2024), 29 pages. DOI: https://doi.org/10.1145/3698148

Mohsen Jenadeleh, **Johannes Zagermann**, Harald Reiterer, Ulf-Dietrich Reips, Raouf Hamzaoui, and Dietmar Saupe. 2023. Relaxed forced choice improves performance of visual quality assessment methods. In *Proceedings of the 2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. DOI: https://doi.org/10.1109/QoMEX58391.2023.10178467

Sebastian Hubenschmid, **Johannes Zagermann**, Daniel Leicht, Harald Reiterer, and Tiare Feuchtner. 2023. ARound the Smartphone: Investigating the Effects of Virtually-Extended Display Size on Spatial Memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, ACM, Hamburg, Germany. DOI: https://doi.org/10.1145/3544548.3581438

Daniel Immanuel Fink, **Johannes Zagermann**, Harald Reiterer, and Hans-Christian Jetter. 2022. Re-locations: Augmenting Personal and Shared Workspaces to Support Remote Collaboration in Incongruent Spaces. In *ACM Hum.-Comput. Interact. 6, ISS, Article 556 (ISS '22)*. ACM, New York, NY, USA. DOI: https://doi.org/10.1145/3567709

Sebastian Hubenschmid*, Jonathan Wieland*, Daniel Immanuel Fink*, Andrea Batch, **Johannes Zagermann**, Niklas Elmqvist, and Harald Reiterer. 2022. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, ACM, New Orleans, LA, USA. DOI: https://doi.org/10.1145/3491102.3517550 *contributed equally to this research.

Jonathan Wieland, **Johannes Zagermann**, Jens Müller, and Harald Reiterer. 2021. Separation, Composition, or Hybrid? - Comparing Collaborative 3D Object Manipulation Techniques for Handheld Augmented Reality. In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR '21)*. Piscataway, NJ:IEEE, pp. 403-412. DOI: https://doi.org/10.1109/ISMAR52148.2021.00057 **Nominated for the Best Conference Paper Award**.

Katja Vock, Sebastian Hubenschmid, **Johannes Zagermann**, Simon Butscher, and Harald Reiterer. 2021. IDIAR: Augmented Reality Dashboards to Supervise Mobile Intervention Studies. In *Mensch Und Computer 2021*. ACM, Ingolstadt, Germany, 248-259. DOI: https://doi.org/10.1145/3473856.3473876

Sebastian Hubenschmid, **Johannes Zagermann**, Simon Butscher, and Harald Reiterer. 2021. STREAM: Exploring the Combination of Spatially-Aware Tablets with Augmented Reality Head-Mounted Displays for Immersive Analytics. In *Proceedings*

of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21), ACM, Yokohama, Japan. DOI: https://doi.org/10.1145/3411764.3445298

Marcel Borowski, **Johannes Zagermann**, Clemens N. Klokmose, Harald Reiterer, and Roman Rädle. 2020. Exploring the Benefits and Barriers of Using Computational Notebooks for Collaborative Programming Assignments. In *Proceedings of the 51st Technical Symposium on Computer Science Education (SIGCSE '20)*, ACM, Portland, OR, USA, 468-474. DOI: https://doi.org/10.1145/3328778.3366887

Jens Müller, **Johannes Zagermann**, Jonathan Wieland, Ulrike Pfeil, and Harald Reiterer. 2019. A Qualitative Comparison Between Augmented and Virtual Reality Collaboration with Handheld Devices. In *Proceedings of Mensch Und Computer 2019*. ACM, Hamburg, Germany, 399-410. DOI: https://doi.org/10.1145/3340764.3340773

Sabrina Jaeger, Karsten Klein, Lucas Joos, **Johannes Zagermann**, Michael de Ridder, Jinman Kim, Jean Yang, Ulrike Pfeil, Harald Reiterer, and Falk Schreiber. 2019. Challenges for Brain Data Analysis in VR Environments. In *2019 IEEE Pacific Visualization Symposium (PacificVis)*, 42-46. DOI: https://doi.org/10.1109/PacificVis.2019.00013

Johannes Zagermann, Ulrike Pfeil, Carmela Acevedo, and Harald Reiterer. 2017. Studying the Benefits and Challenges of Spatial Distribution and Physical Affordances in a Multi-Device Workspace. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*, ACM, Stuttgart, Germany, 249-259. DOI: https://doi.org/10.1145/3152832.3152855

Extended Abstracts (Late-Breaking Work)

Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2018. Studying Eye Movements as a Basis for Measuring Cognitive Load. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*, ACM, Montreal QC, Canada, 1-6. DOI: https://doi.org/10.1145/3170427.3188628

Workshop Papers and Workshop Proposals

Anke V. Reinschlüssel and **Johannes Zagermann**. 2023. Exploring Hybrid User Interfaces for Surgery Planning. In: *"HybridUI: 1st Workshop on Hybrid User Interfaces: Complementary Interfaces for Mixed Reality Interaction" (IEEE ISMAR 2023 adjunct)*.

Abdelrahman Zaky, **Johannes Zagermann**, Harald Reiterer, and Tiare Feuchtner. 2023. Opportunities and Challenges of Hybrid User Interfaces for Optimization of Mixed Reality Interfaces. In: *"HybridUI: 1st Workshop on Hybrid User Interfaces: Complementary Interfaces for Mixed Reality Interaction" (IEEE ISMAR 2023 adjunct)*.

Sebastian Hubenschmid*, Daniel Immanuel Fink*, **Johannes Zagermann**, Jonathan Wieland, Harald Reiterer, and Tiare Feuchtner. 2023. Colibri: A toolkit for rapid prototyping of networking across realities. In: *1st Joint Workshop on Cross Reality (IEEE ISMAR 2023 adjunct)*. *contributed equally to this research.

Sebastian Hubenschmid, **Johannes Zagermann**, Raimund Dachselt, Niklas Elmqvist, Steven Feiner, Tiare Feuchtner, Benjamin Lee, Harald Reiterer, Dieter Schmalstieg. 2023. Hybrid User Interfaces: Complementary Interfaces for Mixed Reality Interaction. Workshop at IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Sydney, Australia.

Sebastian Hubenschmid, Simon Butscher, **Johannes Zagermann**, and Harald Reiterer. 2018. Employing Tangible Visualisations in Augmented Reality with Mobile Devices. In: *Workshop on Multimodal Interaction For Data Visualization (MultimodalVis) (ACM AVI '18 adjunct)*.

Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2016. Measuring Cognitive Load Using Eye Tracking Technology in Visual Computing. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16)*, ACM, Baltimore, MD, USA, 78-85. DOI: https://doi.org/10.1145/2993901.2993908

Johannes Zagermann, Ulrike Pfeil, Mario Schreiner, Roman Rädle, Hans-Christian Jetter, and Harald Reiterer. 2015. Reporting Experiences on Group Activities in Cross-Device Settings. In *Proceedings of Cross-Surface 2015: Workshop on Interacting with Multi-Device Ecologies in the Wild (ACM ISS '15 adjunct)*.

Book Chapters

Florian Geyer, **Johannes Zagermann**, and Harald Reiterer. 2020. Physical meets digital: Blending reality and computational power with digital sticky notes. In *Sticky Creativity*, Bo T. Christensen, Kim Halskov and Clemens N. Klokmose (eds.). Academic Press, 125-151. DOI: https://doi.org/10.1016/B978-0-12-816566-9.00006-9

Contents

Listings and Bibliography

Part I
Introduction
and Background

Parts of the following Chapter 1 have been published as:

Johannes Zagermann, Sebastian Hubenschmid*, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. 2022. Complementary Interfaces for Visual Computing. it - Information Technology 64, no. 4-5, 145-154. DOI: https://doi.org/10.1515/itit-2022-0031 *contributed equally to this research.*

The responsibilities of this joint publication were divided as follows: Sebastian Hubenschmid and I equally spearheaded the writing of the paper. Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, and Albrecht Schmidt contributed to the writing of the paper. Harald Reiterer contributed to the writing and supervised the work.

Supplemental Material

The QR code below allows you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

1

Introduction

"the real power [...] comes not from any one of these devices; it emerges from the interaction of all of them."

Mark Weiser, 1991

Contents

Over the last few years, we have seen an increasing number of interactive devices finding their way into our everyday lives. We employ a smartphone while on the go, a laptop computer in our offices, and, increasingly, an augmented (AR) or virtual reality (VR) head-worn device for immersing ourselves into virtual worlds. Clearly, each device class possesses unique properties that makes it suitable for different contexts and activities.

Despite potential opportunities, most of our workflows are constrained to a single device: Instead of benefiting from an entire device ecology for a given task in a specific context, we often hesitate to incorporate more devices into our workflow and perform entire tasks on a single device, regardless of its suitability for each subtask [316]. The reasons for this constraint are numerous, as to achieve an effective combination of devices for a particular task many factors must be considered, such as the affordances for interaction provided by each device, the continuity of data, or user representations. Therefore, much research has been dedicated to studying

these multi-device ecologies. For example, the field of Cross-Device Interaction [50] examines different constellations of splitting a task across different devices. Even for the nascent field of mixed reality, the area of Hybrid User Interfaces [97] argues for the use of head-worn devices in combination with more common devices (e.g., smartphones or tablets). But what makes these multi-device ecologies worthwhile? Simply adding more devices can be counterproductive, as it might not fit to users' workflow or current activity [316].

1.1 The Need for Evaluation

Recent research streams in Human-Computer Interaction (HCI), such as Cross-Device Interaction [50], Multimodal interaction [394], and Hybrid User Interfaces [97] show parallels to Mark Weiser's famous vision of the computer for the 21st century: "specialized elements of hardware and software, connected [...], will be so ubiquitous that no one will notice their presence." [409]. The technological advances within the last decades allow researchers and designers to study and design new interaction paradigms beyond the boundaries of a single device and modality, leading to various combinations of interfaces that can be used seamlessly in concert.

Indeed, people own several devices [82], and they creatively combine them to "do the UbiComp" [295], but sharing content between nearby devices often results in sending mails, presenting slides on a shared screen fails due to missing display adapters, and heterogenous operating systems lead to improvised setups and a never-change-a-running-system approach – questioning if today's off-the-shelf devices (e.g., tablets) should be a bit less like Swiss army knives – multi-purpose tools with an abundance of possibilities – leading to an overwhelming presence but more "specialized" and "suited to a particular task" as envisioned by Weiser [409].

Similarly, recent research has shown that handling multiple devices can increase cognitive load [325], leading to high transaction costs [153], and users are often not aware of the benefits of including additional devices into their workflow [316].

A potential reason for these issues is that HCI often is "much better at proposing new technologies than at validating them" [149] – especially if novelty is an expected key contribution[1]. Indeed, Brudy et al. [50] highlight the need for validation and evaluation of multi-device environments, as many contributions "re-envision and push the boundaries of interaction possibilities" from a technical perspective, yet they are often "disconnected from findings using empirical studies", leading to a fragmented research landscape.

[1] https://sigbed.org/2022/08/22/the-toxic-culture-of-rejection-in-computer-science/ – last accessed 01.11.2023

1.2 Research Goal

The overarching goal of this thesis is to address this need for evaluation [50, 149] by studying complex Ubiquitous Computing (UbiComp) environments in controlled lab experiments that guarantee for a high internal validity to cope with the elaborated interactions that exceed standardized user study settings [353].

With this, *experiments serve as design-informing activities* [296] for future UbiComp environments (including Cross-Device Interaction, Multimodal Interaction, and Hybrid User Interfaces) – based on meaningful combinations of input and output components that establish **Complementary Interfaces**, a *symbiosis of interfaces*, where each component purposefully increases the quality of interaction and further supports users in their current activity, as "the real power [...] comes not from any one of these devices; it emerges from the interaction of all of them." [409]

In Chapter 2, I will put a spotlight on a variety of UbiComp environments (e.g., Cross-Device Interaction, Hybrid User Interfaces, or Multimodal Interaction) that lead up to the concept of ***Complementary Interfaces****.*

Research presented in this thesis uses typical components of experiments in HCI (see Figure 1.3) as framing regarding the experimental situation itself (e.g., tasks and settings), factors regarding the structure of experiments (e.g., procedures and study designs), and dependent variables (e.g., qualitative and quantitative measurements). In total, three use cases for Complementary Interfaces and a meta-analysis of assessment methods of cognitive workload (as a potential dependent variable of experimental settings for Complementary Interfaces) are presented in this thesis (see Research Objectives 1 and 2).

Addressing this research goal will contribute to and foster our understanding of interaction with Complementary Interfaces and UbiComp environments in general, which will inform their design [296] and evaluation methods [149] likewise.

1.3 Research Questions and Research Objectives

The overall research question is defined as follows:

How can we design and evaluate Complementary Interfaces?

To cover a variety of possible situations for the research goal and research question, this thesis intentionally creates a spectrum of Complementary Interfaces that represent the first research objective: **RO1 – Studying Complementary Interfaces**, with a focus on multi-user (Chapter 3), multimodal (Chapter 4), and multi-device interaction (Chapter 5). This thesis focuses on Complementary Interfaces for knowledge work activities using similar device classes (e.g., tablets). Complementary Interfaces that include immersive components (e.g., head-worn devices as in Hybrid User Interfaces [155]) are not core elements in this thesis.

Specifically, **RO1** will be addressed by three experiments studying three use cases of Complementary Interfaces. These experiments will be guided by the following use case specific general research questions (RQ1-3):

RQ1: *What are the influences of the size of a shared interactive surface on collaboration in a Complementary Interface? (Chapter 3)*

RQ2: *What are the influences of embodiment (represented as meaningful combinations of input and output modalities) on spatial memory in a Complementary Interface? (Chapter 4)*

RQ3: *What are the influences of interaction techniques on users' working behaviors and utilization of components (e.g., devices) in a Complementary Interface? (Chapter 5)*

As research has shown the influences on cognitive workload in multi-device [325] and multimodal environments [439], a particular focus is placed on understanding **RO2 – Measuring Cognitive Workload** by a meta-analysis of relevant related work (Chapter 6) – serving as a potential dependent variable in experimental settings.

Therefore, **RO2** will be addressed by a literature survey on measurements of cognitive workload in HCI in general. The literature survey will be guided by the following general research question (RQ4):

RQ4: *How to measure cognitive workload? (Chapter 6)*

The three use cases for Complementary Interfaces (Chapters 3,4, and 5) each further address individual use case specific research questions based on the general research questions. Here, different design alternatives (e.g., meaningful combinations of in-

put and output components) serve as independent variables for each use case. This allows us to study their effect on general (e.g., time and error) and use case specific dependent variables, such as communication behavior, spatial memory, or utilization of devices – allowing us to focus on the effect, influence, and utility of each meaningful combination, beyond a technical perspective [50]. For RO2, the meta-analysis puts a spotlight on measuring cognitive workload, including its theoretical roots, current assessment methods, and challenges for future research. RO1 and RO2 with their individual use cases (RO1) and insights (RO2) are depicted in Figure 1.1.

Figure 1.1: Research Objective 1 focuses on Studying Complementary Interfaces, and Research Objective 2 addresses Measuring Cognitive Workload.

I introduce research approaches and methods applied for this thesis in the following. Then, I give a brief overview of my research that attributes to the above-mentioned research objectives and questions, and conclude this introduction with my research path.

1.4 Research Approaches and Methods

Research in this thesis follows a multi-layered structure and procedure, combining deductive and inductive reasoning as proposed by Mackay and Fayard [251], as a standard approach for research in HCI. Figure 1.2 illustrates this approach and maps the procedure to the four consecutive stages: Analysis, Design, Empiricism, and Synthesis. In general, for each Complementary Interface that is presented in this thesis (see Chapters 3, 4, and 5), this approach and procedure was followed. The meta-analysis of cognitive workload metrics (see Chapter 6) excludes the design and empiricism steps – focussing on the in-depth analysis of related work and theories to synthesize key findings.

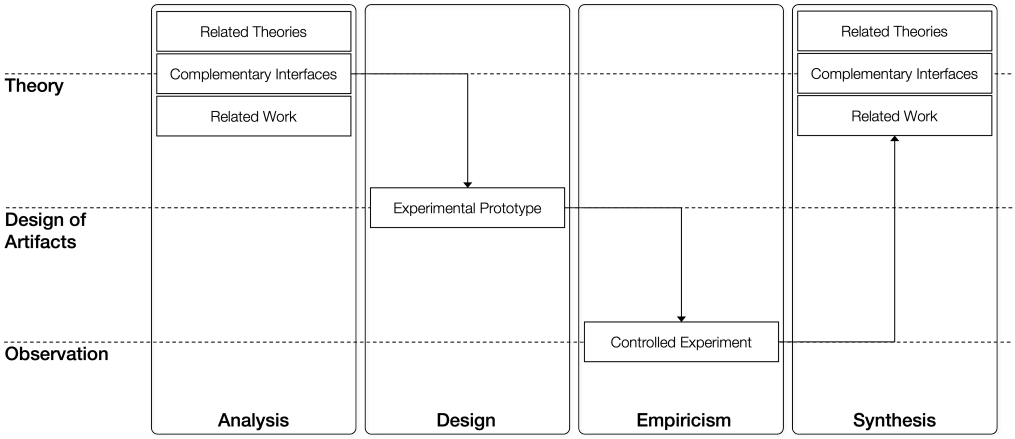


Figure 1.2: Research approach followed in this thesis (based on [251], extended for this thesis).

As shown in Figure 1.2, research on Complementary Interfaces in this thesis (see Chapters 3, 4, and 5) starts with the analysis phase of studying related theories and related work, and identifies research opportunities based on the concept of Complementary Interfaces. In a subsequent design phase, this knowledge is transferred and applied to an experimental prototype that allows studying e.g., design alternatives. Within the following empiricism phase, the experimental prototype is used within a controlled lab experiment to answer use case specific research questions derived from related work and theories. In a concluding step, the knowledge gained via the lab experiment is synthesized to provide insights and implications on research and design, using *experiments as design-informing activities* [296].

In this thesis, several typical components of experiments in HCI as described by Hornbæk [149] (see Figure 1.3) were used as a framing of the presented research: The different design alternatives of each Complementary Interfaces (e.g., differently sized output devices as in Chapter 3) served as **independent variables** that were operationalized each in an individual experimental **situation** (including various character-

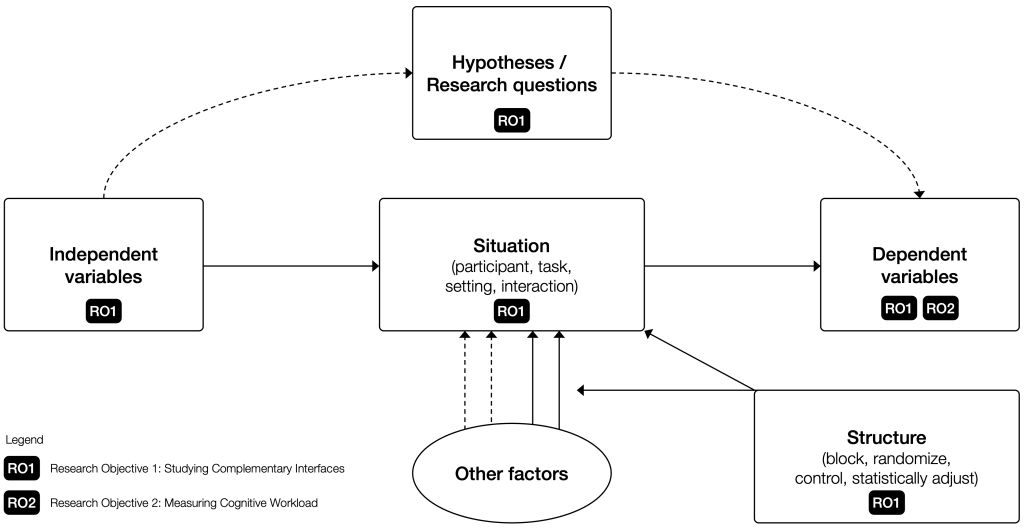


Figure 1.3: Typical components of experiments in HCI (adapted from [149]). Research objectives of this thesis (RO1 and RO2) were added to represent their coverage and allocation of research objectives to components.

istics of participants, tasks, and settings) and studied and analyzed based on use case specific **dependent variables** (i.e., quantitative or qualitative measurements) to inform the design and provide implications [296] for future research of Complementary Interfaces. The research carried out in this thesis purposefully "experimented" with various components of experiments to avoid the trap of becoming a "one method man" [317]. Instead, this resulted in a broad knowledge base for designing, conducting, and analyzing experiments and their outcomes. Additionally, a spotlight was placed on a selected metric (i.e., constituting a potential dependent variable): The in-depth meta-analysis of cognitive workload measurements. The typical components of experiments in HCI will be further depicted in Chapter 2, and all research presented in this thesis will be described accordingly in their chapters.

1.5 Research Ethics

All user studies reported in this thesis were granted ethics approval by the Ethics Committee of the University of Konstanz[2]. Participants received a welcoming and informed consent sheet for each user study, including descriptions of the experiments in everyday language, the course of study, collected data, their rights to abort the user study, and their compensation. The data collection did not start before participants signed the informed consent. All participants were recruited at the University of Konstanz using printed flyers and a social media bulletin board. Participants were

[2] https://www.uni-konstanz.de/en/university/administration-and-organisation/university-bodies-and-committees/university-bodies-for-scientific-integrity/ethics-committee/– last accessed 01.11.2023

compensated for their participation. All presented user studies were conducted as laboratory experiments. The assignment of participants for all user studies to specific conditions and orders was done based on the individual user study design (e.g., via counterbalancing). For user studies involving dyads (see Chapter 3), I carefully assigned participants to their study partner based on pre-assessed demographics.

1.6 Context and Research Path

The overall context of this research is provided by my work as a research assistant and PhD student in Harald Reiterer's Human-Computer Interaction Group at the University of Konstanz and further framed by the research project SFB-TRR 161 (SFB) "Quantitative Methods for Visual Computing". The SFB is a collaborative research center connecting research teams across universities (University of Stuttgart, University of Konstanz, LMU Munich, and Ulm University) and disciplines (e.g., HCI, Information Visualizations, or Cognitive Psychology). Within project C01, "Quantifying Interaction" / "Evaluating Mixed Reality Experiences"[3] we set out to measure, understand, and improve the quality of interaction in visual and immersive analytics environments. This provided me unique opportunities to *experiment with experiments*, collaborate with peers within and across space and disciplines, and apply, test, and finally present the gained knowledge.

During the time of research for this thesis, fortunate circumstances brought many fellows into my life that supported me personally and professionally, and – driven by my curiosity, interest, ~~and a missing ability to say "no"~~ – I was able to also help them e.g., by applying my knowledge to support them in conducting their research. Luckily, helping others further helped me to gain knowledge, e.g., how to design, conceptualize, and evaluate their apparatus – this eventually allowed me to further refine, extend, and test the concept of Complementary Interfaces beyond the research presented in this thesis. With this, research related to the core of this thesis and parallel activities can be regarded as complementary to each other and further helped to avoid the trap of becoming a "one method man" [317]. Section 1.6.6 presents a selection of these collaborations – it is not meant to take any credit away from the tremendous work of my dear fellows, but instead, allow the reader to understand which thesis-related or less related topics attracted me along the way.

The core research journey of this thesis was guided by three use cases for Complementary Interfaces and an in-depth meta-analysis of cognitive workload measurements (Chapter 6). The three use cases are unified in the concept of Complementary Interfaces [433, see Chapter 2] that builds one of their theoretical foundations. I processed the use cases consecutively, starting with the design and evaluation of a

[3]The project title was purposefully expanded during the runtime of the project to reflect its research.

multi-user environment (Chapter 3), followed by the use case of multimodal interaction (Chapter 4), and finally the use case of multi-device interaction (Chapter 5). The selection of these use cases is based on several rationales: First, the use cases complement each other as they focus on different kinds of activities, different combinations of components, and various dependencies of components. Secondly, the use cases allow us to investigate the influences of variants of output modalities [428], input techniques [432], and their interaction [430]. This allowed to design a variety of Complementary Interfaces; and also offered many possibilities to evaluate them in study designs that e.g., prevent or purposefully encourage learning effects between conditions with carefully selected tasks – eventually leading to complementary evaluation techniques. The research is complemented by a meta-analysis of current and past measurements of cognitive workload – a potential dependent variable in experimental settings studying Complementary Interfaces – as prior research has shown the effects of multi-device [325] and multimodal environments [439] on cognitive workload. Next, I will briefly outline the conducted research and conclude this section with an overview of collaborations.

1.6.1 Complementary Interfaces

With increasing complexity in visual computing tasks, a single device may not be sufficient to adequately support users' workflows. Here, current research streams that run in parallel to the vision of Ubiquitous Computing (e.g., Cross-Device Interaction, Hybrid User Interfaces, or Multimodal Interaction) allow to distribute activities across multiple devices – potentially each dedicated to a specific role. But what makes these multi-device ecologies compelling? This question leads to the establishment of the concept of Complementary Interfaces – meaningful combinations of input and output modalities that support users in their current activity and thereby increasing the quality of interaction.

Johannes Zagermann, Sebastian Hubenschmid*, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. 2022. Complementary Interfaces for Visual Computing. it - Information Technology 64, no. 4-5, 145-154. DOI: https://doi.org/10.1515/itit-2022-0031 *contributed equally to this research.*

1.6.2 Studying Multi-User Interaction

Tablets allow for mixed-focus collaboration [128, 387] where users frequently transition from tightly-coupled collaboration to loosely-coupled parallel work, e.g., for reading a document. Here, a complementary shared interactive workspace can greatly benefit tightly-coupled activities. Yet, large surfaces are costly and need space, raising the question of to what extent the physical size of shared horizontal surfaces pays off. To analyze the suitability of smaller-than-tabletop devices (e.g., tablets) as a low-cost alternative, we studied the effect of the size of a shared horizontal interactive workspace on users' attention, awareness, and efficiency during collaboration.

In our user study, 15 dyads performed a sensemaking task with two individual tablets and a complementary horizontal shared display of varying sizes, representing three conditions. Our findings show that different sizes lead to differences in participants' interaction with the shared surface and in dyads' communication styles. We found that larger surfaces do not necessarily improve collaboration, as they can divert users' attention away from their collaborators and towards the shared display.

***Johannes Zagermann**, Ulrike Pfeil, Roman Rädle, Hans-Christian Jetter, Clemens Klokmose, and Harald Reiterer. 2016. When Tablets Meet Tabletops: The Effect of Tabletop Size on Around-the-Table Collaboration with Personal Tablets. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), ACM, San Jose, California, USA, 5470-5481. DOI: https://doi.org/10.1145/2858036.2858224*

1.6.3 Studying Multimodal Interaction

People's ability to remember and recall spatial information can be utilized to improve navigation and search performance. Recent research has shown that both, input and output modalities influence users' spatial memory and navigation performance [e.g., 179, 304, 324]. As previous studies investigated input and output modalities separately, the question arises of how meaningful combinations influence spatial memory.

We investigated how complementary combinations of display size and input modality influence spatial memory, especially in relation to efficiency and user satisfaction. Based on an experiment with 28 participants, we analyze the effect of three input modalities and two display sizes on participants' ability to navigate to spatially distributed items and recall their positions. Our findings show that the impact of input

modality and display size on spatial memory is not straightforward but characterized by trade-offs between spatial memory, efficiency, and user satisfaction.

***Johannes Zagermann**, Ulrike Pfeil, Daniel Fink, Philipp von Bauer, and Harald Reiterer. 2017. Memory in Motion: The Influence of Gesture- and Touch-Based Input Modalities on Spatial Memory. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), Association for Computing Machinery, Denver, Colorado, USA, 1899-1910. DOI: https://doi.org/10.1145/3025453.3026001*

1.6.4 Studying Multi-Device Interaction

Research on multi-device environments is based on the idea that users easily interact with multiple devices, not recognizing the connected devices as single entities but perceiving them as components of the same underlying system. Recent work has shown the benefits of including complementary devices in users' workflows while various interaction techniques allow transferring content across devices. However, users are only reluctantly using multiple devices in combination. At the same time, research struggles to find a frame of reference to compare these techniques or systems.

This research addresses these challenges by studying the interplay of interaction techniques, device utilization, and task-specific activities in a user study with 24 participants from complementary angles of evaluation using an abstract task, a sensemaking task, and three interaction techniques. We found that interaction techniques have a lower influence than expected, that device utilization depends on the task at hand, and that participants value specific aspects of Complementary Interfaces.

***Johannes Zagermann**, Ulrike Pfeil, Philipp von Bauer, Daniel Fink, and Harald Reiterer. 2020. "It's in My Other Hand!" – Studying the Interplay of Interaction Techniques and Multi-Tablet Activities. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Association for Computing Machinery, Honolulu, HI, USA, 1-13. DOI: https://doi.org/10.1145/3313831.3376540*

1.6.5 Measuring Cognitive Workload

The ever-increasing number of computing devices around us results in more and more systems competing for our attention, making cognitive workload a crucial factor for the user experience of interfaces. Research has used various metrics to determine users' mental demands. However, there needs to be a systematic way to choose an appropriate and effective measure for cognitive workload in experimental setups, posing a challenge to their reproducibility. To address this challenge, we present a literature survey of past and current metrics for cognitive workload. By initially exploring what cognitive workload resembles in the HCI context, we derive a categorization supporting researchers in selecting cognitive workload metrics for system design and evaluation. We conclude with three following research gaps: (1) defining and interpreting cognitive workload in HCI, (2) the hidden cost of the NASA-TLX, and (3) HCI research as a catalyst for workload-aware systems, highlighting that HCI research has to deepen and conceptualize the understanding of cognitive workload in the context of interactive systems.

Thomas Kosch, Jakob Karolus*,* ***Johannes Zagermann*****, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. ACM Computing Surveys 55, 13s, Article 283 (December 2023), 39 pages. DOI: https://doi.org/10.1145/3582272 *contributed equally to this research.*

1.6.6 Overview of Collaborations

Research presented in this thesis focuses on Complementary Interfaces using homogenous classes of devices (e.g., a combination of multiple handheld devices). In line with this, I was able to contribute to an *in-the-wild* user study in a primary school in Scotland – with *Construct-A-Vis*, we studied individual and collaborative visualization-related activities of school children [37]. The in-the-wild user study was an essential aspect of this work and further contributed to my methodological knowledge of conducting user studies beyond classical lab experiments. Similarly, in a *longitudinal* user study that was conducted as part of the lecture "Interactive Systems" at the University of Konstanz, I was able to contribute to the exploration of collaborative programming practices using computational notebooks [42].

From a technological point of view – beyond my own specific research – I was interested in how immersive technologies (e.g., mixed reality[4]) can support individual and collaborative activities. Here, I was involved in a lab experiment studying collaborative, co-located 3D object manipulation techniques [414] with augmented reality tablets. In line with this, we used tablets to systematically analyze the influence of level of immersion (i.e., augmented vs. virtual reality) and spatial dispersion (i.e., co-located vs. remote) on collaborative activities [277]. Covering remote collaboration, I was involved in conceptualizing and studying a solution approach to enrich remote collaboration with qualities of co-located collaboration (e.g., spatial awareness) using head-worn augmented reality [100]. Lastly, I helped in exploring the design space of immersive analytics to analyze brain data in virtual reality [173].

Combining the research on handheld and head-worn modalities, I was involved in the conceptualization and evaluation of multiple Complementary Interfaces that include immersive and non-immersive components: With STREAM, we studied the suitability of combining spatially-aware tablets for precise input and head-worn augmented reality for close-to-unlimited output for immersive analytics [69, 154, 155]. With IDIAR, we studied a multimodal interaction concept and combined a spatially-agnostic smartphone for familiar input with interactive dashboard visualizations in head-worn augmented reality [402]. In "ARound the Smartphone" [159], we used a head-worn augmented reality device to virtually extend the screen space of a smartphone to study the influence of the size of these extensions on participants' spatial memory.

Finally, I was able to contribute my knowledge of evaluating user interfaces to the mixed-immersion visual analytics evaluation framework ReLive [157]. ReLive is an evaluation framework that combines in-situ and ex-situ visualizations for the holistic analysis and exploration of mixed reality user study data. ReLive can be used by individual analysts asynchronously [156] or potentially by multiple analysts synchronously, resulting in a collaborative Complementary Interface across and within different levels of immersion – leading to possible research directions that are further outlined in Chapter 7.

1.7 Contributions

With this research approach and applied research methods, this thesis provides a solid foundation for designing and evaluating Complementary Interfaces. To address my research objectives, I divided research into smaller projects and prototypes instead of developing an overarching research application. This was important to avoid confounding factors and achieve valid results; additionally, it allows for the

[4] including both, augmented and virtual reality [365]

generalization of research findings. In detail, this thesis disseminates five types of contributions that are relevant to the HCI community [417]: 1) empirical, 2) artifact, 3) theoretical, 4) methodological, and 5) survey. All contributions are published as part of several publications [e.g., 217, 428, 430, 432, 433].

Empirical Contributions: This thesis provides insights into the benefits of different design alternatives of Complementary Interfaces on aspects such as task completion times and error rates, collaborative behavior, or utilization of devices. In addition, this thesis further investigates trade-offs resulting from different meaningful combinations of Complementary Interfaces. Design alternatives are set as independent variables for each experiment. In total, three experiments studying three use cases of Complementary Interfaces are presented in this thesis.

Artifact Contributions: As enabling technologies for each empirical contribution, artifact contributions represented as experimental prototypes were designed and implemented. Different design alternatives allowed the study of their individual effects and influences within their specific use case.

Theoretical Contribution: This thesis contributes the concept of Complementary Interfaces, one of the theoretical foundations of the presented research in this thesis. Complementary Interfaces aim for meaningful combinations of interfaces that support users in their current task at hand and can be considered as an umbrella term for a set of UbiComp environments.

Methodological Contributions: For most of the research conducted and presented in this thesis, existing research methods had to be altered, and new research methods and approaches had to be developed. Particularly, the methodological approach of integrating the study task as an independent variable in Chapter 5 led to complementary evaluation techniques.

Survey Contribution: As cognitive workload is an important aspect when interacting with multiple interactive components, an in-depth meta-analysis of current assessment methods of cognitive workload was conducted that resulted in a survey contribution that identified gaps for future research.

1.8 Chapter Preview

The following paragraphs provide an overview of all chapters, including brief summaries of their contents. In general, this thesis is structured in 7 chapters, including this introduction chapter. While the first part of this thesis covers the motivation, procedure, and theoretical foundations of this work (Chapters 1 and 2), the second

part provides insights into the design and evaluation of Complementary Interfaces for multi-user (Chapter 3), multimodal (Chapter 4), and multi-device interaction (Chapter 5). The third part puts a spotlight on understanding cognitive workload measurements (Chapter 6). The thesis concludes with a summary of contributions and an outlook on open research directions in Chapter 7.

Chapter 2: Theoretical Background provides a deeper understanding of the theoretical background that guided this research: This includes the way from Ubiquitous Computing towards Complementary Interfaces, an overview of relevant related work for the use cases of Complementary Interface; and a focus on the components of experiments in HCI, emphasizing *experiments as design-informing activities*.

Chapter 3: Studying Multi-User Interaction presents the design of a Complementary Interface to study collaborative activities. Here, an experiment was used to investigate the influences and effects of the size of an interactive shared surface on collaborative sensemaking activities.

Chapter 4: Studying Multimodal Interaction presents the design of a Complementary Interface to study a variety of meaningful combinations of input and output modalities. The experimental situation shed light on their influences on spatial memory and how different meaningful combinations of devices can lead to trade-offs e.g., regarding their usability.

Chapter 5: Studying Multi-Device Interaction presents the design of a Complementary Interface to study multi-device interaction. The experiment includes a methodological approach of treating the experimental task as an independent variable, which leads to complementary evaluation techniques.

Chapter 6: Measuring Cognitive Workload provides an understanding of applied measurements of cognitive workload as a potential dependent variable of experiments for Complementary Interfaces and HCI research in general. Additionally, this chapter shows opportunities for guiding researchers to understand the role of cognitive workload in HCI.

Chapter 7: Conclusion summarizes the contents, integrates findings, and reflects on them. It concludes with research opportunities for future work.

Parts of the following Chapter 2 have been published as:

Johannes Zagermann, Sebastian Hubenschmid*, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. 2022. Complementary Interfaces for Visual Computing. it - Information Technology 64, no. 4-5, 145-154. DOI: https://doi.org/10.1515/itit-2022-0031 *contributed equally to this research.*

The responsibilities of this joint publication were divided as follows: Sebastian Hubenschmid and I equally spearheaded the writing of the paper. Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, and Albrecht Schmidt contributed to the writing of the paper. Harald Reiterer contributed to the writing and supervised the work.

Supplemental Material

The QR code below allows you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

2

Background

"... the whole computing experience is more than the sum of its individual devices."

Chris Harrison et al., 2010

Contents

This chapter provides the theoretical background that informed the work in this thesis. In a first step, the theoretical roots of Ubiquitous Computing environments are presented, which brings us to an overview of related work in the areas of Cross-Device Interaction, Hybrid User Interfaces, and Multimodal Interaction. These streams of research are then combined into the concept of Complementary Interfaces. Finally, and in a second step, empirical research methods with a special focus on experiments in HCI will conclude the background chapter.

The theoretical background for measuring cognitive workload is highly interwoven with current ways to assess this construct and part of the literature survey; therefore, Chapter 6 is self-contained.

2.1 Towards Ubiquitous Computing

Before starting to dig deeper into the topic of Ubiquitous Computing, the historical developments in hardware (e.g., the sheer size) and software (e.g., interaction concepts) – that go hand in hand with the relationships between computers and their users – are shortly presented: From large (1) mainframe computers that fill entire buildings, to (2) personal computers for offices and homes, to (3) mobile computing in our pockets, to the (4) ubiquity area, described by Harper et al. [131] as four computing eras, as shown in Figure 2.1.

Mainframe Era Room-filling mainframe computers dominated this era. To operate these computers, multiple trained technicians were needed to orchestrate punch cards, magnetic tapes, and printouts. Batch processing allowed for a sequential use of these devices; time-sharing allowed for a more efficient use of the computational capacities using multiple terminals simultaneously. Harper et al. [131] describe the dominant use of computing devices as a one computer per many users relationship.

Personal Computer Era The personal computer era brought computing devices into people's homes and offices. The personal computer changed the relationship towards a one-to-one relationship between computers and individuals (or often families). Personal computers became commodity devices, and they were heavily influenced by one of their primary purposes: Office work activities that required word processors or spreadsheets. Here, the desktop metaphor with its folders and documents in combination with the "WIMP" (windows, icons, menus, and pointer) interface can be seen as their key to success. Using a keyboard and a mouse, users could interact with graphical objects on screens. Although this combination of the desktop metaphor and WIMP interfaces dates back to the early 1980s, it is still used in various ways today.

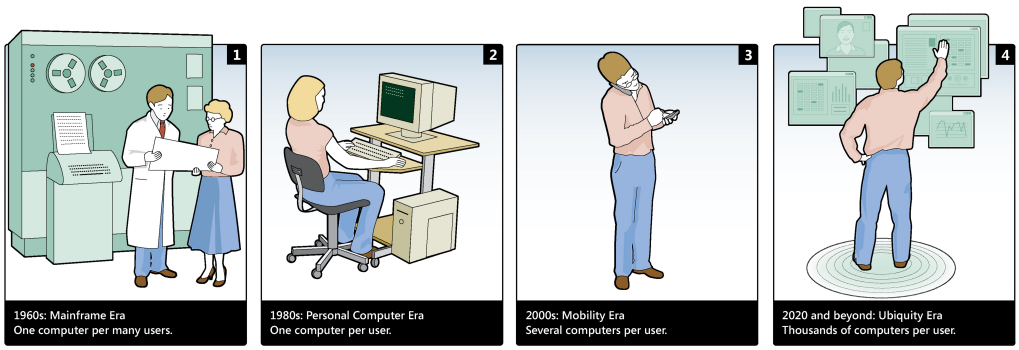


Figure 2.1: Four computing eras as described by Harper et al. [131]: (1) Mainframe era, (2) Personal computer era, (3) Mobility era, and (4) Ubiquity era.

Mobility Era The mobility era brought computing devices another step closer to their users. While mainframe computers were room-filling and PCs colonized desktops [410], the advent of mobile devices such as PDAs (personal digital assistants), mobile phones, tablets, laptops, and other portable computing devices led to another shift of interaction paradigms and relationships between computing devices and its users. Here, users often own several devices with different form factors and capabilities, allowing for their use while on the go. Their size and portability required new ways to interact with them, from input via (soft) keyboards or styluses to (multi) touch or gestures.

Ubiquity Era The ubiquity era expresses another radical shift not only regarding the number, types, and physical or virtual representation of computing devices but also concerning the interaction paradigms. While the form factor and especially the size of computing devices decreased heavily across the previous eras, the ubiquity era highlights that interactivity is not restricted to the physical boundaries of a device, but computing becomes ubiquitous – from smart devices to virtualized computing components that interact with each other and augmented users' spatial surroundings. With this, new interaction concepts are needed that allow for gestural interactions, combinations of explicit and implicit user input, or even physiological sensing – moving an abundance of "computers" even closer to their users.

From today's technology perspective, Harper et al.'s [131] vision for Human-Computer Interaction in the year 2020 was almost accurate. With the advent of more and more off-the-shelf devices that allow for interaction in mixed reality[1] (e.g., Apple's Vision Pro [12]), the technological foundations for the ubiquity era are under heavy development. However, although potential hardware is available, the current limitations of these devices, e.g., technical aspects (e.g., limited field of view), low wearing comfort, or interaction techniques restrict their use cases (e.g., to immersive visualizations or entertainment) but open up the space for future research at the same time.

The ideas behind the two eras *Mobility* and *Ubiquity* have their roots in Mark Weiser's famous vision of the computer for the 21st century: "specialized elements of hardware and software, connected [...], will be so ubiquitous that no one will notice their presence." [409]. During the high time of the PC era in the late 1980s and early 1990s, Mark Weiser and colleagues at XEROX Parc defined their ideas that resulted in a counterpart to the Personal Computer: Ubiquitous Computing.

[1] including both, augmented and virtual reality [365]

The following sections will further dive into the topic of Ubiquitous Computing (Section 2.2), then split into related research that shows parallels to Weiser's vision on Cross-Device Interaction (Section 2.3), Multimodal Interaction (Section 2.5), and Hybrid User Interfaces (Section 2.4) – leading up the concept of Complementary Interfaces (Section 2.6). The chapter concludes with an overview of experiments in HCI (Section 2.7).

2.2 Ubiquitous Computing

Indeed, the vision for the computer for the 21st originates from a "...radical answer to what was wrong with the personal computer: too complex and hard to use; too demanding of attention; too isolating from other people and activities; and too dominating as it colonized our desktops and our lives." [410]

Mark Weiser's answer is described in his vision of the computer for the 21st century: "Specialized elements of hardware and software, connected [...], will be so ubiquitous that no one will notice their presence." [409]

This introducing sentence can be unraveled into three components: (1) "no one will notice their presence", (2) "will be so ubiquitous", and (3) "Specialized elements of hardware and software, connected [...]".

(1) "no one will notice their presence" – Mark Weiser anchors his vision for Ubiquitous Computing not only as an answer to the Personal Computing Era but also in relation to other information technologies (i.e., writing) and psychology. Here, he especially highlights writing as a technology available at any place and any time (i.e., ubiquitous), while not necessarily acquiring active attention and being in the background. Similarly, based on ideas e.g., from psychology and philosophy, Weiser describes this "disappearance" as an ultimate consequence of mastering a technology – absorbing information without active or conscious application of it. Weiser envisions using computers comparable to the way that a "well-balanced hammer '*disappears*' in the hand of a carpenter" [410] – leading to so-called *calm computing*, a state of mind that eventually allows to establish a flow state, instead of "mental gymnastics" [409] to use computers. With this, "they weave themselves into the fabric of everyday life until they are indistinguishable from it", eventually releasing cognitive resources to focus on the task at hand.

(2) "will be so ubiquitous" – While the Mainframe and PC era focused on the many-to-one and later one-to-one relationship of users and computers, Weiser envisioned that computing will be freed from its manifestations of expensive, bulky boxes to – again comparable to technologies such as writing – a high availability and affordability of computing technologies. These cheap computers can then be part of an abundance of everyday objects, tools, and accessories; usable wherever and whenever needed. This eventually implies that there is no need to bring your own device(s), but to use available computing devices ad hoc e.g., in office spaces like sheets of paper.

(3) "Specialized elements of hardware and software, connected . . . "– Similarly as the office environment inspired the desktop metaphor, Mark Weiser also proposed a set of devices (see Figure 2.2) that can make the vision come true: Small-scale **tabs** that resemble sticky notes, medium-sized **pads** that come at the size of sheets of paper, and large **boards** that are equivalent to blackboards or whiteboards – each of them with dedicated purposes. The core of this device ensemble is in its quantity. Weiser suggested that office spaces will typically provide access to tens and hundreds of small and medium-sized devices – leading to usage patterns known from using "scrap paper" – instead of bringing a portable computer to e.g., a meeting, these pads are just available to use everywhere. Weiser argues that the concept of windows allowed to bring multiple applications synchronously to the desktop computer but that these novel devices are counterparts to windows that bring the applications to the real desk, which allows for interactions known from working with paper: Sorting, piling, or reminding. Their interconnectivity further allows them to "seamlessly" sense their users, authenticate them, or share content between and across computing devices.

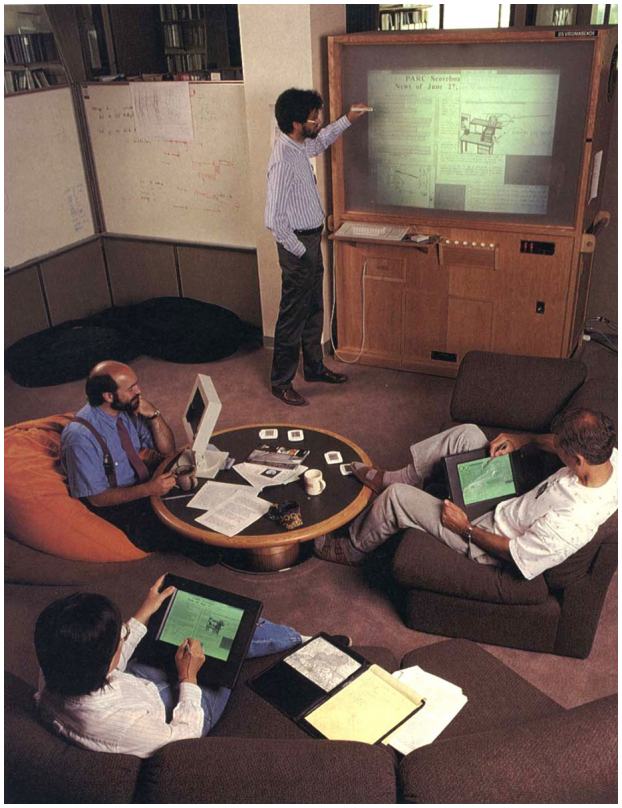


Figure 2.2: Mark Weiser and colleagues showcasing Ubiquitous Computing with tabs, pads, and a board [409].

More than 30 years later, one could ask: Was Weiser right? Or even "*are we there yet?*" [354]

Weiser's vision heavily inspired numerous practitioners and researchers. Technology has come a long way from the early 1990s to present-day computing devices. Mobile devices entertain us on the go, allow us to work in basically any environment, and network technologies (e.g., Wifi or Bluetooth) enable the synchronization of computing devices. The Internet of Things [110] makes artifacts in our environment smart: From the food-ordering fridge to smart toothbrushes, or self-regulating heating systems. Similarly, interactive tangibles [169] bring computing power to everyday objects, almost weaving "themselves into the fabric of everyday life until they are indistinguishable from it" [409]. Although the desktop computer still has its right to exist, it is still quite often a personal device, yet embedded in slightly different circumstances: The personal computer (e.g., today, often laptops) isn't the only device we use in our daily activities. It is rather a whole ecology of personal devices that complement each other for different purposes. These computing devices can be related to Weiser's tabs, pads, and boards: While today's smartwatches and smartphones serve similar purposes as Weiser's tabs and come at similar physical sizes, pads are almost directly resembled by tablets (e.g., Apple iPad), and boards found their way e.g., into our office spaces as shared devices to present and collaborate. Devices within the same ecosystem (e.g., Apple devices, logged in via an Apple ID) further facilitate the transition of data and activities across devices. Yet, they are not abundantly used as envisioned. Individuals own several different devices [82], but not in the numbers as Weiser thought of. Harrison et al. [132] attribute this to the fact that people rather have fewer devices with more capabilities (i.e., quality computing) instead of using a large number with fewer capabilities each (i.e., quantity computing) – a trend that is seized by industry by equipping devices with more and more sensors, memory, and capabilities, enabling its users to do more with fewer devices. Although Mark Weiser heavily argued against Virtual Reality [409], today's mixed reality devices (in this case primarily Augmented Reality) can allow for ubiquitous interfaces in combination with immersive technologies, which is depicted by Newman et al. [283] in their Milgram-Weiser Continuum that allows to e.g., classify interactive systems based on their degree of immersion as well as their quantitative person-computer relationship.

A research stream that strongly parallels Weiser's vision can be seen in today's multi-device and Cross-Device Interaction. Research on Cross-Device Interaction is based on the idea that users easily interact with multiple devices, not recognizing the connected devices as single entities but perceiving them as components of the same underlying system. Typical devices include smartphones, tablets, or larger interactive

surfaces that can be used seamlessly in concert. General and more specific literature on Cross-Device Interaction will be part of the following Section 2.3.

2.3 Cross-Device Interaction

Brudy et al. [50] provide a comprehensive overview of research on cross-device computing. Their goal is to unify research that "transcends the individual device and user" by creating a taxonomy, unraveling the design space, and offering insights into open challenges for this research stream. This section is not meant to report on their taxonomy extensively but to provide the reader with information related to cross-device computing that not only informed the apparatus and research questions for two use cases of Complementary Interfaces (see Chapters 3 and 5) but also to understand their influence on the concept of Complementary Interfaces (see Section 2.6).

Brudy et al. [50] divide the historical development of cross-device computing into three areas: (1) Although research on Cross-Device Interaction often refers to Weiser's vision of Ubiquitous Computing, early work on **Multi-Monitor Workstations** starting in the early 1980s, is already considered as part of the historical roots. Here, unlike Weiser's vision, these workstations extend or mirror the screen space of personal com-

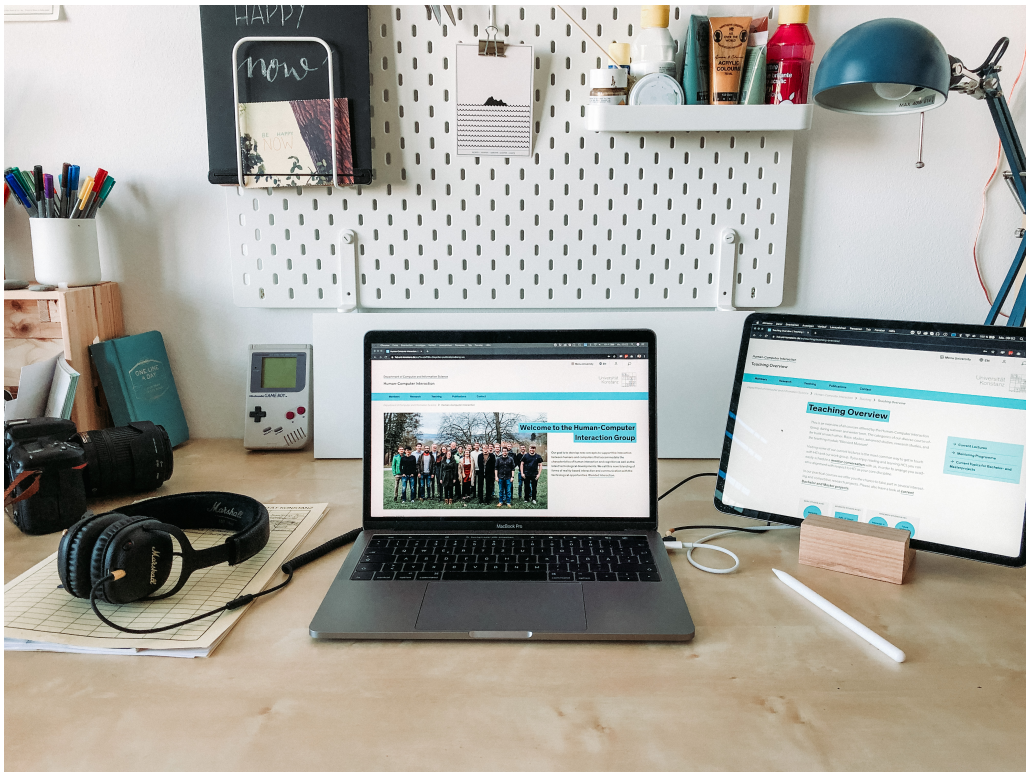


Figure 2.3: Combining multiple mobile devices into a typical office workspace facilitated by the ideas of Cross-Device Interaction.

puters by adding one or more additional screens. Setups like this are pretty common in present-day office spaces – yet these setups are usually static, e.g., with a single primary device and additional extensions. Research in this thesis does not further investigate this area. (2) The second area targets **Multi-Device Environments and Spaces** that are directly inspired by Ubiquitous Computing. Here, research investigates environments that combine a variety of device types such as tablets and large interactive surfaces (cf. Weiser's pads and boards), as well as interaction techniques that cross device boundaries (e.g., Rekimoto's Pick-and-drop technique [333]). These multi-display environments allow for the spatial or logical distribution of user interfaces. The term Distributed User Interfaces [90] was often used to describe and classify these environments before the term "Cross-Device Interaction" was born. Chapter 3 further investigates the meaningful combination of tablets and shared interactive surfaces, and Chapter 5 studies interaction techniques that cross device boundaries. Finally, (3) the third area is framed by **Ad hoc, Mobile Cross-Device Use**. In contrast to the other areas, this work concentrates e.g., on the flexible and spontaneous combination of devices. Here, scenarios include highly dynamic setups often facilitated by tracking technologies. This research strives towards larger device ecologies that can be arranged and incorporated to fit the task at hand. Chapter 5 investigates the usage of tablets – yet excluding additional hardware tracking.

The tracking of devices (e.g., their position or rotation) and the data exchange between and across devices (e.g., to share application states) is a basis for Cross-Device Interaction. Here, researchers investigated ways to track devices without hardware-heavy and costly setups (e.g., marker-based infrared tracking using Vicon or Optitrack cameras) to allow for ad hoc usage scenarios by detecting the e.g., position, orientation, or identity of devices. Brudy et al. [50] distinguish between (1) outside-in tracking: Sensors placed in the environment such as depth cameras are used to detect devices (e.g., [325, 327]) and (2) inside-out tracking: Built-in sensors of devices are used to sense e.g., environmental features like the head of a user [122] – potentially allowing for ad hoc and mobile use of multi-device environments. The tracking of devices is not mandatory to create cross-device experiences. Yet, the data exchange between and across devices is mandatory. Weiser envisioned the connection between devices e.g., to synchronize and exchange data based on wireless networks using infrared technologies or radio frequencies [409]. Today, the most common way to connect devices is through local wifi networks. Here, various open source web-based frameworks evolved that allow for the synchronization of devices (e.g., Metor.js[2]), research investigated ways to detect nearby devices (e.g., [358]), and to share and synchronize content across devices (e.g., using research-driven frameworks like Webstrates [202] or one of its variants [21, 147, 326]).

[2]Meteor.js – https://www.meteor.com – last accessed 01.11.2023

The following subsections are intended to provide a background on this research area that paved the way for the vision of Complementary Interfaces. Additionally, the individual paragraphs link to the corresponding chapters for further reading.

With the growing popularity of Cross-Device Interaction, researchers have proposed working with multiple mobile devices instead of but also in combination with large interactive surfaces (i.e., tabletops). McGrath et al. [264] used a private view on a tablet and a public view on a tabletop for collaborative visual data exploration to support loosely-coupled parallel work and tightly-coupled activities – leading to mixed-focus collaboration [128, 387]. The coupling of views allows to merging individual explorations on tablets with collaborative results. Wallace et al. [405] studied collaborative sensemaking with tablets only, tabletop only, and a combination of them. Here, the presence of a large shared surface led to improved sensemaking performance, supported the ability to prioritize information, make comparisons between task-related information, and support the collaborative procedure.

The combination of tablets for individual and a shared interactive surface for collaborative activities (e.g., [405]), inspired varying the size of the shared interactive surface as an independent variable during a user study. Chapter 3 contributes to a better understanding of the meaningful combination of individual tablets and a shared interactive surface in three different sizes during collaborative sensemaking.

Brudy et al. [49] continued to investigate the role of a shared tablet-size device in collaborative activities and showed that an additional overview device can extend periods of closely-coupled collaboration. Wozniak et al. [419] presented a system that allows for collaborative activities with spatially-aware tablets – and compared it to a paper-based and tabletop condition using a sensemaking task. Their system revealed faster task completion times compared to paper and similar results for task load compared to the tabletop condition. Hamilton and Wigdor [130] have shown that participants can quickly adapt to unfamiliar settings and make use of up to ten mobile devices in a tablet-only setting working on a sensemaking task. Chen et al. [60, 61] investigated active reading activities using a multi-tablet system. Compared to paper, their system provides enhanced navigation within documents and allows for synchronization across multiple devices. Plank et al. [316] built on this research and investigated the usage of mobile devices in a tablet-only setting for collaborative visual analytics. In contrast to previous work [130], Plank et al.'s participants did not exhaust all available devices and solved the tasks with one or two devices. The

authors refer to a legacy bias that influences participants in how they deal with mobile devices, resulting in the usage of fewer tablets than expected.

Researchers studied a wide variety of interaction techniques that cross device boundaries. Rädle et al. [325] investigated if user interfaces for transferring documents across devices should provide spatial information about nearby devices. They compared a menu-based approach (cf. [130]), without information on the location of nearby devices, to two spatially-aware techniques using a camera-based tracking system [323]. Results show that spatially-aware techniques (e.g., dragging an item 'through' a portal-like area to another device) were preferred and lowered participants' task load, yet the menu-based approach was beneficial for settings where the spatial configuration of devices is not changing rapidly.

With portal-like techniques (e.g., [257, 325]), content can be transferred to a nearby device by dragging it 'through' a portal over the device's edge. Building on this research, Jokela et al. [184] compared a technique called *Tray* to two techniques that rely on physical contact by participants with devices (*Transfer Mode*) and devices themselves (*Device Touch*). Participants liked the visual feedback and familiar handling of *Tray*, performed faster with the *Transfer Mode*, and considered *Device Touch* as rather cumbersome. Similar to *Tray* and *Portal*, Scott et al. [361] compared their *Bridge* and a *Pick and Drop* technique for object transfer between a tabletop and mobile devices. Both techniques effectively supported participants in their tasks. *Bridge* resulted in a lower cognitive effort, whereas *Pick and Drop* was seen as more efficient as participants could pick multiple items sequentially and drop them all at once.

Rekimoto originally described *Pick and Drop* [333]: The technique allows to pick up an object with a pen on one display and release it at another device by touching the display – creating the illusion of manipulating a physical object. Paay et al. [301] compared their variant of the *Pick and Drop* technique called *Pinch* with three other techniques called *Swipe*, *Swing*, and *Flick* for interactions between a mobile device and a large display using additional mid-air gestures, taking a first step to find relative strengths and weaknesses of the named interaction techniques.

The versatility of tablets and the variety of interaction techniques inspired us to study the usage of tablets in combination with a set of interaction techniques in Chapter 5. The chapter uncovers relationships between the usage of tablets and interaction techniques in a two-task study design that includes complementary angles of evaluation.

While Cross-Device Interaction is often related to the combination of devices that relate to Weiser's envisioned components (e.g., phones, tablets, or interactive surfaces) of homogeneous device classes, the term Hybrid User Interfaces [97] is used to describe interfaces that combine heterogeneous components. Although the core research of this thesis is focused on combinations of homogeneous device classes, I am involved in designing and evaluating heterogeneous combinations (i.e., Hybrid User Interfaces), and they further contribute to the vision of Complementary Interfaces.

2.4 Hybrid User Interfaces

Feiner and Shamash [97] coined the term Hybrid User Interfaces in the early 1990s. Their ideas can also be, to some extent, regarded as an answer to desktop and personal computers. Feiner and Shamash [97] argue that with the advent of portable computing devices (i.e., laptops), the physical device sizes decrease, which in turn also leads to less interface real estate – including both the output space with smaller display sizes but also the input space with integrated touchpads instead of a "free-ranging mouse" [97]. While these conventional technologies tend to decrease in physical size, they still provide high-resolution and high input accuracy e.g., via mouse input. Immersive technologies like augmented reality on the opposite – in the early 1990s and still partly today – offer a lower resolution for both, input and output space. However, augmented reality holds the potential for virtually large interfaces that exceed the bespoke conventional display technologies. Feiner and Shamash [97] propose to combine these two technologies – by *"taking advantage of the strong points of each"* [97], treating the technologies as "complementary" instead of competing – leading to Hybrid User Interfaces that e.g., can extend the physical screen size of laptops with large virtual content using augmented reality head-worn devices. Tasks that require higher resolutions and accuracy can benefit from the conventional display, while tasks that benefit from a larger virtual space can be moved to the augmented reality space. Feiner and Shamash [97] propose a window managing system to illustrate their concept, showing how actively used content can be worked with on a conventional display, while other windows and applications can wait for their integration into the workflow in the virtual periphery that is aligned with the physical display, allowing for seamless interaction with both technologies. Similarly, Hybrid User Interfaces were studied in the late 1990s and early 2000s e.g., for mobile navigation scenarios [98] and co-located collaboration [31, 54].

Due to recent hardware developments, Hybrid User Interfaces are experiencing a renaissance, showing their relevance and applicability in a variety of use cases, by combining head-worn augmented reality devices with smartwatches (e.g., [123]), smart-

Figure 2.4: The setup of Feiner and Shamash [97]. A Hybrid User Interface combining a head-worn device for augmented reality and a laptop.

phones (e.g., [159, 402]), tablets (e.g., [155]), interactive surfaces (e.g., [53]), and desktop computers (e.g., [157]). The synchronous usage of Hybrid User Interfaces can lead to issues like the split attention effect [114] or to an increase in cognitive demand that was shown in general for multi-device environments [325]. Using Hybrid User Interfaces asynchronously opens up the design space for transitional user interfaces [156], yet leading to specific challenges e.g., loss of context or consistency tradeoffs.

Recent developments in technology suggest bringing these ideas out of research labs and into the consumer market – here, Apple announced its mixed reality device Vision Pro [12] to not only allow for immersive experiences but also to integrate with existing computing devices e.g., by extending screen spaces. Yet, it is necessary to further study these environments, especially as the design space is huge and lacks a clear definition and terminology (e.g., Hybrid User Interfaces vs Transitional Interfaces vs Mixed-Immersion Interfaces vs Cross-Reality Interfaces vs Augmented Displays vs Virtually-extended Screen-aligned Displays) – leading to a fragmented research landscape [158].

Brudy et al. [50] included head-worn AR/VR devices as part of their cross-device taxonomy but did not further elaborate on heterogeneous device combinations as described in this section. Hybrid User Interfaces could therefore be seen as a spe-

cial case of Cross-Device Interaction. However, Hybrid User Interfaces differ in the relationship of components: In Cross-Device Interaction, research has studied how additional devices can be included ad hoc into workflows e.g., by offering additional devices for an activity (e.g., [49, 316]). Users are free to use the extra space or functionality to e.g., externalize ideas to another device, yet they do not have to – leading to a low functional interdependence of components, but potentially to a high task-related interdependence of components. In Hybrid User Interfaces, the relationship between components is part of the design of the interface itself: Here, all components show a tightly coupled relationship, leading to a functional interdependence, as e.g., one component acts as input and the other component as output (e.g., STREAM uses a spatially-aware tablet for input and a head-worn AR device for close-to-unlimited visual output [155]). Therefore, Hybrid User Interfaces and Cross-Device Interaction can be distinguished by the classes of devices (e.g., heterogeneous or homogeneous device classes) that lead to different levels of immersion and device usage classified by the Milgram-Weiser Continuum [283] but also regarding the task-related and functional interdependence of components.

Chapter 5 integrates the idea of task-related and functional interdependence into a user study setting, where participants solve one task with a controlled interdependence of components to see how this is transferred into a second task with uncontrolled interdependence. Also, designing and evaluating meaningful combinations of heterogeneous devices and modalities is one of my ongoing research activities, which opens up a new dimension for Complementary Interfaces and will be further discussed in Chapter 7.

The two previous sections on Cross-Device Interaction and Hybrid User Interfaces have shown technical realizations of multi-device environments that run in parallel to Weiser's vision of Ubiquitous Computing. They focused on the combinations of non-immersive-only components but also of non-immersive with immersive components. Yet, multiple input and output modalities can also be used in combination to solve a task at hand. Therefore, the next section will first look at Multimodal Interaction in general and then direct the reader to Chapter 4 to further study the influences of input and output modalities on spatial memory as another potential dependent variable to study in Ubiquitous Computing environments.

2.5 Multimodal Interaction

Multimodal Interaction often refers to using multiple input modalities, such as voice, gesture, gaze, and touch, to interact with interactive systems. As humans, we interact with our surroundings in a multimodal way (e.g., using modalities in parallel or sequentially), therefore, Multimodal Interaction promises to build on pre-knowledge to eventually provide a high usability and user experience [394]. Likewise, we can use multiple output modalities of an interactive system (e.g., visually or auditory) to address different perceptual channels of users (e.g., to provide feedback). Here, the visual and auditory systems can be leveraged in a complementary way by e.g., by mapping spatio-temporal properties to a specific channel, which can e.g., reduce visual clutter or minimize attention switches [433] but also increase the available information and attention [108].

We often interact in multi-device environments (i.e., Cross-Device Interaction and Hybrid User Interfaces) in a multimodal way: Brudy et al. [50] classify input modalities for Cross-Device Interaction as touch (on-screen), gestures (around the device or body gestures), and device manipulation (motion in 2D/3D or shape change). Hybrid User Interfaces are often multimodal per se, as each component might require its own type of interaction – e.g., a desktop environment requires different input techniques than an immersive environment (e.g., AR or VR), yet they can be described as an asynchronous Hybrid User Interface [156]. Similarly, using a combination of heterogeneous devices (e.g., a tablet and a head-worn device) simultaneously might require gestural and touch input (cf. [155]). In general, the advent of consumer-grade head-worn devices to experience mixed reality expanded the catalog of Multimodal Interaction from meanwhile well-known voice commands (e.g., Siri, Alexa, and friends) to e.g., combinations of eye and hand input (cf. Apple Vision Pro), which is also a topic of interest for recent research (e.g., [249, 404]).

Research has demonstrated benefits of Multimodal Interaction regarding accessibility [299], flexibility [229, 298], task performance [300], cognitive workload [439], social presence [100], and it can even lead to the perception of "magic" [27].

Similarly, research has shown that the choice of input or output modalities can influence users' spatial memory [348]:

Tan et al. [384] focused on the influence of touch and mouse interaction on users' spatial memory. They asked 28 participants to perform a memorization task on an $18.1''$ screen with an invisible grid and concluded with a 19% increased memorization accuracy for touch input. Building on this work, Jetter et al. [179] investigated the effect of multi-touch vs. mouse input on participants' navigation performance and object location memory in *panning* and *panning & zooming* interfaces. Results show

that multi-touch results in better spatial memory and navigation performance for *panning* interfaces but not for *panning & zooming* interfaces. Palleis and Hussmann [304] distinguished between the influence of kinesthetic cues and direct interaction on spatial memory and navigation performance. Results show that neither of these aspects influences spatial memory, but a smaller touch input surface in the indirect touch condition resulted in a better navigation performance.

Focusing on whole body movements, Klinkhammer et al. [201] compared users' spatial memory for panning using single touch vs. body panning on a horizontal user interface. Although body panning (i.e., panning via body movements) did not result in an increased navigation performance, participants performed better with regard to object location memory. Similarly, Rädle et al. [321] investigated the effect of egocentric movements in peephole navigation of zoomable peephole interfaces, concluding with a positive effect of egocentric body movements on spatial memory.

The above-mentioned studies indicate that users' spatial memory can be positively influenced by a suitable input modality, however, results on the characteristics of the optimal input modality are contradicting. Conditions that used touch performed better than mouse [179, 384]. However, when replacing the mouse with an indirect touch modality, findings show no difference in spatial memory and an improvement in navigation performance for the indirect interaction on a small input surface [304]. Conditions including whole body movement (e.g., body panning, egocentric body movements) performed better than those with less movement [201].

Previous research investigated how different visualization techniques influence spatial memory. For example, fisheye lenses [366], peephole navigation [190, 274, 324] and systems that allow for an overview [151] or focus and context view [52, 292] have been analyzed to further understand how these techniques impact on navigation and object relocation. Visualizations like grids and structural alignments of items on a landscape have been found to positively effect spatial memory [233, 348].

Analyzing the influence of the output modality (e.g., screen size) on spatial memory has shown that large displays lead to a greater sense of presence and a higher level of immersion, which positively influences users' spatial memory performance [385, 395]. For example, Tan et al. [385, 386] showed that physically large displays increase performance on spatial tasks such as 3D navigation, mental map formation, and memory tasks. Further, they found that the benefit of using a large display strongly depends on the kind of tasks performed on it: The benefit of large screens was found to be restricted to tasks in which users have to take on an egocentric strategy [385]. However, this finding could not be replicated by Tyndiuk et al. [395], who found that large displays are beneficial for both, ego- and exocentric tasks.

Tan et al. [386] also investigated users' spatial memory performance when navigating a virtual environment. Results show that users perform better in mental map formation and memory tasks on a large display because they feel more immersed in the environment compared to a small screen. Tyndiuk et al. [395] investigated the influence of display size on spatial memory and showed that not all users benefit from the use of large displays as only users with high visual selective attention abilities performed better on large displays. Rädle et al. [324] focused on peephole navigation and analyzed the effect of peephole size on users' map navigation behavior, navigation performance, and task load. They confirmed a sweet spot of a tablet-sized peephole where learning speed, navigation speed, and task load were improved.

As previous research has shown the influence of different input and output modalities on spatial memory, Chapter 4 sheds light on the interplay of different input and output modalities and how these meaningful combinations can affect spatial memory.

The previous sections described how combinations of homogeneous and heterogeneous device combinations and various input and output modalities can influence the user experiences of interactive systems. Taking a step back from these technical realizations allows us to see the bigger picture of Complementary Interfaces – meaningful combinations of devices and modalities – that are further illustrated in the next chapter.

2.6 Towards Complementary Interfaces

For typical knowledge work activities like writing documents, preparing presentations, or calculating costs in spreadsheets, we do rely on the affordances, constraints, and utilities of e.g., different input devices. A mouse for pointing and selecting, a keyboard for text input. While this approach is working for desktop computing, many post-WIMP[3] [398], Ubiquitous Computing [409], and related interfaces (e.g., Cross-Device Interaction) build on self-contained devices, trading complementary peripherals with the convenience of built-in touch interaction and a combined input and output space. This can be advantageous when working on a single device (e.g., for mobility), but if task complexity increases, single devices may no longer be sufficient to adequately support users in their workflows – despite extensive device capabilities.

The technological and methodological advances within the last decades allow researchers to design and evaluate new interaction paradigms beyond the boundaries

[3] post-"**W**indows **I**cons **M**enus **P**ointer".

of a single device and modality, leading to a variety of combinations of interfaces that can be used seamlessly in concert. Yet, simply adding more devices can be counterproductive, as it might not fit to workflows or activities [316], handling multiple devices can increase cognitive workload [325], or lead to high transaction costs [153].

By *pushing the boundaries* of technological possibilities, designers and researchers often try to sophistically blend multiple devices together, following a "the more the merrier" approach – paying the price of potentially not suiting the context of use, users' activities, or scenarios [50], instead of utilizing or "embracing" the individual qualities of each component.

Interestingly, in 1991, two papers pointed to this idea from slightly different angles: Feiner and Shamash [97] described their Hybrid User Interfaces as "complementary [...] technologies [...] that take advantage of the strong point of each" and Weiser [409] highlighted that "the real power [...] comes not from any of the[se] devices; it emerges from the interaction of all of them." Both papers emphasize the importance and benefits but also the dependency of the combination of multiple interactive components. Feiner and Shamash [97] underline the utility of each component in a Hybrid User Interface and Weiser's "it emerges from the interaction of all of them" [409] can be interpreted as aspects of progress, activities, or workflows.

Indeed, in a successful ecology of devices and modalities, each interface component possesses *complementary* characteristics, filling a niche that was not suitably covered before – supporting users in achieving their goals and accomplishing their tasks.

Based on this, we believe that attributing unique roles, properties, and purposes to each device and modality can lead to a *meaningful combination of interfaces that support users in their current task at hand* – emphasizing the importance of users, their task, and how they can be supported.

We call these meaningful combinations of devices and modalities **Complementary Interfaces**: By distributing interaction across devices and modalities, we can establish a *symbiosis of interfaces*, where each component purposefully increases the quality of interaction and supports users in their current activity. Fundamentally, Complementary Interfaces draw their strengths from meaningful combinations across and within the dimensions input and output; and they are dependent on the context.

A Complementary Interface is a meaningful combination of interfaces that support users in their current task at hand.

Figure 2.5 further illustrates this concept: A Complementary Interface strives for meaningful combinations of one or multiple **Input** components and one or multiple **Output** components (each within and across dimensions) to allow one or multiple users to interact with an **Application** that is embedded in a **Context** and shares some connection to the **Rest of the World** – representing a classic model in Human-Computer Interaction where a user is involved in handling an application by an iterative process of input and output activities. As indicated in Figure 2.5, both, **Input** and **Output** can each be composed of single or multiple components (e.g., multiple input modalities and/or multiple output channels). The key to Complementary Interfaces is the meaningful combination of components, where each component can be seen as a piece of the puzzle. This leads to combinations of homogeneous (e.g., Cross-Device Interaction) and heterogeneous (e.g., Hybrid User Interfaces) device classes, but also input (e.g., interaction techniques) and output (e.g., visually or auditory) modalities. Complementary Interfaces may feature some degree of heterogeneity of involved components that complement each other to support the overall system functionality to solve the task at hand. These degrees of heterogeneity may lie in the input or output modality, location (e.g., screen space or input space), or dimensionality of data visualization (e.g., 2D or 3D). However – depending on the **Context** – to solve a task, a user might increase the input space or screen space by e.g., simply adding extra tablet devices – creating a Complementary Interface with components that might not show any degree of heterogeneity at all but support the user during the current activity, which creates a task-related interdependence of components. Likewise, Complementary Interfaces can also feature a functional interdependence of components, e.g., when one component is solely used for input and another component is exclusively used for output (cf. Hybrid User Interfaces). We often see a task-related interdependence for Complementary Interfaces that have multiple components either for input or output and a functional interdependence for Complementary Interfaces with a combination of single input and output components. Yet, the degree of interdependence might be dynamic, depending on the context and workflow.

To following paragraphs will further describe each of the dimensions **Input**, **Output**, **Context**, **Application**, **Rest of the World.**

Input The combination of multiple input modalities for interaction is often considered as Multimodal Interaction (see Section 2.5). Here, a wide variety of input combinations can influence the user experience: Taking the example of combining gaze and hand input from above [404], a human's eyes and hands can form a symbiosis of interfaces by e.g., explicitly selecting a target via gaze and confirming it with an additional hand or finger gesture (cf. Apple Vision Pro). Alternatively, the input by a user's eye gaze can be utilized as an implicit input (e.g., different reading patterns as

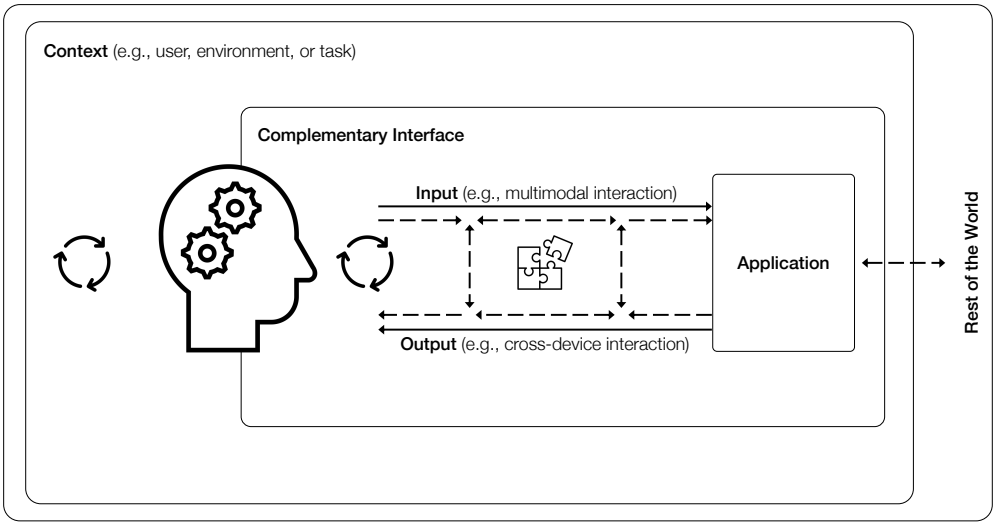


Figure 2.5: An abstract model of the concept of Complementary Interfaces. The figure is based on the implicit Human-Computer Interaction model [353], yet adapted to illustrate the concept of Complementary Interfaces.

indicators for language proficiency [186]) to adapt a user interface that then allows the user to interact with it explicitly.

Output Although off-the-shelf mobile devices such as tablets have a combined input and output space, the physical output size can differ or be extended. This also relates to traditional multi-display environments. Yet, using tablets and other mobile devices potentially allows for ad hoc scenarios as initially envisioned by Mark Weiser [409]. Here, extra devices can extend the space to think (cf. [7]), allow to divide and conquer a problem [325], or to share insights in collaborative activities (see Chapter 3). Further, the output can differ regarding the number of dimensions: Hybrid User Interfaces that e.g., combine 2D and 3D visualizations enable analysts to gain complementary insights into data [157]. Also, multiple output channels can complement each other in (remote) collaborative scenarios, where a limited visual field of view of augmented reality head-worn devices can be complemented using spatial audio to increase the sensation of presence [100].

Context A Complementary Interface is embedded into a context: Schmidt et al. [355] divide a context into (1) Human Factors (e.g., user, environment, or task) and (2) Physical Environment (e.g., conditions, infrastructure, or location). The physical environment certainly influences the availability of e.g., interactive components and handling them. Yet, for Complementary Interfaces, the human factors play a major role. Here, as common for human-centered approaches, users and their current activity are the focus of attention. This includes a user's knowledge, preferences,

or emotional and physiological states and their current task at hand – both of these aspects are further framed by the (social) environment (e.g., the interaction with others). An activity that is placed in a social environment might lead to certain needs for each user. Here, different meaningful combinations might be suitable and useful to allow users to achieve their goals. For example, showing recent photo albums on a shared screen to family and friends comes with different requirements than collaboratively presenting the outcome of an investigation as a slideshow in a professional work environment – here, each scenario allows for different meaningful combinations of input and output components to best support users in the current activity.

Application and Rest of the World A Complementary Interface needs an application that can cope with a variety of input and output components. Yet, as research prototypes are usually tailored for an empirical user study, research presented in this thesis does not focus on universal interaction with any given application. Potentially, a given task requires a specific application that further requires a well-balanced set of input and output components. Similarly, a Complementary Interface is embedded in a context that is further part of the rest of the world. This allows e.g., for communication with others that are not involved in the given activity, but has less priority for Complementary Interfaces that are presented in this thesis.

The described dimensions frame the concept of Complementary Interfaces and allow to use it in two intertwined ways: (1) As a conceptual design framework for composing meaningful Complementary Interfaces that have the potential to create a *symbiosis of interfaces*; and (2) as an evaluation framework to study the effects and influences of different component combinations (e.g., design alternatives). Both of them put an emphasis on users, their task at hand, and how to support this with meaningful combinations of input and output components. This eventually allows to design task and user specific interactive systems and further *study the utility of meaningful combinations.*

The concept of Complementary Interfaces aligns with existing frameworks that support designers and researchers to design, study, or explain Ubiquitous Computing environments or interactive spaces in general (e.g., Blended Interaction [180], Reality-Based Interaction [172], or other frameworks that make use of users' pre-existing knowledge via blends or metaphors). Reality-Based Interaction (RBI) [172] introduces themes of reality (i.e., näive physics, body awareness and skills, environment awareness and skills, and social awareness and skills) and provides means to analyze, compare, and evaluate design tradeoffs [117]. RBI suggests anchoring interaction on the named themes to build on existing pre-knowledge of the real world and users' skills, thereby reducing the mental effort to use an interactive system [172]. Yet, Jacob et al. [172] highlight that an interactive system requires design tradeoffs be-

tween these themes and desired qualities (e.g., efficiency or ergonomics), depending e.g., on the current context. Jetter et al. [180] describe these themes of reality as "basic-level experiences and building blocks of human thought" [180] and consider them among other aspects (e.g., sensorimotor experiences or well-established digital concepts) as part of a user's reality in their Blended Interaction (BI) framework. BI strives for "considerate blends" [180] between this existing knowledge of a user about reality and the "magic" of the computational power of today's user interfaces. Jetter et al. [180] emphasize that designers have to find "good design tradeoffs" (cf. [172]) that are situated between the two extremes of (1) very powerful – yet hard-to-learn – concepts and (2) familiar real-world concepts that are easy-to-learn but e.g., lack precision or efficiency. Therefore, Jetter et al. [180] refer to embodied cognition [113, 363] as their theoretical foundation and describe how conceptual integration [96] facilitates the learnability and ease of use of an interactive system – by *blending* new concepts from existing ones. Ideas from embodied cognition and related theories also relate to the concept of Complementary Interfaces. Here, cognition is not only embodied (e.g., as part of multimodal interaction techniques), but can also be distributed "across humans and [interactive] artifacts" [242] as well as "resources and materials in the environment" [144]. The meaningful combinations of input and output modalities aim to distribute cognition across the above-described dimensions to create tools that ultimately "disappear" [410].

The described frameworks can help to design and evaluate Complementary Interfaces by extending the notions of design tradeoffs and considerate (or *meaningful*) blends to different input and output components, which can increase the user experience. However, Complementary Interfaces strive for meaningful combinations of input and output components that can support users in working on the task at hand – showing a rather pragmatic perspective on interaction design that focuses on the utility of components.

This thesis presents three use cases of Complementary Interfaces: Chapter 3 contributes to a better understanding of the meaningful combination of individual tablets and a shared interactive surface in three different sizes during collaborative sensemaking. Chapter 4 sheds light on the interplay of different input and output modalities and how these meaningful combinations can affect spatial memory. Chapter 5 uncovers relationships between the usage of tablets and interaction techniques in a two-task study design that includes complementary angles of evaluation.

Designing and evaluating Complementary Interfaces can be challenging. The combination of multiple interactive components for input and output activities has to be carefully integrated into users' workflows. Here, simply adding another component can result in a *loss of context*, as users have to re-orientate themselves e.g., in the data. Similarly, *switching attention* between different devices can incur a significant overhead [331]. To overcome this, devices in the environment could tune their attention to the users and *adapt to the users' needs*. Yet, as each interactive component comes with its individual strengths and drawbacks, but also affordances and constraints, we need to provide ways to achieve a *consistent user experiences* during workflows. For collaborative scenarios, this also includes the *continuity of user representations*. Ultimately, *finding suitable modalities* is a core challenge of Complementary Interfaces. Therefore, this thesis studies three use cases of Complementary Interfaces, integrating potentially suitable modalities as design alternatives – allowing to use experiments as design-informing activities.

Chapter 7 will transfer the described challenges of Complementary Interfaces into opportunities for future research with a focus on collaborative environments.

The previous sections focussed on the conceptual and rather design-oriented aspects of Complementary Interfaces. The remainder of this Chapter 2 puts a spotlight on empirical research methods in HCI and how experiments can be used as a design-informing activity for Complementary Interfaces.

2.7 Experiments as Design-Informing Activities

As a highly interdisciplinary research discipline, HCI research borrows research methods from various fields and adapts them to the individual context [230]. This adaptation of methods is an ever-moving target that goes hand in hand with the technological evolution of computing devices. This leads to new use cases for these technologies but also to new ways to assess the quality of interaction (e.g., eye tracking). While research in the 1980s often focussed on performance-related aspects (e.g., time and error), more recent research added an emphasis on human-related aspects (e.g., immersion or joy) – providing researchers toolboxes to assess the usability and user experience of interactive systems. Research proposed a variety of ways to evaluate the usability of simple interaction techniques, e.g., the selection of items using pointing devices (e.g., Fitts' Law [101, 252]). Here, aggregated performance metrics or raw measures (e.g., time and error) work especially well to evaluate simple interaction instances like pointing or dragging, yet they are usually insufficient to fully character-

ize more complex activities in Complementary Interfaces (or other post-WIMP and Ubiquitous Computing environments) – resulting in an elaborated interaction that exceeds standardized user study settings [230].

Brudy et al. [50] classified evaluation strategies for multi-device environments: Evaluation through (1) demonstration, (2) usage, (3) heuristics, (4) informative studies, and (5) technical evaluation. The authors do not break down these strategies into specific research methods (e.g., interviews), yet, they can be related to usability testing (1 and 2), analytical methods (3 and 5), and experimental research (4). Usability testing is usually a formative evaluation method to e.g., catch design flaws during the design process. Here, a small number of participants (e.g., 6 to 8) are invited to use an interactive system. Identified usability problems and feedback are addressed during the iterative design process [137]. Analytical methods have a similar goal. However, for analytical methods, experts identify problems (e.g., based on heuristics), prioritize them, and report them, which also feeds the iterative design process. Experimental research follows a different approach. The goal of experiments is not to identify problems with an existing interactive system to improve its design but to "identify causes of a situation" [230] – or in other words: **to understand cause and effect**.

Does this make experiments less design-oriented? Experiments allow us to study how design alternatives of an interactive system (e.g., interaction techniques) influence certain key aspects (e.g., user satisfaction) to eventually answer research questions. Experiments follow the *ceteris paribus* principle [149, 296], which means that in highly-controlled study situations, only one aspect (e.g., the interaction technique) gets varied by the experimenter, while all other factors (e.g., study situation or measurements) stay the same. In a typical experiment, e.g., one group of participants completes a pre-defined task with one interaction technique and another group of participants completes the same task with another interaction technique, while the experimenter measures the resulting user satisfaction of using the individual interaction technique to complete that task. With this, a causal relationship can be established: In this example, a typical finding would be that one interaction technique might result in a higher user satisfaction than the other, as every other factor was the same. This cause-and-effect relationship allows us to define implications for future research and design, making **experiments a design-informing activity** [296].

This thesis follows the idea of using experiments as a design-informing activity for three use cases of Complementary Interfaces in Chapters 3, 4, and 5.

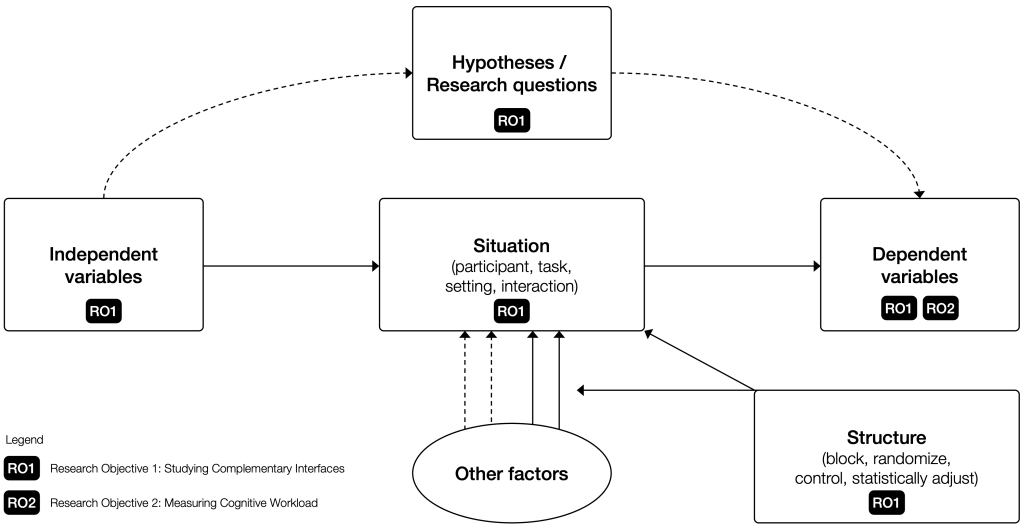


Figure 2.6: Typical components of experiments in HCI (adapted from [149]).

This thesis uses experiments as a research method to study design alternatives for three use cases of Complementary Interfaces. Following Figure 2.6, each component of experiments in HCI will be further described below. These typical components serve as a framing regarding the study situation itself (e.g., tasks and settings), factors regarding the structure of experiments (e.g., procedures and study designs), and dependent variables (e.g., qualitative and quantitative measurements). In total, three experiments studying Complementary Interfaces and a meta-analysis of assessment methods of cognitive workload (as a potential dependent variable of experimental settings) are presented in this thesis (see Research Objectives 1 and 2). Each chapter will align its individual experiment accordingly.

2.7.1 Experiments and their Components

Before describing individual components of experiments, a definition of an experiment is presented:

Experiment: A study in which an intervention is deliberately introduced to observe its effects [74].

Cook et al.'s definition allows to split an experiment into three essential parts: For typical experiments in HCI, (1) the *intervention* can be an interaction technique, but also participants' pre-knowledge or a task. (2) The *observed effects* are related to the measurement. (3) The experimental situation of the user *study* allows to expose participants to an intervention and observe its effects. Hornbæk [149] builds on this

and provides a pragmatic-yet-detailed description of some *dos and don'ts of designing experiments* and further breaks down the three named parts into a set of components that is shown in Figure 2.6. For this thesis, the focus is placed on providing the reader with an overview of all components.

Hypotheses and Research Questions are theoretical connections between independent and dependent variables. Both of them "connect variation in independent variables to expectations about variation in the dependent variables" [149]. Hypotheses are assumptions backed up with literature and therefore, build on prior work: Data from earlier findings or theories might allow us to predict the effect of an intervention. Thus, hypotheses must be "(a) testable, (b) concise, and (c) name key constructs" [149]. Similarly, research questions also name the key constructs of the experiment. However, research questions are formulated as questions and omit the dependency on prior work, which allows for rather explorative experiments if e.g., no prior work is available.

Independent variables are constructs that are not directly exposed to participants. Yet, independent variables are operationalized and represented as multiple levels to participants: Conditions. A typical experiment in HCI consists of a single independent variable (e.g., input technique) that is broken down into different conditions. Usually, a baseline condition refers to the state-of-the-art approach that is challenged by a new technique. Often, the baseline condition works as a control condition that allows to study the effects of a novel approach in comparison. This can be critical as the choice of conditions might bias the experimental situation and measurements. Therefore, the comparability of conditions is essential for experimental study design.

Experimental Situation is the actual situation under study. Here, (1) participants use one or more of the conditions to complete some sort of structured (2) task in a certain (3) environment, while the experimenter uses various data collection methods to assess participants' activities and results. The selection of (1) participants can influence or even bias the outcome of the experiment. Their pre-knowledge, experiences, abilities, and motivation can influence the measurements. Therefore, the selection of participants should be based on the research questions, which further allows to generalize the effects. The number of participants can be estimated using methods such as power analysis. Yet, a common practice is to refer to local standards for sample sizes in HCI experiments [56], which usually leads to approximately 20 participants per condition for typical experiments in HCI. During the experiment, participants complete a (2) task. These tasks are usually designed due to the goal of the experiment and can range from simple pointing or selection tasks (e.g., Fitts' Law [101]) to knowledge work activities that take hours to complete. Tasks should be representative and a valid activity for the scenario. Participants have to under-

stand the task to avoid confounding factors, and some research has used the task as an additional independent variable in the study setting. Using well-established tasks potentially allows for comparisons across individual experiments. Typical (3) environments for experiments include lab settings that allow to control all environmental aspects and minimize external influence to collect fine-grained data to achieve a high internal validity of data. Alternatively, field settings allow for more realistic scenarios, yet they are prone to environmental noise (e.g., disturbance of participants).

Dependent variables are the counterparts to independent variables. These constructs have to be operationalized in a similar way. Usually, hypotheses or research questions mention the dependent variables. Typical dependent variables of experiments in HCI are e.g., time, workload, or accuracy. These constructs can be measured e.g., using data logging, questionnaires, or concluding interviews. The resulting data can be statistically analyzed to identify statistically significant differences between the effects of independent variables. This, in turn, allows us to draw conclusions and answer research questions.

Structure and Other factors are essential for experiments in HCI. Experimenters purposefully define the structure of an experiment up front. A typical experiment includes multiple conditions and tasks. In general, we can distinguish between between-subjects design and within-subjects design. For between-subjects experiments, each participant experiences only one of the conditions. For within-subjects experiments, each participant experiences all of the conditions. Having the same order of conditions and tasks for all participants can lead to problems such as learning effects or fatigue. To overcome these issues, experimenters can randomize the order of conditions or use counterbalancing methods (e.g., Latin square). Other factors, such as participants' demographics, can further influence the experimental situation. Yet, counterbalancing methods can help to minimize their influence on the outcome.

Work presented in this thesis purposefully *experiments with experiments*: All three use cases for Complementary Interfaces include **research questions** to frame their research. Research questions were chosen instead of hypotheses as prior research either was only taken as motivation to conduct the research or findings by prior work were contradictory. Also, the use of research questions allows for more explorative research, which is represented in Chapters 3, 4, and 5. The **independent variables** differ for each use case. The individual independent variables are the size of a shared interactive surface for collaborative activities (Chapter 3), the type of input modality and size of the output modality (Chapter 4), and the interaction technique and task (Chapter 5). Each use case consists of its own **experimental situation**. The number of participants was selected based on the number of conditions and the chosen counterbalancing method. Participants were invited using flyers and posting at the

University of Konstanz, leading to a majority of students as participants. For collaborative activities, great care was taken to find dyads that did not know each other (i.e., to avoid confounding factors) and to achieve an equal gender distribution per condition. All tasks are representatives of knowledge work activities, and each of them is well-established and was previously used by related works. For Chapter 3, a collaborative sensemaking task based on the IEEE VAST challenge [315] was included. For Chapter 4, a well-established task to study spatial memory was slightly adapted to fit the study apparatus and research questions. For Chapter 5, a well-established sorting activity was adapted and complemented with an additional sensemaking task. All experiments were conducted in a lab environment to control for a high internal validity. The **dependent variables** for each experiment with a Complementary Interface in this thesis are individually targeted based on the research questions and independent variables. However, all of them include qualitative and quantitative measurements that allow for data triangulation. Chapter 6 is a special case: Here, no experiment is conducted, yet, the chapter puts a spotlight on the measurement of cognitive workload – a highly relevant metric for Complementary Interfaces. Each experiment follows its own **structure** e.g., to account for other factors. Here, the experiments purposefully use a between-subject design (Chapter 3), a within-subjects design (Chapter 5), and a combination of both experimental designs (Chapter 4).

The experimental design for each Complementary Interface is described in a higher level of detail in each corresponding chapter.

2.8 Chapter Conclusion and Outlook

This chapter described the theoretical background of this thesis. The historical evolution of computing devices led to the description of Weiser's vision for the computer of the 21st century: Ubiquitous Computing [409]. Based on this, relevant research streams that run in parallel to this vision were used to illustrate and further frame the research described in this thesis: Cross-Device Interaction, Hybrid User Interfaces, and Multimodal Interaction conceptualized the vision of Complementary Interfaces – meaningful combinations of interfaces that support users in their current task at hand. The chapter concluded with a new perspective on empirical research and how experiments can be used as a design-informing activity.

The second part of this thesis will showcase three use cases of Complementary Interfaces, describe their experiments, and report on findings. The third part of this thesis will put a spotlight on the measurement of cognitive workload as a potential metric for Complementary Interfaces. Part four concludes this thesis and provides an outlook.

Part II

Studying Complementary Interfaces

Parts of the following Chapter 3 have been published as:

***Johannes Zagermann**, Ulrike Pfeil, Roman Rädle, Hans-Christian Jetter, Clemens Klokmose, and Harald Reiterer. 2016. When Tablets Meet Tabletops: The Effect of Tabletop Size on Around-the-Table Collaboration with Personal Tablets. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), ACM, San Jose, California, USA, 5470-5481. DOI: https://doi.org/10.1145/2858036.2858224*

The responsibilities of this joint publication were divided as follows: I formulated the research questions, designed and implemented the research prototype, designed and conducted the user study, analyzed the study data, and spearheaded the writing. Ulrike Pfeil helped in formulating the research questions, analyzing the study data, and writing the paper. Roman Rädle, Hans-Christian Jetter, and Clemens Klokmose equally helped in formulating the research questions and writing the paper. Harald Reiterer supervised the work.

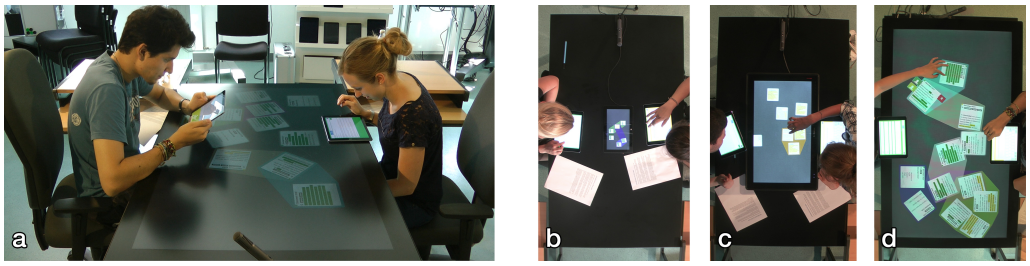
Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

Video Figure

Conference Talk



In this Complementary Interface, dyads can engage in knowledge work activities – freely transitioning from loosely-coupled parallel work to tightly-coupled activities and back – leading to mixed-focus collaboration [128, 387]. While their personal device (i.e., a tablet) enables them to search for documents, read, and annotate them, a complementary shared interactive surface provides opportunities to visually and spatially cluster ideas, share insights, and discuss possible solutions. The meaningful combination of personal and shared devices is tailored towards knowledge work activities and is designed to support dyads within and across all phases of their process [314].

Study Design:	Between-Subjects
Task:	Sensemaking Task
Independent Variable:	Size of Shared Interactive Surface
Dependent Variables:	Interaction with Devices, Communication, Group Awareness, User Satisfaction

---

The described Complementary Interface and experimental setting are used to address the following general research question. The chapter will integrate this general research question as its overall research goal and further address it in specific use-case related research questions.

RQ1: *What are the influences of the size of a shared interactive surface on collaboration in a Complementary Interface?*

3

Studying Multi-User Interaction

Contents

Shared interactive surfaces (e.g., tabletops) can greatly benefit collaborative activities [168, 178, 270], mainly because of enabling natural face-to-face communication around a shared workspace [172], providing a high level of awareness and fluid interaction [152, 335], and supporting natural territoriality during collaboration [360]. However, a large surface is also costly and difficult to maintain compared to other common off-the-shelf devices such as tablets. This is particularly true for environments like libraries or classrooms in which these large surfaces are often unavailable due to budget and space restrictions, whereas tablets are quite common. Some researchers, therefore, aim to support knowledge work activities (e.g., sensemaking) with mobile devices *instead of* these large interactive surfaces [130, 246, 323], while Tang et al. [387] recommend combining mobile devices *with* them. Such systems (e.g., [264, 293, 322, 405]) are thought to be helpful for *mixed-focus collaboration* [128, 387] where users frequently transition from tightly-coupled collaboration to loosely-coupled parallel work, e.g., for reading or annotating a document or reviewing a long list of items on their own. Indeed, Wallace et al. [405] observed that in such scenarios, the presence of a shared surface leads to better sensemaking performance than with tablets alone.

To better understand why, we built a Complementary Interface that allows dyads to engage in sensemaking activities. The meaningful combination of tablets for personal activities (e.g., reading) and a shared interactive surface for collaborative activities

(e.g., sharing and discussing insights) allows for a mixed-focus collaboration [128, 387] and covers all phases of the sensemaking process [314].

We used this Complementary Interface as the apparatus for an experiment. Here, we closely studied the effect that three interactive surfaces with different sizes had on users' attention, awareness, and efficiency during mixed-focus collaboration. Investigating these effects enabled us to understand size-related advantages and disadvantages and to compare large interactive surfaces with lower-cost, smaller devices that could potentially serve as design alternatives. We conducted an experimental study of a collaborative sensemaking task that was executed by 15 dyads with two personal tablets (9.7″) and a shared workspace. We varied the size of the shared workspace over three conditions corresponding to off-the-shelf devices (10.6″, 27″, and 55″). To observe authentic collaboration during the study, we provided participants with our Complementary Interface for solving a complex analytical task based on the "Stegosaurus" data set from the VAST 2006 Challenge [121].

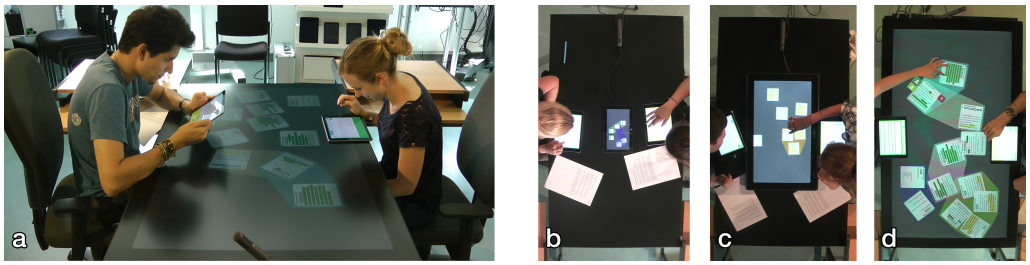


Figure 3.1: This Complementary Interface includes personal tablets and a shared interactive surface (a). Here, the size of the shared interactive surface is the independent variable of the experimental design (b,c,d).

Our goal for the study was to better understand the implications of the size of the shared interactive surface and to contribute a nuanced description of its role in collaborative sensemaking activities. We address this overall research goal by investigating the following research questions both qualitatively and quantitatively:

RQ1.1: *How does the size of a shared interactive surface affect users' interaction with the devices?*
RQ1.2: *How does the size of a shared interactive surface affect the communication between the participants?*
RQ1.3: *How does the size of a shared interactive surface affect group awareness, the quality of the results of the group work, and participants' satisfaction with the support by the system?*

3.1 Experiment

To investigate the effect of the size of a shared interactive surface on collaborative sensemaking, we designed and implemented a Complementary Interface that enabled dyads to search, read, and make sense of documents. The system served as an experimental platform for observing (close to) authentic collaborative behavior.

We chose a between-subjects design with the size of the shared surface as an independent variable with three conditions (SMALL: 10.6″, MEDIUM: 27″, and LARGE: 55″). We investigated five dyads in each condition, resulting in 15 dyads and 30 participants, respectively. Like Isenberg et al., we chose dyads, because in their study, two group members were already sufficient to identify differences in coupling styles and collaboration strategies [168].

3.1.1 Pre-Study

A pre-study informed our research questions, our study design, and the system design. It was conducted with an earlier version of our system, where 15 dyads were asked to work on a general search-related group task. The initial analysis of the collected video data indicated that the efficiency gain from a larger display was far smaller than expected – or even negative. To look into this phenomenon more systematically, we refined our system for the main study and focused the user interface and study design on our identified sub-topics: the interaction with the devices (RQ1.1), the communication activities (RQ1.2), and the quality of group work results and users' satisfaction with system support (RQ1.3).

3.1.2 Task

To provide an authentic and sufficiently demanding collaborative sensemaking task, we used the "Stegosaurus" data from the VAST 2006 Challenge [121] that was already used by prior research [7, 130, 168, 174, 282]. VAST 2006 describes a fictional story with 238 news articles, images, and data sheets. Dyads have to search and find relevant information from this data, filter non-important articles, and connect important facts that lead to new insights and suppositions to come up with a solution to a hidden plot. The task involves individual activities like searching for and reading relevant information as well as collaborative activities like sharing and sorting relevant documents to solve the case. The task is difficult enough to require participants to work together for an extended period of time (we limited it to 90 minutes).

We expected dyads to only partially solve the task within the time frame. However, participants were handed questions about the plot at the beginning of each session to focus their behavior and stimulate an authentic collaboration.

3.1.3 System Description

Our system supported dyads during all phases of individual and collaborative work. On their tablets, participants could execute full text searches in the VAST 2006 data by typing one or more keywords into a search box. Search results were presented in a ranked list (right tablet in Figure 3.2). Each item in this list represented a document and contained a document ID for clear verbal reference, a preview of sentences that contained the keyword(s), and icons indicating whether a document was already read, favored, or edited before. To access the document, users tapped on the respective search list item to open a document view (left tablet in Figure 3.2). This view enabled them to read and highlight text in a document, add comments to it, share it to the shared surface, or send it directly to the tablet of their partner, on which the document then was also displayed in a document view. Each participant was assigned with an individual color (green or yellow). This color coding was used for the background color of their tablet, as the highlight color of selected parts of documents, and as an indicator for the origin of each shared document.

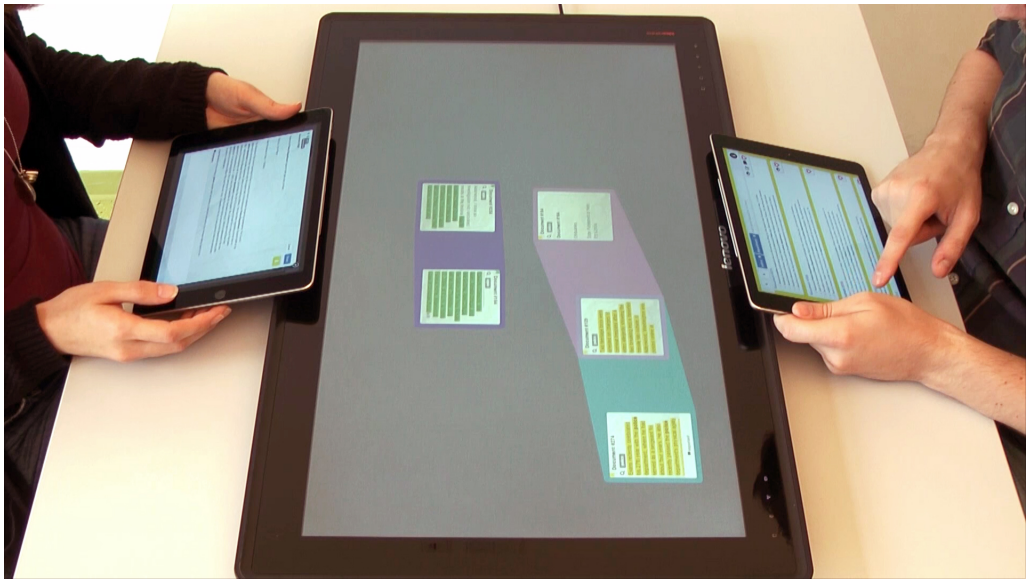


Figure 3.2: Overview of our Complementary Interface with two tablets (left and right) and one of the shared interactive surfaces (center).

Shared documents appeared on the shared surface in boxes containing the document ID, an excerpt of its content, and buttons for document-related actions (center of Figure 3.2). The possible actions enabled participants to re-read documents on their personal tablets or to remove documents from the shared interactive surface. Each box was additionally marked with a bookmark icon with a participant's color to indicate who shared the document. Boxes could be moved and rotated freely to allow for spatial arrangements or to form tableaux to embody the dyads' working

hypotheses [405]. Multiple boxes could be grouped into visual clusters using touch input to encircle them with a lasso. This created a colored convex hull around the selected boxes as a visual grouping. This hull remained around the boxes, and its shape was updated whenever the boxes were re-arranged by participants. For a rapid exchange of information, users could send documents to each other's tablet. Adding documents to the shared interactive surface enabled users to externalize ideas, store documents for later use, and to create a common ground for communication. Clustering of snippets helped to find connections between relevant events, persons, and places. The color coding provided a visual connection between the participants, their tablets, and their contributions to the shared interactive surface.

3.1.4 Apparatus

Figure 3.1a shows the experimental setup, which consisted of a conventional office table (1.4 × 0.8 m) on which the interactive surface that served as the shared space was placed. For the study, we chose three multi-touch devices of different sizes as shared interactive surfaces (Figure 3.1b, c, and d), representing the conditions of the user study: 10.6″ (Microsoft Surface 2 Pro), 27″ (Lenovo Horizon), and 55″ (Microsoft Perceptive Pixel). Each of them represented a typical size for a personal mobile device, a personal workspace, and a multi-user workspace. All three devices had the same display resolution of 1920 × 1080 pixels to guarantee for a high internal validity of the user study.

Both participants sat on office chairs and were facing each other. Each one was provided with an Apple iPad (9.7″) as their tablet. Next to each participant was a small mobile desk for filling out the questionnaire and as a space for notes or to place the tablet.

We used two cameras to record the study. One camera was installed with a bird's-eye view above the table (Figure 3.1 right), and a second camera recorded the scene from the side to be able to see gestures, postures, and interactions of both participants (Figure 3.1 left).

3.1.5 Participants

Thirty participants (15 female, 15 male, 0 non-binary) were recruited for the experiment. The mean age was 22.4 years (SD = 2.9, aged 15-28). Twenty-nine participants were right-handed, and one was left-handed. None of the participants had color vision deficiency, consequently, they had no problems with the employed color coding. Twenty-five participants were undergraduate students and 4 PhD students from non-technical subjects such as psychology, biology, or politics. One participant did not answer this question. We used postings and flyers looking for "detectives"

to recruit participants. By this, we hoped to find participants interested in puzzles, crime stories, or mysteries and thus were motivated by the VAST 2006 data and task. We assigned five groups of two participants to each condition. Two groups were both female, two were both male, and one group was mixed. Group members did not know each other and had not worked together before. The single underaged participant (aged 15) was a gifted student who studied at university at school age. The parents signed a consent form to assign this student all the rights of an adult, including study participation.

3.1.6 Procedure

At the beginning of each study, participants were seated at the two mobile desks. They were asked to fill out a questionnaire about demographics and tech-savviness. Then, they were seated facing each other on the long sides of the table. Participants received an introduction to all features of the system and were given time to explore its functionality with a sample data set until they felt comfortable enough using it. The experimenter then introduced them to the VAST 2006 task. During the 90-minute session, they could look up task instructions at any time on their tablet or on a printed sheet of paper. The experimenter answered general questions about the procedure or prototype but did not answer questions about participants' performance or how to solve the given task. Participants were free to move the mobile desks and office chairs as desired. The session stopped after 90 minutes or when participants felt that they had solved the case and found the plot. Then, both group members were asked to individually fill out a second questionnaire about their findings about the plot. Finally, we conducted a short semi-structured interview to ask participants about the strategy that they employed during the task and how much they felt that the system supported their activities. Each session lasted about 2 hours in total, and participants were compensated for their time.

3.2 Data Analysis

We employed several data collection methods to investigate our research questions: We gathered video data to analyze participants' collaborative behavior, data logs to analyze and quantify participants' interactions with the included digital devices, and a questionnaire and interviews to investigate participants' results of group work and to what extent participants felt supported by the system.

3.2.1 Video Data

To address the topics of our research questions, we developed a coding scheme to analyze our video data. In doing so, we focused on participants' interaction with

devices, their communication behavior, as well as the object (e.g., one of the devices) they focused on during the task. After initial viewings of our videos, we started to describe and distinguish between different codes that apply to our research questions, as well as offering a sound description of the video data. This coding scheme was refined until we felt that it was saturated. After this, an inter-coder reliability test was conducted for three 10 minute samples of video data to ensure objective coding by our two independent coders. Cohen's Kappa was calculated individually for the categories Focus ($\varkappa = 0.96$), Communication ($\varkappa = 0.85$), and Interaction ($\varkappa = 1.0$), revealing a high inter-coder reliability in each category for the final coding scheme that is presented below.

0.0 Focus: To investigate where participants placed their visual focus during the task, we distinguish between the codes *0.1 Tablet*, *0.2 Surface*, *0.3 Person* and *0.4 Other*.

1.0 Communication: To analyze the extent to which participants were talking, we code *1.1 Talk* and *1.2 Silence*. Utterances and mumblings were coded as *1.2 Silence*.

2.0 Interaction: To investigate participants' interaction with the shared interactive surface, we code *2.1 add document* for sending documents from the tablet to the shared surface, *2.2 delete document* for deleting documents from the shared surface, *2.3 move document* for moving documents to another place on the shared surface as well as *2.4 cluster documents* for participants' activities to collect documents in a cluster to sort them.

We used the coding scheme for both qualitative and quantitative investigations. In our qualitative analysis, the creation of the coding scheme helped us to elicit the main aspects of the collaborative work. Also, it allowed us to focus on specific sub-parts of the video for a more detailed observation. We could, for example, only view sequences when participants were focusing on the shared surface to further investigate this event by describing its context and related events in more detail. In addition, we also calculated the frequencies and durations for the coded events to investigate how the qualitative findings are reflected in the quantitative data.

3.2.2 Data Logs

In addition to the video data, we also collected data logs of users' interaction with the system. We logged the number of document manipulations (e.g., moving or rotating) on the shared interactive surface and how often documents were transferred from the shared surface to a tablet or vice versa. We also logged the number of times a document was transferred from one tablet to another. We did this to analyze whether the size of the shared surface had an effect on how and how often users interacted with documents.

3.2.3 Questionnaire and Interview

We used a questionnaire with questions about the solution (i.e., plot) of the VAST 2006 challenge to understand (1) the quality of the results based on individual answers to the questions and (2) the extent to which group members had the same knowledge of their results and solutions to the task. Also, we concluded each session with a semi-structured interview focusing on group strategies when working on the task and to what extent participants felt supported by the system.

3.3 Findings & Discussion

In this section, we report on our findings and discuss them in relation to our research questions. In particular, we focus on the differences across the three sizes of the shared surface (SMALL: 10.6″, MEDIUM: 27″, LARGE: 55″) with respect to users' interaction with the devices (RQ1.1), the group's communication activities (RQ1.2), and the differences regarding user perceptions and the quality of group work results (RQ1.3). For each research question, we elaborate on specific subtopics based on our qualitative observations and interview data. Quotes from participants in the following sections were slightly modified to prevent publishing the solutions of the VAST 2006 Challenge tasks.

In addition, we report our quantitative results in relation to our observations. In our analysis, we employed a mixed-effects model approach to analyze individual scores, taking participants' group membership as an additional source of variance into account. We used the nmle package by Pinheiro et al. [313] for R.

3.3.1 RQ1.1: Device Interaction

We distinguish between three aspects of participants' interactions with devices that could be influenced by the size of the shared surface: (1) participants' visual focus and attention, (2) participants' frequency of searching and transferring documents, and (3) participants' manipulation of documents on the shared surface.

Focus and Attention

In our qualitative analysis of video data, we observed that participants used the system mostly as we expected and intended in all three conditions. They used the tablets for searching or reading activities and the shared surface for collecting and grouping documents. But we also observed the following differences between the conditions: Participants in the SMALL condition used the shared surface in a more goal-oriented fashion and only focused on it when actively using it. In comparison, participants in the LARGE condition stayed visually focused on the shared surface, even during

phases of reflection or for visually scanning over shared documents. Interestingly, this was notable even during phases that did not require visual focus on the shared surface (e.g., when reflecting on the next steps) but during which the shared surface apparently attracted most participants' visual attention. For example, participants sometimes looked at and "played" with documents on the shared surface (i.e., playfully moving them back and forth with their fingers without the intention of actually reordering them) while thinking about how to proceed with the task and even while verbally coordinating next steps with their partner ("What could we search for that we don't already have on the table?") or reflecting on the progress so far ("What do have we already? How is all this related?").

In contrast, participants in the SMALL condition were visually focusing on and directly referring to their group partner when they were reflecting or coordinating their activities (e.g., one participant who was finished reading a document on his tablet looked up and directly addressed his group partner: "Did you find anything yet?"). These observations are also reflected in our quantitative results. Table 3.1 shows the average percentage of the session time that participants focused on the tablet, shared surface, group member, and other objects across the three conditions.

	Focus on Tablet	**Focus on Shared Surface**	**Focus on Person**	Focus on Other
SMALL	71.47 (9.57)	6.40 (2.32)	14.21 (10.06)	7.92 (2.83)
MEDIUM	77.27 (7.62)	5.69 (2.36)	8.77 (6.11)	8.27 (3.91)
LARGE	78.37 (4.97)	9.57 (2.64)	4.43 (2.25)	7.63 (3.38)

Table 3.1: Mean percentage of time that participants focused on tablets, shared surface, other person, or other objects. Standard deviation (SD) is shown in brackets, and statistically significant differences ($p<.05$) are shown in bold.

Our analysis revealed no statistically significant effect of "focus on other person" ($F(2,12) = 2.74$, $p = .10$). But as Table 3.1 shows, participants using the SMALL condition focused three times longer on the other person (14% of the time) than participants using the LARGE condition (4% of the time). With a statistically significant difference between LARGE and SMALL ($b = 9.77$, $t(12) = 2.22$, $p < .05$). There were, however, no statistically significant differences for this between the SMALL and MEDIUM conditions and no statistically significant differences between the MEDIUM and LARGE conditions.

The analysis also shows statistically significant differences for "focus on shared surface" ($F(2,12) = 6.50$, $p<.05$) between LARGE and SMALL ($b = -3.17$, $t(12) = -2.63$, $p < .05$) and LARGE and MEDIUM ($b = -3.88$, $t(12) = -3.21$, $p < .05$). Participants

using the LARGE condition focused almost twice as long on the shared surface (10% of the time) than in the MEDIUM condition (6% of the time). There were, however, no statistically significant differences for this between the SMALL and MEDIUM conditions. Focusing on 'other things' (e.g., the information sheet or just looking around) happened to the same extent in all conditions.

Finding 1.1. *Our results show that participants' visual attention in Complementary Interfaces can be drawn towards a large shared surface, even if there is no pragmatic reason to focus on the shared surface. This can result in non-"face-to-face" verbal communication without eye contact between collaborators. On the contrary, a small shared surface received less visual attention and was used in a more goal-oriented fashion. Users did not "play" with objects on the small shared surface and focused significantly longer on their collaborator.*
The choice of the size of a shared surface should therefore not be taken too lightly. Larger does not necessarily mean better. Depending on the context of use, it can be desirable to increase eye contact (e.g., to build trust in collaborators or create a sense of intimacy) or to increase the focus on the shared surface (e.g., to support collaborative activities).

Search and Transfer

Our qualitative data revealed differences in the strategies that participants employed to access documents. We observed that participants worked more *filing-oriented* in the LARGE condition and more *search-oriented* in the SMALL condition.

By *filing-oriented*, we mean that the shared surface served as a space to collect, file, and pile found documents in space and to use visual and spatial cues to quickly access and then transfer documents to the tablets. We observed that the large shared surface was used as external memory and "space to think" as described in [7].

By *search-oriented*, we mean that even after a found document was filed or piled on the shared surface, participants frequently accessed these previously found documents by repeated searches with the tablet using memorized keywords rather than looking up documents on the shared surface and sending them to the tablets.

We found evidence for these two different strategies in our video data (e.g., one user commented "I will just search for 'basement' - this will get me the document about the groundskeeper." Another user asked "Where did we put the documents with the wine export?" while trying to find a document on the shared surface). However, we could

not directly confirm these findings with a quantitative analysis. For this analysis, we attempted to use the number of document transfers across devices as indicators for the two strategies. We distinguished between (1) the transfer of documents from a tablet to the shared surface, (2) the transfer of documents from the shared surface to a tablet, and (3) the transfer of documents between tablets. We expected that a filing-oriented strategy should lead to more frequent shared surface-to-tablet transfers than a search-oriented strategy.

	Tablet to Tablet	Tablet to Shared Surface	Shared Surface to Tablet
SMALL	3.3 (4.76)	6.2 (2.32)	12.6 (8.06)
MEDIUM	5.1 (5.09)	6.0 (3.74)	14.3 (10.41)
LARGE	1.3 (1.25)	6.5 (3.21)	23.7 (12.6)

Table 3.2: Document transfers across devices per participant. The standard deviation (SD) is shown in brackets.

Table 3.2 shows the mean number of document transfers per participant across the three conditions. The results show differences, but none are statistically significant. There is only a non-significant tendency that the larger the shared surface, the more documents were generally transferred from the shared surface to tablets (e.g., almost twice as many in condition LARGE than in condition SMALL).

Finding 1.2. *There was qualitative evidence in our data that a larger shared surface affords more spatial and visual organization schemes for digital content, which is also considered one of the key qualities of large high-resolution displays [7] or tabletops [178] for sensemaking. Nonetheless, we observed how users skillfully used the search function and memorized keywords to successfully compensate for the lack of space in conditions MEDIUM and SMALL. This is noteworthy, since (as we discuss later) there was no significant difference in the objective quality of group results for the three conditions.*

Document Manipulation

During video analysis, we found that document manipulation differed across the three conditions. As mentioned above, participants in the SMALL condition used the shared surface in a very goal-oriented fashion, primarily for quick collection and organization of documents. In contrast, participants in the LARGE condition did not always interact purposefully with documents but tended to “play” with documents

by moving them on the shared surface. Participants also spent a lot of time browsing and arranging the content of the shared surface while seemingly getting sidetracked from the task at hand. This could imply that the larger shared surface affords a more spatial, but also more playful and serendipitous interaction with the documents, while the smaller shared surface affords a limited but more goal-oriented interaction.

Table 3.3 summarizes our quantitative analysis of participants' document manipulation on the shared surface. It shows the mean number of different manipulations (i.e., *add*, *delete*, *move*, *cluster*) per session.

	Add	Delete	**Move**	Cluster
SMALL	5.1 (2.43)	0.6 (0.80)	9.7 (6.12)	4.1 (2.84)
MEDIUM	5.8 (3.74)	1.5 (1.20)	9.7 (6.72)	2.5 (3.26)
LARGE	5.8 (2.75)	0.9 (0.94)	21.7 (14.51)	2.8 (1.72)

Table 3.3: Mean number of different manipulations of documents on the shared surface per session. Standard deviation (SD) is shown in brackets, and statistically significant differences ($p<.05$) are shown in bold.

The analysis shows how the condition did indeed affect the number of document movements on the shared surface. It revealed a significant effect of the number of *move*-manipulations ($F(2,12) = 3.98, p<.05$) between LARGE and SMALL ($b = -12.0$, $t(12) = -2.32, p<.05$) and LARGE and MEDIUM ($b = -12.0$, $t(12) = -2.32, p < .05$).

Participants in the LARGE condition moved documents almost twice as often as participants working with MEDIUM or SMALL while there were no significant differences in *add*, *delete*, or *cluster*. This hints at a more exploratory working style with frequent rearrangements of documents and a more playful style of interaction.

Finding 1.3. *We found that a larger shared surface can invite a more explorative or playful interaction with objects on the shared surface that is also reflected in a significantly higher number of movements. Other kinds of manipulations (e.g., clustering) showed no statistically significant differences.*
This more frequent movement of objects is not necessarily connected to the purposeful rearrangement of objects in space for sensemaking. Instead, larger shared surfaces also seem to invite more playful and serendipitous interactions that happen between and during the actual goal-oriented manipulations or conversations.

3.3.2 RQ1.2: Communication Activities

Our analysis differentiates between the *amount* of communication and *characteristics* of communication. Both are discussed in the following.

Amount of Communication

Our qualitative observations indicate that the amount of communication differs across the three conditions. The larger shared surfaces seemed to attract and hold attention and reduce verbal communication.

For example, we observed that participants in the SMALL condition often explicitly dedicated time to engage in communication with their group partner. Doing so, they would lean back in their chairs and shift their attention to the other participant and away from the devices. This "leaning back" was intended to signal readiness to communicate with the other group member. Typically, this was followed by direct verbal communication about the current results and the next steps. It occurred during group work, when participants stepped out of their current activity to reflect on the work done so far and think about the next steps. Thus, it can be regarded as an active measure to create group awareness.

In the LARGE condition, participants only rarely took the time to directly address their group partner without focusing on the shared surface. They reflected on their work not by directly addressing their group partner but through activities mediated by the shared surface, such as browsing and rearranging the content. This was not necessarily accompanied by verbal communication.

	Talk
SMALL	16.51 (9.03)
MEDIUM	13.68 (4.95)
LARGE	8.39 (4.68)

Table 3.4: Percentage of time that participants were talking to each other. The standard deviation (SD) is shown in brackets.

The differences in the amount of verbal communication were also part of our quantitative analysis. Table 3.4 shows the percentage of time that participants were talking as opposed to being silent during the session. Participants talked most in the SMALL condition (16% of the time) and least in the LARGE condition (8% of the time). However, the analysis revealed that the differences were not statistically significant ($F(2,12) = 2.55$, $p = .12$) with $p = 0.057$ for LARGE vs. SMALL being only marginally greater than .05 ($b = 8.12$, $t(12) = 2.11$).

Finding 2.1. *Based on our qualitative analysis, we found that participants talked more and more directly to each other when using the SMALL condition than when using a large shared surface. We observed that this face-to-face talk seemed to be more important when using a smaller shared surface and that participants introduced phases exclusively dedicated to verbal communication for coordinating actions and creating awareness. This suggests that verbal communication was used to compensate for a smaller shared surface that provided less overview and only allowed for less effective deictic references. However, a quantitative analysis did not reveal a statistically significant difference in the total amount of communication.*

Characteristics of Communication

To investigate what mediates communication across the three conditions, we analyzed where participants were looking while they were talking to each other.

Based on our video observations, we also found differences in communication with respect to how participants used deictic references by pointing to documents on the shared surface for elaborating their arguments. For example, participants working with the LARGE condition were often already engaged in sensemaking discussions while they were interacting with the documents on the shared surface. This was also reflected in the prevalence of deictic references that accompanied their communication. For instance, "What do we have here? We have the groundskeeper cleaning the basement (participant *pointing* at the shared surface), we have the event at the town hall (*pointing*), and the sick grandparents (*pointing*). How could these three be related?" In the MEDIUM and LARGE conditions the two group work activities *structuring documents on the shared surface* and *talking about and making sense of the gathered information* blended into each other, and participants frequently switched between them. When group partners talked during such activities, this happened often without eye-contact and was mediated by the shared surface.

In contrast, the two activities *structuring documents on the shared surface* and *talking about and making sense of the gathered information* rarely overlapped in the SMALL condition. After participants clustered the documents on the shared surface according to document topics, they would disengage from the shared surface and address their group partner to reflect on the information. For example, one participant was silently clustering documents on the shared surface while the other participant was reading a document on his tablet. Once the participant was finished with his sorting

activity, he withdrew from the shared surface, rubbing his face, and addressing his partner by saying: "Let's think about what we have so far," upon which his partner looked up and they started a discussion about the information that they have read and found so far without using deictic references.

Our quantitative results confirm these differences for the visual focus. Table 3.5 shows the average percentage of time that participants focused on the tablet, shared surface, group member, and other objects while talking. Results indicate that participants mostly focused on their tablets while talking. This is not surprising and lies in the nature of the task, in which the tablets were the source of the information that participants wanted to convey to their partner. For instance, they looked at it to read important information aloud or to verify that their contribution was correct.

	Talk and Focus Tablet	**Talk and Focus Shared Surface**	Talk and Focus Person	Talk and Focus Other
SMALL	49.27 (10.13)	11.53 (6.65)	29.76 (15.41)	9.44 (4.66)
MEDIUM	53.73 (11.53)	12.33 (8.08)	24.96 (11.92)	8.98 (5.07)
LARGE	49.36 (14.84)	24.61 (8.88)	18.11 (7.60)	7.92 (5.00)

Table 3.5: Individuals' focus while talking. Standard deviation (SD) is shown in brackets, statistically significant differences ($p<.05$) are shown in bold.

The analysis showed a significant effect on the time they focused on the shared surface while talking ($F(2,12) = 8.03$, $p < .05$) for LARGE and SMALL ($b = -13.08$, $t(12) = -3.39$, $p < .05$) and LARGE and MEDIUM ($b = -12.27$, $t(12) = 3.18$, $p < .05$). In the LARGE condition, participants focused almost twice as long on the shared surface (24.6%) than in the MEDIUM (12.33%) or SMALL condition (11.53%). In addition, Table 3.5 also shows that participants in the SMALL condition focused more on the other person while talking, whereas participants in the LARGE condition focused more on the shared surface than on the other person.

Finding 2.2. *The large shared surface attracted the visual attention during verbal communication. It enabled a close integration of structuring and manipulating items on the shared surface with talking about and making sense of items including the use of deictic references. This close integration did not happen for the SMALL condition, where active structuring and talking rarely overlapped. Instead, these activities happened sequentially with frequent phases that were exclusively dedicated to verbal face-to-face communication (including eye contact) without interacting with the shared surface.*

3.3.3 RQ1.3: User Perceptions and Group Results

We analyzed data from our interviews and questionnaires to understand (1) participants' collaboration strategies during group work, (2) participants' satisfaction with the system support, and (3) the objective quality of the results of group work.

Group Strategy and Awareness

Concerning collaboration strategies, participants in the SMALL condition reported that they had not explicitly distributed tasks among themselves. Rather they worked mostly individually while keeping their partner informed about their current progress. They also stated that most of the time they were aware of their group partners' activities and that "there was no need to communicate the strategy – only findings needed to be communicated." They stated that their main focus was on reading and that clustering and sorting on the small shared surface was considered as taking "too much time."

Participants in the MEDIUM condition reported that they occasionally distributed tasks among each other, but that a more explicit assignment of subtasks would have improved group work. Despite the division of labor, they were mostly aware of their partner's activities, but they sometimes needed to ask their partner for a progress update ("What are you currently working on?") to actively create group awareness.

In the LARGE condition, participants stated that they frequently distributed tasks among themselves. They also reported that they were only partially aware of their group partner's activities, but some awareness was mediated by the shared surface. Clustering and sorting activities were perceived to help them in sensemaking and reflection. One participant said: "If I didn't know how to proceed, I just looked at the table."

User Satisfaction with System Support

Participants in the SMALL condition valued the support for direct document transfer between tablets and the highlighting function to emphasize important parts of a document. They were also satisfied with the ability to cluster documents. Other advantages of the shared surface were not mentioned. They reported that they did not divide the shared surface into personal and group territories [360], but, in contrast, explicitly valued the clear distinction between personal activities on the tablet and shared activity on the shared surface (as suggested by [387]).

Participants in the MEDIUM condition mainly valued the shared surface as an information and memory pool that allowed them to quickly access relevant documents. Furthermore, they reported using the space to sort documents thematically. Most

of them stated that they were satisfied with the size of the shared surface, however, some mentioned that a larger shared surface might have allowed for additional personal spaces on the interactive surface.

Participants in the LARGE condition explicitly stated their appreciation of the shared surface as an efficient but also playful way of accessing documents, while providing a good overview. They additionally stated that they divided the shared surface into different areas that they used to distinguish between kinds of documents. For example, one group reported that they "stuck all the unimportant stuff in the corner and focused on the main documents of the plot in the center of the table" which resonates with the observations of Andrews et al. [7] about sensemaking on large screens. Participants used the shared surface to structure documents and to compare them with documents displayed on the tablets. Furthermore, they valued being able to work on the shared surface simultaneously and reported the interaction with the system to be fun. Regarding the size of the display, participants were satisfied and positively reported that it even allowed them to have a personal territory on the shared surface as well. Some of the participants expressed a wish to also have a reading function on the shared surface in addition to that on the tablet.

Results of Group Work

To assess the results of the participants' group work, we asked them to fill out a questionnaire about facts on the plot they were investigating. This questionnaire consisted of nine multiple-choice questions. In total, participants were able to score 44 points, depending on their answers.

Table 3.6 summarizes participants' general performance across the three conditions as well as the level to which they agreed on an answer, irrespective of whether it was right or wrong. For the Performance Score, significance was tested based on individual scores using the mixed-effects model approach. For the Conformity Score, we applied a one-way ANOVA based on group scores that show how consistently the group partners performed.

	Performance Score	Conformity Score
SMALL	36.5/44 (4.86)	36.0/44 (3.08)
MEDIUM	33.3/44 (3.67)	32.4/44 (2.30)
LARGE	33.3/44 (5.16)	33.8/44 (3.27)

Table 3.6: Mean Performance Scores (mixed-effects model, based on individual scores) and Levels of Conformity per group (one-way ANOVA, based on group scores). The standard deviation (SD) is shown in brackets.

The results show no statistically significant differences for the Performance Score nor for the Conformity Score. This lets us conclude that in all three conditions, the participants were provided with sufficient support and flexibility to establish successful collaboration styles and strategies. Neither the system nor the task unfairly privileged one condition compared to another.

The qualitative analysis of how participants made use of the different shared surface sizes confirms the benefit that is usually attributed to large displays for increased awareness [270, 335] or for spatial sensemaking [7, 168]. The appreciation of the large shared surface by participants and the desire for a larger shared surface by the participants in the MEDIUM condition fit into this picture and resonate with earlier findings about the desire for more space during around-the-table collaboration [344]. The positive attitude towards a clear separation between private territories on mobile devices vs. group territories on shared surfaces is in accordance with earlier work on territories and coupling styles around tabletops [360, 387].

However, we found it surprising how the subjective preference for larger shared surfaces was not reflected in the objective quality of the results of the group work. While there were clear differences in group awareness and task distribution, there were no statistically significant differences in Performance and Conformity scores. This indicates that the participants were perfectly able to perform sensemaking and maintaining awareness using face-to-face communication, rather than it being primarily mediated by the shared surface.

Strategies included the introduction of different styles of communication with eye-to-eye contact, using search-oriented strategies instead of filing-oriented strategies, and relying more on direct exchange of information between tablets than sharing documents via the shared surface.

Finding 3. *From our observations, we can conclude that participants working with small shared surfaces were perfectly capable of maintaining awareness and performing sensemaking primarily mediated by face-to-face communication. They were able to adapt and improvise and succeeded in achieving an objective quality of work results that was not significantly different from those working with medium-sized or large shared surfaces.*

3.4 Limitations

As mentioned in the study description, the system design and the given task naturally influenced the collaboration behavior that we observed. There are specific aspects of our user study that limit the generalizability of our findings: First, the VAST 2006 Challenge made it necessary to frequently search for documents and to closely read them. This shifted the overall nature of the collaboration towards many phases of individual parallel work and fewer phases of tightly-coupled collaboration. Users focused on their tablets at least for 71% of the time (Table 3.1). A different task (e.g., collaborative design or gaming) might afford more tightly-coupled collaboration and thus emphasize the role and the importance of the shared surface. This could also lead to a generally greater importance and benefit of larger shared surfaces than we observed. Second, our groups consisted of two users. We chose dyads like Isenberg et al. [168], who demonstrated that dyads are sufficient to identify differences in coupling styles and collaboration strategies. Choosing larger groups might reveal a greater need for display space than in our user study. In larger groups, there will also be an increased importance of group awareness – whether it is actively created through verbal conversation or is mediated by the size and physical affordances of a shared surface. Third, our participants were students whose analysis skills differ from that of professional analysts. But as our user study focused on observed differences between the three conditions and their user interface and demands for technical proficiency were almost identical, our findings should be independent from education and technical skills of participants. Observed differences should not be substantially different for non-students. However, age-related differences in memory, eye-sight, or motor skills could have an impact (better eye-sight and memory could benefit SMALL, motor impairments could hamper LARGE).

Nonetheless, we believe that our user study employed a balanced task set and also realistic setting for collaboration. That none of the conditions can be considered unfair or unsolvable becomes evident through the absence of statistically significant differences in the groups' performance scores.

3.5 Implications for Future Research

Our findings show the need for a more nuanced understanding of the role of a shared surface and its size as part of a Complementary Interface for co-located collaboration. We have observed that for the particular kind of collaborative sensemaking task that we studied, a larger shared surface did not lead to a better task performance or user experience. In particular, we found evidence that the large shared surface attracted significantly more attention than the smaller ones – for good and for ill.

However, a major part of our sensemaking task consisted of individual work with personal tablets, e.g., searching for documents and reading text. It would be interesting to contrast our results with a study of a task where the major part of the work requires tightly-coupled collaboration. It would also be interesting to see to what extent the found strengths of SMALL shared surface sizes (i.e., increased group awareness, focused work strategy) hold true for group sizes larger than two. Some effects that we considered to be positive in our study setting (e.g., focus on other person) might be detrimental to the quality of collaboration in larger groups.

Although we did not analyze territoriality in our study, findings suggest that the size of the shared surface influences how people manage personal and shared work spaces in cross-device settings. Research investigating the effect of shared display size on territoriality in group collaborations would further help to understand the role of the shared surface in relation to personal devices.

In addition, we are interested in how our findings relate to a study investigating the effect of a shared display size in a vertical setting. Changing the orientation of the shared display also implies changes to the physical user position and interaction with the system, which might strongly influence the effects that we found. We would especially expect different results in how participants would refer to the shared display if it were vertically oriented.

Research on shared interactive surfaces has primarily focused on large surfaces or large interactive wall displays with up to 200″ [29]. Our findings imply that under certain conditions small shared interactive surfaces could outperform large ones. While tablets are typically designed for personal tasks, our results encourage research on re-purposing tablet-sized devices as shared interactive surfaces for mixed-focus collaboration, or even to ad hoc expand the shared surface using additional devices.

3.6 Chapter Conclusion

We built a Complementary Interface that allowed dyads to engage in sensemaking activities. The meaningful combination of tablets for personal activities and a complementary shared interactive surface for collaborative activities allowed for mixed-focus collaboration. We studied the effect of the size of the shared surface on participants' attention, awareness, and efficiency during an authentic sensemaking activity. We studied 15 dyads working with our Complementary Interface consisting of two personal tablets and a shared interactive surface with three sizes. We found that larger shared surfaces do not necessarily improve collaboration, as they can divert the attention away from their collaborator towards the shared surface. This has implications for participants' interaction and communication style. The results show

that each condition can be seen as a valid design alternative in this specific Complementary Interface. Yet, we conclude that the choice of the size for this shared surface is not obvious and should depend on the task at hand – emphasizing the embedding of a Complementary Interface into a context.

Parts of the following Chapter 4 have been published as:

***Johannes Zagermann**, Ulrike Pfeil, Daniel Fink, Philipp von Bauer, and Harald Reiterer. 2017. Memory in Motion: The Influence of Gesture- and Touch-Based Input Modalities on Spatial Memory. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), Association for Computing Machinery, Denver, Colorado, USA, 1899-1910. DOI: https://doi.org/10.1145/3025453.3026001*

The responsibilities of this joint publication were divided as follows: I formulated the research questions, designed the user study, and spearheaded the writing. Ulrike Pfeil helped in analyzing the results and the writing process. Daniel Fink implemented the research prototype. Philipp von Bauer conducted several user study sessions. Harald Reiterer supervised the work.

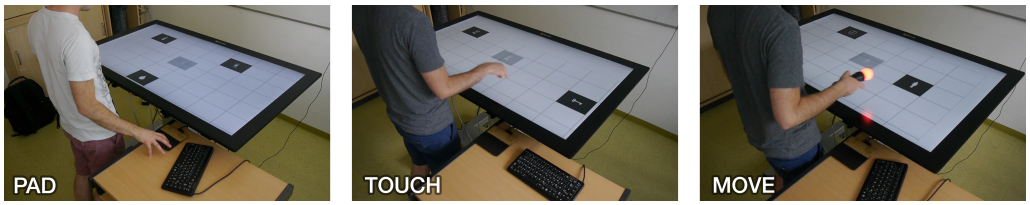
Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

Video Figure

Conference Talk



This Complementary Interface allows to study multimodal interaction. In contrast to the first Complementary Interface that allowed to study the role of the size of a shared interactive surface on collaborative activities, this Complementary Interface enables the investigation of the influence of input and output modalities on individual spatial memory: Specifically, it allows for an experimental comparison of a variety of meaningful combinations of input and output modalities regarding their influence on efficiency, user satisfaction, and spatial memory in an abstract knowledge work task.

Study Design:	Between-Subjects and Within-Subjects
Task:	Navigation Task, Object Location Task
Independent Variable:	Type of Input Modality, Size of Output Modality
Dependent Variables:	Spatial Memory, Efficiency, User Satisfaction

---

The described Complementary Interface and experimental setting are used to address the following general research question. The chapter will integrate this general research question as its overall research goal and further address it in specific use-case related research questions.

RQ2: *What are the influences of embodiment (represented as meaningful combinations of input and output modalities) on spatial memory in a Complementary Interface?*

4

Studying Multimodal Interaction

Contents

Spatial memory is a crucial aspect of successful interaction design as it takes advantage of users' ability to memorize object locations to ease navigation and search processes. Interfaces that harness users' spatial memory can profoundly reduce cognitive and physical load, support fluid interaction, and free resources to allow for better task performance [348]. Spatial memory is a fundamental component across a wide span of everyday interactive systems. For example, it can be utilized in desktop applications to support menu navigation and icon search [319, 348]. In particular, spatial memory is an important aspect for interfaces consisting of a large digital landscape that extends the visible display space (e.g., pan and zoom interfaces in public displays, overview + detail interfaces in navigation systems). In regular use, spatial memory eventually reduces visual search activities, leading to automatic navigation and freeing cognitive resources for the task [348].

Previous research has shown that the size of the output modality (i.e., a display) and the input modality impact on users' spatial memory and navigation performance [179, 304, 324, 348, 384, 386]. Yet, these studies investigated input and output modalities separately, making it difficult to apply the results to real-world scenarios.

We believe that we need to consider the interplay of different input and output modalities to provide valuable insights on spatial memory for everyday interactive systems (e.g., desktop applications, navigation systems, or public displays). In this chapter,

we built a Complementary Interface that allows us to study the meaningful combinations of input and output modalities that map current off-the-shelf interactive systems (e.g., tablets, laptops, and larger surfaces supporting input via trackpad, direct touch as well as gesture). Also, previous work suggests that different aspects of embodiment (e.g., touch vs. mouse [179, 384], body vs. touch panning [201], body movement [321], direct vs. indirect touch [304]) positively affect spatial memory performance. However, existing studies mostly compare input modalities that differ profoundly in their interaction techniques, not explaining how the underlying aspects of embodiment influence spatial memory. Applying the findings to settings beyond the specific conditions used in these studies is difficult. Understanding how aspects of embodiment (e.g., haptic feedback, kinesthetic cues, or direct interaction) influence spatial memory can help system designers to apply the findings to the input and output modalities of their specific setting. Furthermore, previous research has focused on navigation path length and object location when investigating spatial memory. However, for designers to be useful, it is essential to understand how spatial memory relates to usability aspects. Looking at the definition of usability, which is divided into effectiveness, efficiency, and user satisfaction [148], we argue that current methods can be mapped to effectiveness (e.g., how well users perform a task). However, the relationship between spatial memory, efficiency, and user satisfaction is unclear. In real-life scenarios, spatial memory is not detached from usability, but designers have to gauge these aspects depending on the context and the task at hand (e.g., prefer efficiency in a desktop application or user satisfaction in leisure activities). In these cases, relating spatial memory to usability aspects facilitates the incorporation of spatial memory into system design.

This chapter contributes a detailed understanding of how meaningful combinations of different input and output modalities impact on spatial memory. We also investigate the role of embodiment (e.g., kinesthetic cues, haptic feedback, direct interaction) in these effects. Analyzing the influence of these aspects of embodiment not only on spatial memory but also on efficiency and user satisfaction, we tackle the question of to what extent these measurements lead to coherent results. Our multi-faceted analysis accounts for complex yet meaningful combinations of input and output modalities. Thus, results can be mapped onto real-world scenarios, and implications can explain user behavior and support navigation and search processes in interactive systems.

The goal of our study was to better understand how aspects of embodiment (represented as meaningful combinations of input and output modalities) influence spatial memory. In addition, we were interested in the relation of these measures to the efficiency of task completion and user satisfaction. To this end, we defined input

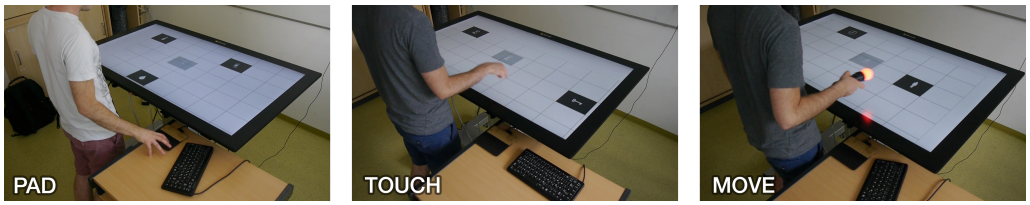


Figure 4.1: This Complementary Interface allows to study multimodal interaction. This figure shows three input modalities: PAD, TOUCH, and MOVE that represent the conditions of the within-subjects independent variable.

modality and display size as independent variables. To address our dependent variables, we focused on spatial memory, efficiency, and user satisfaction, resulting in the following three research questions:

RQ2.1: *How do meaningful combinations of input and output modalities influence spatial memory? (measured in the navigation performance and object location memory performance)*
RQ2.2: *How do meaningful combinations of input and output modalities influence the efficiency of task completion? (measured in navigation speed and task completion time)*
RQ2.3: *How do meaningful combinations of input and output modalities influence the user satisfaction? (measured in perceived task load and user experience)*

4.1 Experiment

Concerning the input modality, we defined the three conditions: (1) **PAD** (using a touchpad), (2) **TOUCH** (using direct single touch), and (3) **MOVE** (using a gesture-based motion controller). We chose these three conditions as they constitute three types of state-of-the-art input modalities: Touchpads are common features of laptops and can replace the use of a mouse, especially to provide mobility. Touch technology can be seen as the primary input modality for mobile devices, tabletops, and public displays. As the number of consumer devices that allow for gesture-based interaction either using additional sensing hardware like controllers (e.g., Sony PlayStation 3 Move motion controller or Nintendo's Wii Remote) or based on computational image processing (e.g., Microsoft Kinect or laptops with built-in Intel RealSense cameras) increases, we chose gesture-based interaction as a third input modality. In addition to off-the-shelf devices that allow for the individual use of the named input modalities, there is also a growing market of devices like e.g., laptops, that enable all three input

modalities to be used in sequence or even in combination, bringing up the question of when to use which of them. We tested several arrangements and carefully considered comparable conditions to ensure participants can handle the conditions equally well to ensure comparability and prevent problems with target acquisition. Regarding **MOVE**, we decided for an additional sensing hardware as the controller showed a higher accuracy and robustness for panning operations compared to the computational detection of a hand's position. This was also the case when dealing with arbitrary and task-dependent movements like panning or clutching. The activation of the panning mode was more robust using the trigger of the controller compared to the detection of an additional finger gesture.

	PAD	TOUCH	MOVE
Kinesthetic Cues	—	×	×
Direct Interaction	—	×	—
Haptic Feedback	×	×	—
Activation	Touch	Touch	Trigger

Table 4.1: Comparison of the three different input modalities

Table 4.1 shows an overview of the characteristics of the three different input modalities with respect to embodiment. Our choice of input modalities allows us to compare them on the basis of four aspects: (1) *kinesthetic cues* – an awareness of body's (or parts of it) position with respect to itself or to the environment [384]. For our study, we describe *kinesthetic cues* as the extent to which the user's arm movement corresponds to the actual path length, (2) *direct interaction* – here we distinguish between direct panning manipulations (TOUCH) and indirect panning manipulations using a mediator (PAD and MOVE), (3) *haptic feedback* – we consider this as direct feedback to panning a specific surface, and (4) *activation* – switching the mode between arbitrary and task-oriented movements.

The independent variable *input modality* is a within-subjects variable allowing for inner-subjects ratings of comparisons and preferences.

Concerning the output modality, we define the conditions (1) **SMALL** (display size: 10.6″, a typical size for a personal device) and (2) **LARGE** (display size: 55″, a typical size for a public display) to account for two different scales to exhaust the respective level of embodiment. This independent variable is a between-subject variable. We expect that especially the positive influence of kinesthetic cues will result in greater spatial memory performance on the LARGE condition compared to the SMALL condition.

4.1.1 Apparatus

Figure 4.1 shows the setting of our study with LARGE and the three input modalities: PAD, TOUCH, and MOVE. We chose two differently sized output devices: a 10.6″ Microsoft Surface 2 Pro and a 55″ Microsoft Perceptive Pixel. Both devices had the same display resolution of 1920×1080 pixels to guarantee for a higher internal validity of the study. In the condition with the 10.6″ display, the device was placed on top and in the center of the 55″ display as our pre-test showed that it eased interaction with MOVE and TOUCH (people stepped back when placing the tablet closer). Also, this placement ensured that participants' postures were comparable for both display sizes. Next to the display, we placed a small mobile desk, which could be easily reconfigured by participants depending on their handedness. This desk was used to fill out questionnaires and to place the input devices on.

In the **TOUCH** condition, participants used the output device-specific touch capabilities. For our study, we asked participants to use single touch only to ensure comparability between the three conditions. We used a Logitech T650 trackpad for the **PAD** condition, adjusting it in three ways: from a hardware perspective, we limited the area where participants could interact with it to a 16:9 ratio. This allowed us to have the same ratio of participants' movements in all conditions as the displays have the same ratio. We did this by placing an overlay on top of the trackpad itself. From a software perspective, we changed the control-display-ratio. Panning from the outermost left position of the trackpad to its outer-most right position was equivalent to the same panning operation using TOUCH, resulting in a 1:1 mapping with the output device. We adjusted the trackpad to work without additional pressing to have a comparable haptic feedback with TOUCH. These three design decisions regarding the trackpad led to a better comparability between the three conditions.

For the **MOVE** condition, we used a Sony PlayStation 3 Move motion controller (weight: 277g). We attached a PlayStation Eye Camera in bird's-eye-view directly above the display and adjusted the distance to the display to allow for an interaction space corresponding to the size of the display, which allowed us to have a comparable movement range to the TOUCH condition. The camera detected the position of the visual marker of the controller, and participants were able to pan the canvas by pulling and holding a trigger at the backside of it.

4.1.2 Task

We used a well-established and validated task [159, 179, 201, 233, 274, 275, 277, 321] consisting of a navigation phase and a subsequent object location memory phase. Using one of the input modalities, a canvas could be navigated using panning operations. The canvas itself contained a 24×18 grid with a spatial configuration of

18 items in total. In the default position, an 8×6 section of the grid was visible – containing no items.

In the navigation task, the system started at the default position in the center of the canvas, where no items were located. In the middle of this default position, a semi-transparent overlay of the item to be searched was shown. Participants had to find this item on the canvas and pan to its position using the condition-specific panning possibilities. Items to be searched were auto-selected as participants panned to their position with a tolerance of 100px to their center. After the item was found, the task continued at the default position with the next item. For each input modality, participants had to find 8 out of 18 items. The same sequence was repeated 6 times (6 blocks), resulting in 48 search trials. In total, each participant had three of these 48 search trials (one per input modality), resulting in 144 trials per participant and 4032 trials over all sessions.

To avoid possible learning effects, we used several item sets: (1) a demo item set containing 12 letters of the alphabet, (2) an exercise item set containing 18 symbols, and (3) three task item sets containing 18 symbols each. All item sets were exclusive without overlaps. Additionally, the items were divided equally between the different item sets with respect to their complexity and theme (e.g., devices, animals, vehicles, etc.). Also, the position of the different items and, thus, the total optimal navigation path length was taken into consideration with a similar length in all item sets (16729px, 16125px, and 15956px) with a randomization factor of +-3% to prevent possible influencing factors like symmetry or memorization of specific areas.

In the second phase – the object location phase – participants had to locate the presented items on their specific position on the canvas. In this phase, the entire canvas was shown, with no need for panning interaction. To locate the items, participants used the arrow keys of a keyboard to avoid influences of the specific input modality (cf. [179]). After placing an item to its position, the specific item was invisible to avoid possible mental constructs. We provide the application with all item sets and a supplemental video online.[1]

4.1.3 Participants

Twenty-eight participants (12 female, 16 male, 0 non-binary) were recruited for the experiment. As a requirement, we only invited participants who had German as their mother tongue. By this, we avoided possible influences on the memorization of the items due to different linguistic representations of foreign languages similar to prior work [179, 233]. The mean age was 24.57 years (SD = 3.62, aged 18-32). All participants were right-handed and had normal or corrected to normal eye-sight,

[1] https://github.com/hcigroupkonstanz/Panning-Evaluation-Touch-Pad – last accessed 01.11.2023

consequently, they had no problems with the employed size of items. Twenty-one participants were undergraduate students, 4 PhD students, and 3 university employees. Participants had a mixed background ranging from computer science to psychology, biology, educational science, and law. All participants were familiar with touch interfaces, and 18 participants reported to have already used gesture-based input modalities, for example Nintendo's Wii Remote. To recruit participants, we used postings and flyers looking for people who are interested in playing games like Memory/Pairs, which helped us to find participants who were motivated by our chosen task. We assigned 14 participants to each display size.

4.1.4 Procedure

At the beginning of the study, each participant was asked to fill out a demographic questionnaire, including questions about tech-savviness and eye-sight. Then, each participant received an introduction to the navigation and object relocation phases with their first input modality and a demo item set. Participants were given time to explore the functionalities of the system and input modality with an exercise data set until they felt comfortable using it. After that, participants were asked to start the navigation phase with the first input modality with the remark to navigate to the required items as fast as possible. After that, participants were asked to fill in a NASA-TLX [136] and User Experience Questionnaire (UEQ) [228]. Then, participants had to relocate the position of all items on the canvas as precisely as possible. This sequence was the same for all three input modalities. To avoid learning effects, we counterbalanced the order of input devices and the assignment of item sets by using a random selection of 14 out of 36 possible combinations for each output modality. Finally, each participant filled out a questionnaire concerning subjective preferences. Each session lasted about one hour in total, and participants were compensated for their time.

4.2 Analysis

We chose a 2×3 factor split-plot counterbalanced study design consisting of a between-subjects part with the display size as the independent variable with two conditions (SMALL and LARGE) and a within-subjects part with the input modality as the independent variable with three conditions (PAD, TOUCH, and MOVE)

4.2.1 Data collection

We employed several data collection methods to investigate our research questions:

RQ2.1 - Spatial Memory: We focused on the task-dependent aspect of spatial memory: Navigation performance and object location memory performance. For the naviga-

tion performance, we calculated the ratio between users' actual navigation path to the shortest possible path (excluding the first search block as results for this are influenced by the randomness of initial navigation trials, cf. [179]). For the object location memory, participants were asked to position items on an empty canvas at the correct position. Here, we logged the position and calculated the Euclidean distance to its actual position in pixels.

RQ2.2 - Efficiency: To analyze the efficiency of task completion, we logged the duration participants needed to find the specific items (i.e., the task completion time in seconds for each item) and the navigation speed (pixels per second) that participants operated at.

RQ2.3 - Satisfaction: To analyze the user satisfaction for the investigated conditions, we used three questionnaires in total: a NASA-TLX [136] was used to measure the perceived task load. The UEQ [228] provided information on the experience using the system in the different conditions. Additionally, we asked participants at the end of the experiments to rank the three input modalities and justify their preferences.

4.3 Results

In this section, we report our findings in relation to our three research questions. In particular, we focus on the differences across the three different input modalities (PAD, TOUCH, and MOVE) and output devices (SMALL and LARGE) with respect to participants' spatial memory (RQ2.1), their efficiency (RQ2.2), and their subjective satisfaction (RQ2.3). As the data violated the assumption of normal distribution, we decided to use a non-parametric approach to analyze the data statistically. For the overall tests on statistical significance concerning the three input modalities, we used a Friedman test. If this test showed statistically significant differences, we employed a Wilcoxon signed-rank test as a post-hoc analysis, including Bonferroni correction. Additionally, a Mann-Whitney U-test was used to test for statistically significant differences between the two display sizes.

4.3.1 RQ2.1: Spatial Memory

For spatial memory, we differentiate between navigation path and object location memory.

Navigation Path Length

Mean values for *Navigation Path Length* are shown in Table 4.2. Results show a statistically significant difference in the navigation path length depending on the input modality, $\chi^2(2) = 6.929$, $p < .05$, irrespectively of the display size. However, a

Wilcoxon signed-rank test showed no statistically significant effect between pairs after applying Bonferroni correction. A Friedman test showed no statistically significant differences on the levels of SMALL and LARGE for the three input modalities.

A Mann-Whitney-U-Test showed a statistically significant difference for the display size when comparing navigation path length for TOUCH ($U = 55.0$, $p < .05$) and MOVE ($U = 52.0$, $p < .05$).

Focusing on the navigation path across blocks, Figure 4.2 shows that participants' performance increased over time with all input modalities. Whereas with PAD and MOVE, the navigation path seems to level off in block 5, this was found to occur earlier in TOUCH as here, the curve already bends in block 4. In addition, the lines indicating the decrease in mean navigation path length per block differ for TOUCH and MOVE when comparing SMALL and LARGE.

	PAD	TOUCH	MOVE
M_{TOTAL}	2.65 (1.10)	2.11 (0.85)	2.43 (0.97)
M_{SMALL}	2.74 (1.07)	2.43* (0.93)	2.69* (0.75)
M_{LARGE}	2.56 (1.17)	1.78* (0.64)	2.18* (1.13)

Table 4.2: Mean navigation path length (1.0 = optimal path length). Standard deviation (SD) is shown in brackets; statistically significant differences are indicated via asterisks (for between).

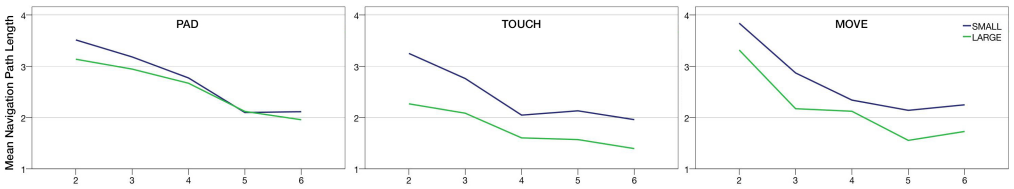


Figure 4.2: Mean Navigation Path Length for PAD, TOUCH, and MOVE per block. Optimal navigation path length: 1.0

Object Location Memory

Mean values for *Object Location Memory* are shown in Table 4.3. Results show no statistically significant difference in the object location memory depending on the input modality ($\chi^2(2) = 3.429$, $p > .05$), irrespectively of the display size. However, there was a statistically significant difference in the object location memory for LARGE, ($\chi^2(2) = 7.429$, $p < .05$) – but no difference for SMALL. A Wilcoxon signed-rank test showed a statistically significant difference for a pairwise comparison of PAD and MOVE for LARGE ($Z = -2.919$, $p < .016$).

A Mann-Whitney-U-Test showed no statistically significant difference for the display size when comparing object location memory.

	PAD	TOUCH	MOVE
M_{TOTAL}	1175.12 (457.97)	1075.80 (352.63)	1059.50 (534.19)
M_{SMALL}	1107.45 (521.41)	1156.96 (399.52)	1250.64 (638.24)
M_{LARGE}	1242.79^{g} (392.26)	994.64 (290.61)	868.36^{g} (326.53)

Table 4.3: Mean Euclidian distance in pixels. SD is shown in brackets, and statistically significant differences are indicated via raised letters (for within).

4.3.2 RQ2.2: Efficiency

To provide a nuanced description of efficiency, we differentiate between navigation speed (in pixels per second) and average task completion time for each item search (in seconds).

Navigation Speed

Mean values for *Navigation Speed* are shown in Table 4.4. There was a statistically significant difference in the navigation speed depending on which type of input device was used, $\chi^2(2) = 30.071$, $p < .05$, irrespectively of the display size. A Wilcoxon signed-rank test showed that the navigation speed among all three input modalities was statistically significantly different: PAD-TOUCH: ($Z = -3.575$, $p < .016$), PAD-MOVE: ($Z = -4.486$, $p < .016$), TOUCH-MOVE: ($Z = -4.144$, $p < .016$).

Results show statistically significant differences in the navigation speed across the input modalities for SMALL ($\chi^2(2) = 9.000$, $p < .05$) and for LARGE ($\chi^2(2) = 24.571$, $p < .05$). A Wilcoxon signed-rank test showed that navigation speed using SMALL was statistically significantly different for pairwise comparisons of PAD and MOVE ($Z = -2.919$, $p < .016$) and TOUCH and MOVE ($Z = -2.982$, $p < .016$). A Wilcoxon signed-rank test showed that the navigation speed on LARGE was statistically significantly different for all pairwise comparisons: PAD-TOUCH: ($Z = -3.296$, $p < .016$), PAD-MOVE: ($Z = -3.296$, $p < .016$), TOUCH-MOVE: ($Z = -2.982$, $p < .016$).

A Mann-Whitney-U-test showed a statistically significant difference for the display size when comparing navigation speed for TOUCH ($U = 2.0$, $p < .05$) and MOVE ($U = 1.0$, $p < .05$).

Task Completion Time

Mean values for *Task Completion Time* for single items are shown in Table 4.5. Results show a statistically significant difference in the task completion time depend-

	PAD	TOUCH	MOVE
M_{TOTAL}	1421.96[a, c] (456.62)	1159.53[b, c] (394.74)	967.79[a, b] (327.92)
M_{SMALL}	1499.03[d] (581.62)	1474.18[e, *] (299.29)	1224.72[d, e, *] (266.30)
M_{LARGE}	1344.89[g, h] (286.30)	844.89[g, i, *] (144.30)	710.87[h, i, *] (101.20)

Table 4.4: Mean navigation speed (in pixels per second). SD is shown in brackets; statistically significant differences are indicated via asterisks (for between) or raised letters (for within).

ing on the input modality, $\chi^2(2) = 22.071$, $p < .05$, irrespectively of the display size. A Wilcoxon signed-rank test showed that task completion time using MOVE was statistically significantly longer than using PAD ($Z = -4.099$, $p < .016$) and TOUCH ($Z = -2.755$, $p < .016$). There was no statistically significant difference between PAD and TOUCH.

Results also show a statistically significant difference in the task completion time depending on the input modality – for SMALL ($\chi^2(2) = 15.857$, $p < .05$) and for LARGE ($\chi^2(2) = 9.000$, $p < .05$). A Wilcoxon signed-rank test showed that the navigation time on SMALL was statistically significantly different for pairwise comparisons of PAD and MOVE ($Z = -2.480$, $p < .016$) and for TOUCH and MOVE ($Z = -2.480$, $p < .016$). A Wilcoxon signed-rank test showed that only the pairwise comparison for the task completion time using LARGE for PAD and MOVE are statistically significantly different ($Z = -3.233$, $p < .016$).

A Mann-Whitney-U-Test showed a statistically significant difference for the display size when comparing navigation time for TOUCH ($U = 55.0$, $p < .05$).

	PAD	TOUCH	MOVE
M_{TOTAL}	3.71[a] (1.21)	3.90[b] (1.54)	5.31[a, b] (2.17)
M_{SMALL}	3.56[d] (0.89)	3.37[e, *] (1.14)	4.48[d, e] (1.02)
M_{LARGE}	3.87[g] (1.48)	4.43[*] (1.74)	6.13[g] (2.70)

Table 4.5: Mean task completion time (in seconds). SD is shown in brackets; statistically significant differences are indicated via asterisks (for between) or raised letters (for within).

4.3.3 RQ2.3: User Satisfaction

For user satisfaction, we differentiate between participants' task load, their user experience, and the subjective preferences of input modalities.

Task Load

Mean values for *Task Load* are shown in Table 4.6. Results show a statistically significant difference in the task load measured by the NASA-TLX questionnaire depending on the input modality, $\chi^2(2) = 14.721$, $p < .05$, irrespectively of the display size. A Wilcoxon signed-rank test showed that the task load using MOVE was statistically significantly higher than for PAD ($Z = -3.952$, $p < .016$) and for TOUCH ($Z = -3.063$, $p < .016$). There was no statistically significant difference between PAD and TOUCH ($Z = -0.445$, $p > .016$).

Results also show a statistically significant difference when comparing the overall task load for the three different input modalities on the level of SMALL, $\chi^2(2) = 10.145$, $p < .05$. Here, a Wilcoxon signed-rank test revealed statistically significant differences for a pairwise comparison of PAD and MOVE ($Z = -3.077$, $p < .016$). However, no statistically significant difference was found on the level of LARGE.

A Mann-Whitney U-test revealed no statistically significant differences between the two display sizes.

	PAD	TOUCH	MOVE
M_{TOTAL}	31.15[a] (15.36)	32.13[b] (12.91)	44.96[a, b] (19.46)
M_{SMALL}	32.44[d] (15.76)	33.37 (9.35)	46.45[d] (19.88)
M_{LARGE}	29.87 (15.43)	30.88 (15.99)	43.46 (19.65)

Table 4.6: Mean user task load. SD is shown in brackets, and statistically significant differences are indicated via raised letters (for within).

Analysis of mean values in the TLX subscales showed statistically significant differences for *physical demand* ($\chi^2(2) = 22.691$, $p < .05$), *effort* ($\chi^2(2) = 16.074$, $p < .05$), and *frustration* ($\chi^2(2) = 15.340$, $p < .05$). Table 4.7 shows the mean values for these subscales.

A Wilcoxon signed-rank test showed that the *physical demand* was statistically significantly higher for MOVE when compared to PAD ($Z = -4.251$, $p < .016$) and TOUCH ($Z = -3.354$, $p < .016$). The *effort* was also statistically significantly higher for MOVE when compared to PAD ($Z = -3.321$, $p < .016$) and TOUCH ($Z = -3.153$, $p < .016$). The level of *frustration* was statistically significantly higher for MOVE when compared to PAD ($Z = -2.909$, $p < .016$) and TOUCH ($Z = -2.791$, $p < .016$). Subscales that indicated statistically significantly different mean values are shown in Table 4.7.

A Mann-Whitney-U-Test showed a statistically significant difference for display size when comparing the subscale *performance* ($U = 54.0$, $p < .05$). For MOVE, participants rated their performance better on LARGE with $M_{LARGE} = 33.57$, (SD = 28.04) and $M_{SMALL} = 49.86$, (SD = 20.29).

	PAD	TOUCH	MOVE
Physical Demand	19.89[a] (15.76)	31.57[b] (23.56)	53.50[a, b] (24.58)
Effort	33.54[a] (21.55)	35.07[b] (19.75)	50.54[a, b] (19.94)
Frustration	23.54[a] (20.88)	22.82[b] (17.38)	39.21[a, b] (25.41)

Table 4.7: TLX subscales. SD is shown in brackets, and statistically significant differences are indicated via raised letters (for within).

User Experience

User Experience was evaluated based on the subscales of the UEQ [228]: *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation*, and *novelty*.

Irrespectively of the display size, results show no statistically significant difference in *attractiveness* depending on the input modality, $\chi^2(2) = 3.448$, $p > .05$. However, we found statistically significant differences in *perspicuity* ($\chi^2(2) = 14.673$, $p < .05$), *efficiency* ($\chi^2(2) = 12.250$, $p < .05$.), *dependability* ($\chi^2(2) = 18.780$, $p < .05$.), *stimulation* ($\chi^2(2) = 6.928$, $p < .05$.), and *novelty* ($\chi^2(2) = 21.189$, $p < .05$.). Mean values for these subscales are shown in Table 4.8.

A Wilcoxon signed-rank test showed that the *perspicuity* was ranked better for PAD compared to MOVE ($Z = -2.916$, $p < .016$), and for TOUCH compared to MOVE ($Z = -3.078$, $p < .016$). Also, the *efficiency* was evaluated likewise with $Z = -3.623$, $p < .016$ for the comparison of PAD with MOVE and $Z = -3.180$, $p < .016$ for the comparison of TOUCH and MOVE. The *dependability* was ranked higher for PAD in comparison to MOVE ($Z = -3.746$, $p < .016$) and for TOUCH in comparison to MOVE ($Z = -3.993$, $p < .016$). The post-hoc analysis showed no statistically significant differences for the scale of *stimulation* after Bonferroni correction. The *novelty* was ranked highest for MOVE and lowest for PAD. Pairwise comparisons showed statistically significant differences between PAD and TOUCH ($Z = -2.502$, $p < .016$), PAD and MOVE ($Z = -3.695$, $p < .016$), and TOUCH and MOVE ($Z = -3.182$, $p < .016$).

Additionally, a Mann-Whitney U-test showed a statistically significant difference for display size when comparing *novelty* for PAD ($U = 48.5$, $p < .05$, with $M_{LARGE} = 0.57$, (SD = 0.89) and $M_{SMALL} = 0.30$, (SD = 0.84)), and TOUCH ($U = 47.5$, $p < .05$, with $M_{LARGE} = 1.00$, (SD = 0.73) and $M_{SMALL} = 0.09$, (SD = 1.02)). In these two conditions, LARGE led to a more novel impression of PAD and TOUCH.

	PAD	TOUCH	MOVE
Perspicuity	1.87^{a} (1.16)	2.12^{b} (0.78)	$1.21^{a, b}$ (1.21)
Efficiency	1.23^{a} (0.92)	1.16^{b} (0.88)	$0.21^{a, b}$ (1.04)
Dependability	1.65^{a} (0.76)	1.81^{b} (0.56)	$0.69^{a, b}$ (1.15)
Novelty	$0.13^{a, c}$ (1.27)	$0.55^{b, c}$ (0.99)	$1.29^{a, b}$ (1.12)

Table 4.8: Subscales of the UEQ. SD is shown in brackets, and statistically significant differences are indicated via raised letters (for within). Semantic Differential: Min: −3, Max: 3.

Subjective Preferences

Concerning the subjective preferences of input modalities, we asked participants to rank them. For analysis, we gave input modalities scores based on their ranked position (e.g., 1 for the input device ranked best, 2 for the one ranked second, and 3 for the least preferred input modality). Table 4.9 shows the mean scores based on the given ranks.

	PAD	TOUCH	MOVE
M_{TOTAL}	1.71 (0.71)	1.68 (0.77)	2.61 (0.63)
M_{SMALL}	1.79 (0.70)	1.50 (0.76)	2.71 (0.47)
M_{LARGE}	1.64 (0.74)	1.86 (0.77)	2.50 (0.76)

Table 4.9: Subjective Preferences. SD is shown in brackets.

Results show that participants preferred PAD and TOUCH compared to the MOVE condition. PAD was rated best by 12 participants. Concerning the PAD, participants specifically preferred the "quick, efficient and easy interaction" (P14) and the possibility for "fast navigation" (P19). In addition, participants liked the fact that it required "less movement" (P10) than the other conditions. Fourteen participants rated TOUCH as their favorite input modality. Concerning TOUCH, participants specifically liked the "intuitive interaction" (P20) and the "direct manipulation" (P4) that allowed for a "good overview over the canvas" (P28) and a "direct way of navigating towards the items" (P11). Out of 28 participants, 19 ranked MOVE to be the least preferred input device. Reasons for that were mostly the "cumbersome" (P32) interaction as participants often felt that navigation was "jerky" (P9). In particular, using the trigger was perceived as "disruptive" (P15).

4.4 Discussion

In our discussion, we address the three research questions (RQ2.1 - RQ2.3) regarding the aspects of spatial memory, efficiency, and user satisfaction.

4.4.1 Spatial Memory

Regarding the performance of spatial memory, we measured the navigation performance (ratio of actual navigation path and shortest possible path) and participants' object location memory (distances of the placed items to the correct location in the reconstruction task). Our results concerning the two measurements match, however, they contradict the results regarding the efficiency of task completion.

Our results show that the display size has a positive impact on the navigation performance in the conditions TOUCH and MOVE. With these two input modalities, participants performed statistically significantly better in the LARGE condition. These findings are similar to previous research that a larger display positively influences spatial memory [386], however, our findings show that this is only the case for interaction modalities that provide kinesthetic cues (TOUCH and MOVE), suggesting that this effect only occurs when arm movements are sufficiently large. We could also observe that TOUCH resulted in a steeper learning curve for navigation path length across the blocks, leveling off earlier than those for PAD and MOVE (see Figure 4.2). Our findings are in line with previous research that kinesthetic cues and direct interaction improve navigation performance in panning interfaces [179, 384].

Results regarding the object location memory go even further, showing more accuracy for MOVE. This effect was statistically significantly different between PAD and MOVE for LARGE. However, differences between TOUCH and PAD were not statistically significant. This finding resonates with Palleis and Hussmann [304], as spatial memory was not found to be influenced by direct interaction in touch-based interaction. Nonetheless, the statistically significant improvement for MOVE in LARGE suggests that kinesthetic cues in gesture-based interaction indeed positively impact on spatial memory.

In summary, we conclude that embodiment impacts on the performance of spatial memory, as kinesthetic cues in the LARGE condition lead to better navigation performance when working with TOUCH and MOVE. In addition, object location memory performance is increased when interacting with MOVE, suggesting that – in gesture-based interaction – the involvement of kinesthetic cues does not only lead to a better navigation performance but also positively influences object location memory. This finding could not be replicated for touch-based interaction.

4.4.2 Efficiency

Regarding efficiency, we were particularly interested in how the different aspects of embodiment impact on participants' navigation speed and task completion time. Although we assumed that these two measurements should be correlated, our results

showed unexpected differences regarding these two measurements between the input modalities and display sizes.

As expected, PAD resulted in statistically significantly higher navigation speed compared to TOUCH and MOVE. We believe that this result is due to the 1:1 mapping of PAD to the display size. Thus, kinesthetic cues can hamper efficiency by forcing the user to engage in larger arm movements. Interestingly, there was also a statistically significant difference in navigation speed between TOUCH and MOVE, showing that TOUCH was statistically significantly faster. This finding is interesting as the difference in motor activity between these conditions is marginal. However, in addition to the arm movement, participants had to handle MOVE (e.g., pull the trigger), which might have caused them to move at a slower speed. In the TOUCH condition, there was no such device interference. In addition, MOVE required participants to lift their arms, which might have caused fatigue as another possible reason for slower navigation speed.

Our results show that display size positively influences spatial memory when using TOUCH and MOVE, as a smaller display increased the navigation speed for these modalities. This finding was expected as SMALL requires less motor activity. This was not the case for PAD, which allowed for equally fast interaction across SMALL and LARGE.

The increase of navigation speed was only partially transferred into task completion time. Although interacting with PAD still resulted in faster task completion time, TOUCH reached similar results, and the statistically significant difference found between these input modalities concerning navigation speed diminished when measuring task completion time. This is in line with previous research that did not find statistically significant differences in task completion time between direct and indirect touch input [304]. In our study, PAD and TOUCH resulted in statistically significant faster task completion times than MOVE. These results suggest that the haptic feedback in TOUCH allowed for similar task completion times by compensating for the slower navigation speed. We suspect that it might be a more accurate way of navigating to the item that allowed participants using TOUCH to reach similar task completion times than participants using the PAD. Alternatively, another interpretation might be that PAD allowed for navigation at such a fast speed that users were not able to keep up with. This resulted in the fact that the head start in navigation speed could not be held concerning the task completion time.

MOVE resulted in statistically significant longer task completion times than both, PAD and TOUCH. This shows that participants were not able to apply a strategy that compensated for the slower navigation speed, resulting in correlating results

of slower navigation speed and slower task completion time for MOVE. We suspect that they had to focus on the interaction with the device when using MOVE, which negatively affected both, the navigation speed and task completion time.

In summary, we conclude that embodiment also affects the efficiency of task completion, as larger motor activities result in slower navigation speed. However, we also showed that parts of the slower navigation speed can be compensated for in TOUCH but not in MOVE. This suggests that haptic feedback allows participants to compensate for slower navigation speed resulting in faster task completion times.

4.4.3 User Satisfaction

Regarding the user satisfaction, we were particularly interested in users' subjective rating of their task load (measured using the NASA-TLX [136]) and their user experience (measured using the UEQ [228]). Results of these two questionnaires show that irrespectively of the display size, participants rated MOVE differently compared to PAD and TOUCH.

Results of the NASA-TLX show that there are statistically significant differences regarding the task load across the three input modalities. Post-hoc tests revealed that MOVE resulted in a statistically significantly higher task load. Looking at the subscales, MOVE was rated with a statistically significantly higher task load onto the user with respect to *physical demand*, *effort*, and *frustration*. Physically handling the device in addition to the kinesthetic movement might have caused the physical demand. We believe that the novel way of interaction required users to put more effort into using the device and problems during usage might have caused frustration. When looking at the results of the UEQ, we can observe similar patterns: Here, TOUCH and PAD scored statistically significantly better in the subcategories *perspicuity*, *efficiency*, and *dependability*, reflecting the benefits of the approved and long-learned interaction modalities. As MOVE was also rated to be statistically significantly more novel than the other two modalities, we believe that it might be due to this novelty that it was scored lower on the other scales of user experience. As expected, the display size did not profoundly impact on the user satisfaction, as the only difference was that PAD and TOUCH were rated statistically significantly more novel in the LARGE condition.

In summary, we conclude that both TOUCH and PAD are rated equally well, both regarding the subjective task load and the user experience. However, MOVE was perceived to acquire a higher task load and resulted in a lower user experience. Although this is in line with our findings regarding the efficiency of interaction, it contradicts our findings regarding spatial memory. Higher efficiency and user satisfaction for PAD and TOUCH were not reflected in the spatial memory performance.

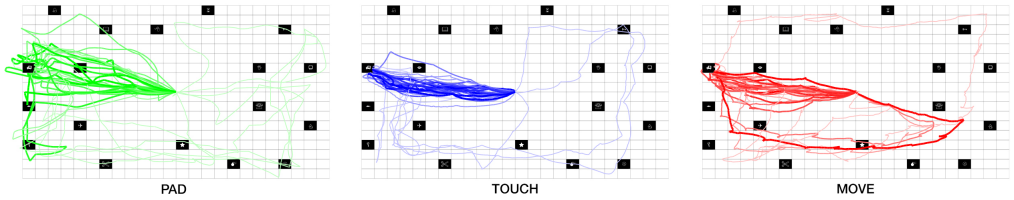


Figure 4.3: Visualizations of navigation paths for the subset of participants navigating to the same item on LARGE using PAD, TOUCH, and MOVE. The thickness of the paths visualizes the sequence of trials. Early trials are thinner.

4.5 Limitations

The accuracy and the general behavior of the input modalities might have influenced participants when solving the task: The 1:1 mapping of PAD and, thus, the resulting possible navigation speed might have been unfamiliar to participants as most of them are used to classic control-display ratios of off-the-shelf laptops. Also, the novelty and unfamiliar experience using MOVE might have influenced participants' interaction behavior. The visual representation of items, which differed in their size regarding the two display sizes, can be seen as another possible influence. The perceived visual representation of items on LARGE was larger in contrast to SMALL, which might have biased the memorization of items.

4.6 Implications for Design and Research

Summarizing, our results show that embodiment indeed impacts on spatial memory, however not in a straightforward way. Our measurements regarding efficiency showed that larger kinesthetic cues (arm movements) resulted in slower navigation speed, which also negatively influenced the overall task completion time. However, TOUCH (with its direct interaction and haptic feedback) mediated this effect by guiding towards a more accurate way of interaction. We name this phenomenon the *Efficiency vs. Spatial Memory* trade-off. The longer it took participants to use the modality and complete the task, the more they engaged with the items and remembered them – resulting in better spatial memory performance. In addition, our results also showed that embodiment, as implemented in MOVE (gesture interaction with an additional device), was rated quite negatively with regard to user experience and task workload. However, it also had a positive impact on the performance of spatial memory (maybe due to the inconvenience of the interaction). We name this phenomenon the *User Satisfaction vs. Spatial Memory* trade-off.

In the following sections, we discuss how the two trade-offs can be exploited to influence users' interaction patterns with regard to spatial memory. We discuss the

implications of our results for practitioners and show how they can be used to inform design decisions.

Efficiency vs. Spatial Memory

This trade-off is especially reflected in the strategy employed when working with PAD. For PAD, participants made use of the increased interaction speed, resulting in fast interaction and short task completion times. However, this high speed came at a cost regarding spatial memory performance, as users seemed to be less accurate when navigating towards an item and performed less accurately in object relocation. Regarding navigation performance, Figure 4.3 (participants had to navigate to the item at the left) exemplifies that participants using PAD tended to overshoot and diverted from the optimal path toward the end (when they could already see the item). This suggests that with PAD, participants were less careful to navigate accurately. Contrary to that, participants working with TOUCH and MOVE interacted at a slower pace, however, being more careful in their navigation precision (compare Figure 4.3 in the middle and on the right). Working with TOUCH, this precision led to surprisingly fast task completion times. Working with MOVE, it resulted in high object location memory.

Regarding the display size, participants working with TOUCH and MOVE benefit from larger kinesthetic movements and thus a larger display as it improves their navigation performance, however, no statistically significant differences between the two display sizes could be found regarding the object location memory. Using PAD, no benefits could be gained when working on a larger display.

For designers, knowing about the trade-off between efficiency and spatial memory can bear important implications: Depending on the task and context, designers need to account for the specific input modality. For example, desktop applications (especially for laptops) are often designed for the interaction with a touchpad. Knowing that this modality scores high on efficient interaction, designers are encouraged to compensate for the flip side of the coin: a lower performance in spatial memory. As the indirect mapping of this input leads to more speed, long navigation distances can be easily compensated for. However, this comes at the cost of a less accurate navigation and a lower accuracy in object relocation. Thus, great care should be taken to support spatial memory from an application perspective, for example, by easing target acquisition (e.g., with iceberg targets [415, p. 76], accounting for low accuracy) or by providing additional visual cues like e.g., landmarks [400], Halos [28], or Wedges [127], accounting for lower object location memory performance.

On the other hand, when designing systems for large interactive surfaces in public settings (e.g., in a museum or in a tourist information store), touch interaction is

often chosen to allow for direct manipulation. Our findings show that such a direct mapping of input and output leads to slower navigation speed, especially for large displays. On the other hand, we could also show that the haptic feedback in TOUCH supports a high precision in navigation performance. We therefore encourage designers to compensate for the slower navigation speed. We recommend to allow for interaction on different levels of detail to increase the efficiency of task completion (e.g., by using interactive overview + detail visualizations [71]).

User Satisfaction vs. Spatial Memory

This trade-off is best reflected in the interaction patterns with MOVE. Interacting with MOVE not only resulted in a low efficiency but also led to frustration. The novelty of the device hampered user interaction and caused the user to focus more on the device than on the task. This led to slower task completion times and lower user satisfaction. Although using MOVE on LARGE led to even lower efficiency (slower navigation speed and longer task completion time), it had a positive impact on spatial memory (navigation path length and object location memory). Especially when working on LARGE, the combination of kinesthetic cues and additional effort to interact with the device led to a shorter navigation path and better object location memory. This discrepancy in our findings between user satisfaction and spatial memory performance is in line with the "performance-preference dissociation" [150, 176, 286] that states that "people are not necessarily performing best with the interfaces they prefer" [176]. Although spatial memory performance was best when interacting with MOVE, participants clearly rated this input modality as the lowest.

We strongly recommend designers to carefully take into account the trade-off between user satisfaction and spatial memory in the design of interactive systems. Especially in use cases that work with gesture input (e.g., in-car settings or in sterile environments like the operating theatre), care has to be taken to make use of the high spatial memory performance to alleviate the low subjective rating concerning task load and user experience. We, therefore, suggest to allow for the customization of the interface by the user as this increases satisfaction [348] but also nurtures spatial memory due to conscious and explicit individual configuration activities.

Our findings show that, especially with gestural input, participants can well remember the navigation path to get to an item as well as the absolute position of that item on a digital landscape. However, the complexity of the input procedure itself was considered to be cumbersome and frustrating. The realization of a gesture should, therefore, place as little effort as possible on the user [348] to allow for easy and smooth interaction. Future research could also focus on the role of performing gestures without an additional device.

4.7 Chapter Conclusion

We built a Complementary Interface to study the influences of meaningful combinations of input and output modalities on participants' spatial memory, from complementary perspectives that formed our research questions: spatial memory performance, efficiency, and user satisfaction. We conducted an experiment using a well-established task and item set and studied individuals working with three input modalities and two output devices. In detail, we focused on aspects like navigation paths and object location memory (spatial memory), navigation speed and task completion time (efficiency), subjective preferences, user experience, and task load (user satisfaction).

We were able to identify the effects of meaningful combinations of off-the-shelf input and output modalities, which helped us to come up with a nuanced description of the influences of different aspects of embodiment. Relating traditional measurements for spatial memory to additional, usability-based measurements provides a first step towards interactive systems that rely on both: Exploitation of a user's spatial memory while providing efficient and satisfying interaction.

Rather than rating interaction modalities and display sizes with respect to their impact on spatial memory, our findings show that participants engage in different interaction strategies when working with different combinations. Depending on the interaction modality and the display size, we identified two trade-offs: The *Efficiency vs. Spatial Memory* trade-off shows that navigation speed negatively correlates with the navigation and spatial memory performance, and the *User Satisfaction vs. Spatial Memory* trade-off shows that a subjective user satisfaction score negatively correlates with the performance of spatial memory. Thus, the interrelation of user satisfaction, efficiency, and spatial memory leads to different usage patterns that participants take on to fulfill the task at hand.

Parts of the following Chapter 5 have been published as:

***Johannes Zagermann**, Ulrike Pfeil, Philipp von Bauer, Daniel Fink, and Harald Reiterer. 2020. "It's in My Other Hand!" – Studying the Interplay of Interaction Techniques and Multi-Tablet Activities. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Association for Computing Machinery, Honolulu, HI, USA, 1-13. DOI: https://doi.org/10.1145/3313831.3376540*

The responsibilities of this joint publication were divided as follows: I formulated the research questions, implemented the research prototype, and designed the user study. Ulrike Pfeil helped analyzing the results and supported the writing process. Philipp von Bauer and Daniel Fink conducted the user study. Harald Reiterer supervised the work.

Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

Video Figure

Pre-recorded Talk

This Complementary Interface allows to study multi-device interaction. The first Complementary Interface allowed to study meaningful combinations of multiple devices, each with a specific purpose (e.g., supporting individual and collaborative activities), and the second Complementary Interface enabled an in-depth investigation of meaningful combinations of off-the-shelf input and output modalities; also with dedicated purposes of each component. In this third Complementary Interface, this dependency is indirectly part of the user study: In the first step, interaction techniques and their usage in relation to a large number of tablets are studied from a usability perspective, which allows e.g., for design recommendations regarding combinations of interaction techniques with tablets. In the second step, however, the dependency is purposefully reduced, as participants are able to use their preferred interaction techniques with an unrestricted usage of tablets for a less restricted task – this allows us to study how participants create purposes for interaction techniques and components of this Complementary Interface.

Study Design:	Within-Subjects
Task:	Sorting Task, Sensemaking Task
Independent Variable:	Interaction Technique, Task
Dependent Variables:	Performance, User Behavior, Interrelations

The described Complementary Interface and experimental setting are used to address the following general research question. The chapter will integrate this general research question as its overall research goal and further address it in specific use-case related research questions.

RQ3: *What are the influences of interaction techniques on users' working behaviors and utilization of components (e.g., devices) in a Complementary Interface?*

5

Studying Multi-Device Interaction

Contents

Building on prior work in the area of Cross-Device Interaction, we built a Complementary Interface to shed light on the interplay of interaction techniques, device utilization, and task-specific work behavior. Although many interaction techniques exist and have been studied extensively, the difficulties of successfully interweaving the techniques with the setting (e.g., tasks and devices) and ensuring viable interaction remain [153, 256, 278]. Prior research on multi-device interaction has shown various strategies to evaluate systems and techniques that connect and combine multiple devices [50]. While different analytical and empirical approaches for various stages during the design process exist, the overall evaluation methodology for multi-device environments is still struggling to find a frame of reference to compare or evaluate existing and future techniques and systems, resulting in a rather fragmented research landscape. Brudy et al. [50] highlight the need for improving evaluation strategies to finally "unify empirical and technical [. . .] work into one stream of research."

We tackle this challenge with an experiment that provides different but complementary angles of evaluation using an abstract classification task and a less restricted sensemaking task – treating the experimental task as an independent variable. This methodological combination results in a holistic evaluation approach with different types of tasks and both, qualitative and quantitative measurements that provided insights into the comparison of interaction techniques, their impact on device utilization, and their adoption across tasks.

In contrast to the Complementary Interfaces of the previous chapters, this Complementary Interface does not pre-define the purposes of each component. In Chapters 3 and 4, each device and component fulfilled a pre-defined purpose as well as task-related and functional interdependences: e.g., input, output, or to support shared activities – representing the independent variables of the specific experiments. In this Complementary Interface, however, the methodological approach of using the task as an additional independent variable allows to control the interdependence of components in the first task; removing this interdependence for the second task provides insights into participants' needs and how they assign purposes to individual components (e.g., interaction techniques and devices).

Figure 5.1: This Complementary Interface allows to study multi-device interaction with tablets. This figure shows a person using multiple tablets in combination to sort documents.

The goal of our experiment was to better understand the implications of different interaction techniques and to contribute a nuanced description of their effects and influences on users' working behaviors and utilization of components (i.e., interaction techniques and devices). We address this overall research question by investigating the following research questions both qualitatively and quantitatively:

RQ3.1:*How do interaction techniques influence work performance?*
RQ3.2: *How do interaction techniques influence work behaviors?*
RQ3.3: *How are interaction techniques, device utilization, and task-specific activities interrelated?*

A distinctive attribute of this experiment is that we tackle these questions not only using qualitative and quantitative metrics but also from different perspectives: 1) Using an abstract task (see *Task 1 – Abstract Classification Task*) allows us to investigate ***RQ3.1*** and ***RQ3.2*** to guarantee for a high internal validity. 2) Providing the ability to choose which technique to use in combination with a sensemaking task (see *Task 2 – The Murder Mystery*) provides a less restricted and more realistic scenario that allows to study participants' activities with higher ecological validity (***RQ3.3***). Additionally, this allows us to see how participants include their preferred technique a) in general or for specific sub-tasks, b) in combination with other techniques, and c) to explore how interaction techniques, utilization of devices, and task-specific activities are interrelated. This methodological combination results in a holistic evaluation approach that does not only provide qualitative and quantitative insights into relative strengths and weaknesses of interaction techniques (similar to [146]) and the usage of a large ecology of devices but can also serve as a first step of an evaluation framework [50].

5.1 Experiment

We conducted an experiment with the two within-subjects factors *task* and *interaction technique*. The order of *task* was predefined, with all participants finishing task 1 before starting task 2. We counterbalanced the factor *interaction technique* with the order of the techniques only for task 1 and allowed participants to use the techniques as favored in task 2 (similar to [146]). This study design allowed us to compare the techniques independently (task 1) and explore how participants use the techniques and utilize the provided large ecology of devices (task 2). The remainder of this section provides an overview of the interaction techniques, the system in general, the two tasks, participants, procedure, apparatus, and data collection.

5.1.1 Interaction Techniques

Concerning the interaction techniques, we defined three conditions: 1) MENU, 2) TRAY, and 3) PICK'n'DROP. They constitute three types of state-of-the-art and robust-to-use interaction techniques. These techniques do not require a) additional user engagement with devices (e.g., rotating or tilting) or b) extra hardware augmentation (e.g., tracking cameras), which allows for a valid comparison of techniques and

an investigation of technique-independent device utilization across both tasks. We decided to allow participants to transfer multiple sticky notes at the same time as Scott et al. [361] have shown that this increases efficiency.

Interaction Technique 1 – MENU

The first interaction technique was inspired by context-menus: Pop-up and context-menus are standard features of everyday WIMP-applications (e.g., copying and pasting text from one application to another), mobile devices (e.g., copying and pasting locations from map apps to messaging apps), across different types of devices (e.g., using Apple's Universal Clipboard[1]), and context-menus have been used as state-of-the art baseline conditions [325]. Similarly, we designed MENU as follows (see Figure 5.2): The MENU appears by performing a long press interaction (200ms) on a sticky note. Tapping on the now visible cut button cuts this sticky note. Now, users can perform another long-press interaction on any position on any tablet – triggering the paste button to appear. Tapping on it will paste the sticky note at this position. To cut multiple sticky notes, the MENU has to be used sequentially. Multiple pasted sticky notes are piled with a randomized offset of a few pixels to increase visibility.

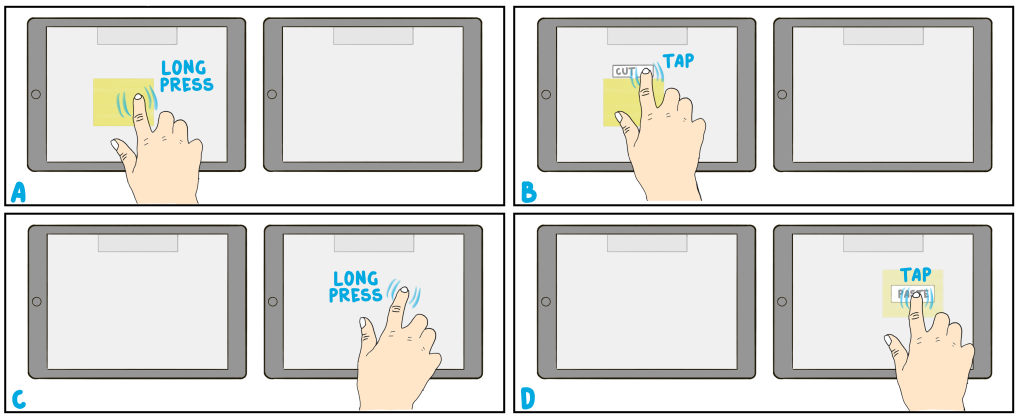


Figure 5.2: Transferring a sticky note using MENU.

Interaction Technique 2 – TRAY

The second interaction technique was inspired by previous research (e.g., [37, 257, 361, 367]) that has shown that participants can benefit from *drag and drop* interactions, where the drop target creates a digital bridge between devices. We designed TRAY as follows (see Figure 5.3): A digital tray was shown in the middle of and at the upper edge of each tablet. Sticky notes could be transferred by dragging and dropping them on the TRAY. As a result, the size of the sticky note was decreased to fit the TRAY – maintaining the readability of the sticky note's text – and the sticky

[1]Apple Universal Clipboard: https://support.apple.com/en-us/HT209460, – last accessed 01.11.2023

note was then visible on every tablet in the area of the TRAY. By dragging the sticky note out of the TRAY and dropping it on any surface, it is transferred to the currently used tablet. With this technique, users are also able to transfer one or multiple sticky notes at the same time across multiple devices.

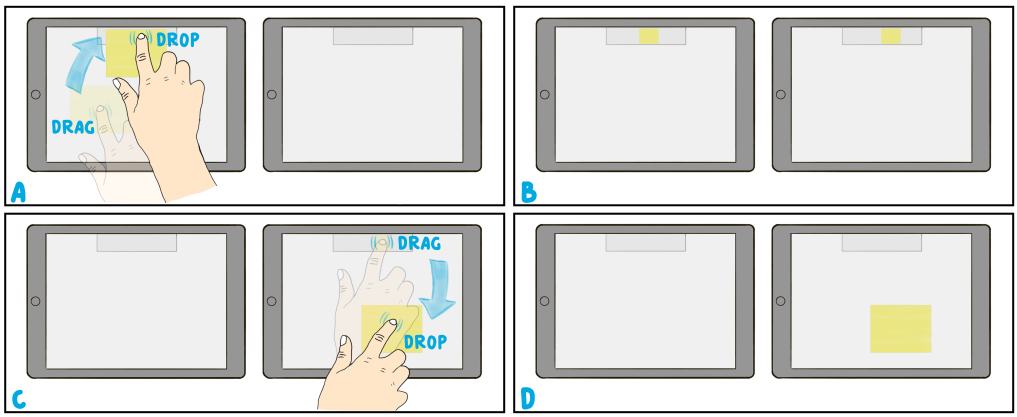


Figure 5.3: Transferring a sticky note using TRAY.

Interaction Technique 3 – PICK'n'DROP

The third interaction technique was inspired by the same-named technique by Rekimoto [333]. In line with recent research [163, 301], we allow users to pick and drop sticky notes with their fingers. In our case, PICK'n'DROP makes use of the well-known *pinch-to-zoom* technique (see Figure 5.4). Sticky notes are picked up (i.e., cut) when a pinch-in gesture is performed on them. Performing a pinch-out gesture on, e.g., another tablet, drops this sticky note at that position. With this technique, users are also able to transfer one or multiple sticky notes at the same time across devices by picking them sequentially and dropping them. Multiple pasted sticky notes are piled with a randomized offset of a few pixels to increase visibility.

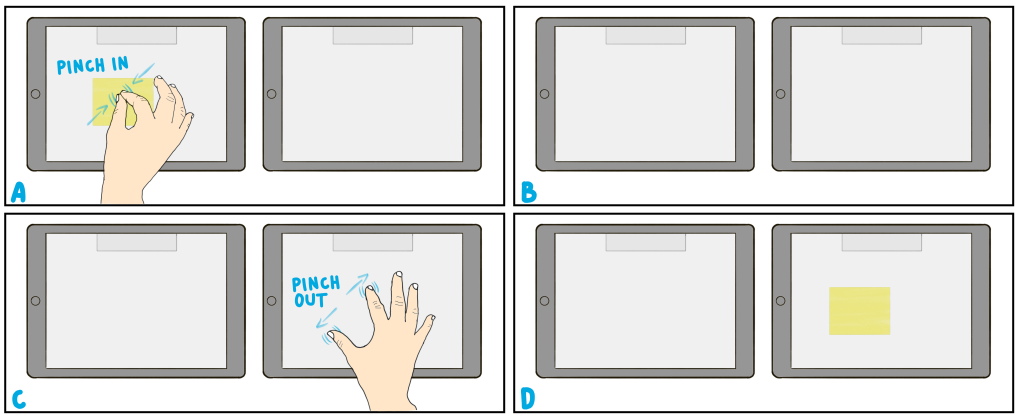


Figure 5.4: Transferring a sticky note using PICK'n'DROP.

5.1.2 System Description

This Complementary Interface enables users to use multiple tablets to work with pieces of information – represented as digital sticky notes – on two levels: Users can a) freely move and arrange sticky notes on a single tablet and b) transfer them across multiple tablets using the described interaction techniques. Especially for the second task (see *Task 2 – The Murder Mystery*), this allows for spatial arrangements to prioritize and compare information [405] and to find connections between different events, persons, and places. We implemented a web-based system using the JavaScript framework Meteor without additional hardware augmentation (e.g., camera tracking systems) that allows for ad-hoc use of multiple devices [50].

5.1.3 Tasks

We used two different tasks for our study: In the first task, we chose to use an abstract classification task as it guaranteed a high internal validity when comparing different interaction techniques. We were also interested in the effects and influences of the different interaction techniques on participants' working behavior using a more realistic sensemaking task. We pre-defined the order of tasks (first task 1, then task 2) to study revealing intended learning effects.

Task 1 – Abstract Classification Task

To provide a documented and sufficiently demanding task, we were inspired by the abstract classification task used by Liu et al. [240], who studied the influence of display size on single users' interaction with a wall-sized display. The main components of the tasks are discs – labeled with different letters. Participants have to move these discs to their specific container (i.e., a display) to classify them correctly. The color of the discs helps participants to solve the task: If not correctly classified (i.e., in the wrong container), discs show a red background color and turn green when transferred into a display where all discs share the same label. The same task was used, e.g., by Jakobsen and Hornbæk [175], who investigated the influences of locomotion on single users' interaction with a wall-sized display. This task was also used to study collaborative activities, either co-located [241] or remote [17].

In light of the prevalent use of this task, we took it as inspiration for our first task. As the task itself was initially developed for wall-sized displays and mostly used for studies in such settings, we decided to adapt it to multi-device settings: Following previous research (e.g., [130]), we decided to set the number of available tablets to 8 (see section *Apparatus* for detailed descriptions). Participants were allowed to have as many items on their tablets as they wished, yet we defined a total number of 40 items (represented using 8×5 letters), resulting in a final solution of 5 items

per tablet in an ideal number of 35 item movements. This was a trade-off decision of having enough data without overwhelming the participants. Items are represented as sticky notes with a size that allows the arrangement of 6 sticky notes per tablet with no overlap (in line with [240]). We decided to visualize items as sticky notes instead of discs to increase the comparability with the second task. Sticky notes could be moved across tablets using one of the interaction techniques. We labeled sticky notes with eight different letters: C, D, E, F, H, K, N, and R. These letters were chosen as they guarantee equal legibility according to the BS 4274-1:2003 vision test standard [164]. Each letter was used to label five sticky notes with a standard computer font (font size: 12pt).

We decided against a pre-configuration of correctly classified items (i.e., having 3 to 4 correctly classified items and only 1-2 wrong ones per tablet) as our pre-tests have shown that participants solved the task on different levels of difficulty in less than a minute. Thus, we decided to omit such a pre-configuration in favor of keeping the task sufficiently demanding. Also, a pre-configuration could influence workplace configurations. As we were interested in our participants' utilization of the tablets (e.g., spatial configurations), we decided to place all sticky notes initially on one tablet that was on top of the stack of the tablets. To ensure equal difficulty, we randomized the order in which sticky notes were represented initially.

Task 2 – The Murder Mystery

As a counterpart to our first task, we opted for a sensemaking task as a real-world activity. The "Stegosaurus" task and data set of the 2006 VAST challenge [121] was often used in previous publications (e.g., [282]), and Chapter 3 has shown that this task indeed offers insights into the usage of multiple devices, yet solving the task can be challenging and task completion times of up to 90 minutes or longer are no exceptions. Thus, we decided to opt for a less time-consuming but also complex sensemaking task: The murder mystery suggested by Stanford and Stanford [374] is a crime-solving task and was also used before by Wozniak et al. [419] in their multi-tablet system. In the task, multiple people were involved in a murder mystery, which happened during the night. Participants had to find the murderer, the weapon, the time and place of the murder, and the motive by combining clues and hints written on sticky notes (similar to [419]). As suggested [374, 419], we increased the difficulty by adding distracting information. In total, there were 40 pieces of information represented as sticky notes (font size: 12 pt), which allows for better comparability between our two tasks. Participants had to understand the content, filter non-important information, and connect essential facts that lead to new insights and suppositions to come up with answers to the murder mystery.

5.1.4 Participants

Twenty-four participants (14 female, 10 male, and 0 non-binary) were recruited for the user study. The mean age was 23.88 years (SD = 4.00). All participants had normal or corrected to normal eyesight; consequently, they had no problems with the size of the sticky notes or the font size. Twenty participants were undergraduate students – the other four participants were university employees or Ph.D. students. Participants had a mixed background ranging from psychology to biology, educational science, and law. All participants were familiar with touch interfaces and reported the frequent handling of mobile devices like tablets and smartphones. To recruit participants, we used postings and flyers looking for people who are interested in new technologies and crime mysteries at our local campus. This helped us to find participants who were motivated by our chosen tasks.

5.1.5 Procedure

At the beginning of the study, each participant was asked to fill out a questionnaire about demographics and tech-savviness. Then, each participant received an introduction to the first task. After that, the first interaction technique was shown using a demo item set. Participants had time to explore the functions of the system and the specific interaction technique with a demo item set until they felt comfortable using it. After that, the experimenter asked participants to start the first task with the first interaction technique with the remark to be as quick and accurate (e.g., as few interaction steps) as possible. After that, participants were asked to fill in a NASA-TLX questionnaire [136]. This sequence was the same for all three interaction techniques. To avoid learning effects, we counterbalanced the order of interaction techniques and randomized the sequence of the shown sticky notes. After that, participants were asked to complete a questionnaire concerning subjective preferences. Then, the experimenter introduced participants to the second task: They were asked to solve the murder mystery and were free to use any of the three different interaction techniques (cf. [146]) and as many tablets as desired. Finally, each participant filled out a post-questionnaire concerning clarification of the usage of the interaction techniques in this phase. Each session lasted about one hour in total, and we reimbursed participants for their time.

5.1.6 Apparatus

The setting consisted of an office table (1.20×0.80 m) on which we stacked eight Apple iPads (9.7″) in the middle of and at a long side of the table (cf. Figure 5.1). We attached felt pads to the bottom of each tablet for smoother movements (cf. [316]). Participants sat on a chair on the opposite side of the table. Next to them was a small

additional desk for filling out questionnaires. We used a camera with a bird's-eye view to see interactions, working behaviors, and device utilizations.

5.2 Data Analysis

We employed several data collection methods to investigate our research questions:

RQ3.1 – **Performance**: We measure task completion time, number of object transfers, the ratio of object transfers per interaction, tablet utilization, and subjective rankings.

RQ3.2 – **Work Behavior**: To explore work behavior, we analyzed participants' activities using video codings.

RQ3.3 – **Interrelations**: We analyzed frequencies of interactions, utilization of tablets, and rationales of using the techniques for the murder mystery.

5.2.1 Video Data

We developed a coding scheme to analyze the video data of the first task. We focused on participants' work behavior (e.g., sorting sticky notes). After the initial viewings of our videos, we started to distinguish between different codes that offer a sound description of the video data and apply it to our research questions. This coding scheme was refined until we felt that it was saturated. An inter-coder reliability test was conducted for six samples (two per technique; in total, about 25 minutes) of video data to ensure objective coding by our two independent coders. Cohen's Kappa of 0.70 revealed substantial inter-coder reliability for the final coding scheme presented below.

S – **Sorting**: A participant was explicitly sorting sticky notes (e.g., piling all sticky notes with the letter "*C*") on one tablet.

D – **Distributing**: A participant distributed one or multiple sticky notes to another tablet.

E – **Exploring**: Similar to *S – Sorting*, but in this case, participants were implicitly sorting sticky notes while explicitly searching for a sticky note in a pile of sticky notes.

We defined all codes as mutually exclusive state events. Thus, we were able to compare the duration of activities and their sequences. The video codes helped us to elicit the main aspects of work behavior. Also, it allowed us to focus on specific sub-parts of the video for detailed qualitative observations.

5.2.2 Data Logs

We collected data logs of participants' interaction with the system, namely when and how often participants transferred sticky notes from one tablet to another (including an internal ID for each tablet), the ratio of object transfers per interaction, and task completion times. We did this to analyze whether the interaction technique affected how and how often participants interacted with sticky notes and tablets.

5.2.3 Questionnaires

We used the NASA-TLX questionnaire [136] to investigate the subjective task load caused by the three interaction techniques. In addition to this, we used one questionnaire for subjective preferences and another questionnaire after the final task to clarify participants' choice of used interaction techniques.

5.3 Findings

We report our findings in relation to our research questions. We focus on differences across the three interaction techniques with respect to participants' performance (***RQ3.1***), their work behavior (***RQ3.2***), and the interrelations of interaction techniques, device utilization, and task-specific activities (***RQ3.3***). As the data violated the assumption of a normal distribution, we decided to use a non-parametric approach to analyze the data statistically. For the overall tests on statistical significance, we used a Friedman test. If this test showed statistically significant differences, we employed a Wilcoxon signed-rank test as a post-hoc analysis, including Bonferroni correction.

5.3.1 RQ3.1 – Performance | Abstract Classification Task

We differentiate between task completion time, the number of sticky note transfers, the ratio of object transfers per interaction, tablet utilization, task load, and subjective preferences.

Task Completion Time

We measured participants' task completion times: M_{MENU}=241.00s (SD=58.58), M_{TRAY}=177.54s (SD=28.31), and $M_{PICK'n'DROP}$=176.92s (SD=36.27). This difference was statistically significantly different ($\chi^2(2)$=25.00, $p < .05$). A post-hoc analysis revealed statistically significant differences between MENU and TRAY (Z=3.972, $p < .016$) as well as between MENU and PICK'n'DROP (Z=3.972, $p < .016$).

Number of Transfers across Devices

We measured how often participants transferred sticky notes across the available tablets. The optimal number of moves is 35. A Friedman test revealed no statistically significant differences when comparing the three different interaction techniques with M_{MENU}=37.04 (SD=3.13), M_{TRAY}=36.13, (SD=1.36), and $M_{PICK'n'DROP}$=38.04 (SD=2.80).

Ratio of Object Transfers per Interaction

As our interaction techniques allowed participants to change the ratio of object transfers per interaction (e.g., picking up three sticky notes sequentially and then dropping them all on another tablet), we analyzed if this ratio varied across the conditions. Here, we define *1* as a 1:1 ratio (i.e., picking up one sticky note and dropping it afterward) and *5* as a 5:1 ratio (i.e., picking up five sticky notes and dropping them afterward at once). For TRAY, the ratio could be varied by e.g., dragging multiple sticky notes via multiple fingers out of it. A statistical analysis revealed no statistically significant differences with M_{MENU}=1.64 (SD=0.67), M_{TRAY}=2.40 (SD=1.57), and $M_{PICK'n'DROP}$=1.79 (SD=1.14).

Time-weighted Tablet Utilization

We analyzed participants' utilization of the provided tablets using the time-weighted tablet utilization score (TUS [316]). The TUS relates the number of used tablets to their usage time and the overall task completion time. In our case, a TUS of 1.0 would indicate the usage of one tablet for the entire session. Similarly, a TUS of 8.0 would show a 100% usage of eight tablets. Our analysis revealed a mean TUS of 5.75 (SD=1.07) for MENU, 5.01 (SD=1.63) for TRAY, and 5.74 (SD=1.45) for PICK'n'DROP. Statistical analysis revealed no statistically significant differences.

Task Load

We used a NASA-TLX questionnaire [136] to investigate participants' task load. A statistical analysis revealed no statistically significant differences for the overall scores with M_{MENU}=27.47 (SD=16.44), M_{TRAY}=19.79 (SD=12.16), and $M_{PICK'n'DROP}$=25.73 (SD=17.35).

We found a statistically significant difference for *Physical Demand* ($\chi^2(2)$=7.649, $p < .05$). A post-hoc analysis showed statistically significant differences for a comparison of TRAY and PICK'n'DROP ($Z = -2.732$, p<.016) with M_{TRAY}=16.46 (SD=16.58), and $M_{PICK'n'DROP}$=27.50 (SD=24.89).

Subjective Rating

We asked participants to rank the techniques based on their subjective preferences. For analysis, we gave interaction techniques scores based on their ranked position (e.g., 1 for the technique ranked best, 2 for the one ranked second, and 3 for the least preferred technique). The mean scores are: *MENU*: 2.71 (SD=0.55), *TRAY*: 1.67 (SD=0.70), and *PICK'n'DROP*: 1.62 (SD=0.71). The overall analysis showed a statistically significant difference ($\chi^2(2)=18.083$, $p<.05$). A post-hoc test showed a statistically significant difference for pairwise comparisons of TRAY and MENU ($Z=-3.376$, $p<.016$) and PICK'n'DROP and MENU ($Z=-3.542$, $p<.016$).

PICK'n'DROP was rated best by 12 participants. Participants described it as "*like picking up real objects*" and that it was "*fun*", "*fast*", and its "*fluent interaction*" was "*intuitive*". Eleven participants rated TRAY as their favorite interaction technique – they liked the "*familiar handling*" of drag interactions, described it as "*like a game*", and found it "*interesting to see sticky notes appear on other tablets*". 18 ranked MENU to be the least preferred technique. Although described as "*you don't have to think, it's like on a PC*", participants disliked the additional sub-tasks to invoke the context MENU itself.

***RQ3.1 – Findings**: The three interaction techniques partly influenced participants' performance: TRAY and PICK'n'DROP are participants' choice, and both are faster than MENU – with PICK'n'DROP showing a higher subjective physical demand than TRAY.*

5.3.2 RQ3.2 – Work Behavior | Abstract Classification Task

To investigate work behavior, we focus on our video analysis and the resulting types of work behaviors.

Results of Video Analysis

Table 5.1 shows an overview of our three video codes in relation to the three interaction techniques. All conditions show an emphasis on *Distributing*, which can be directly mapped to the task: It was necessary to distribute all sticky notes to solve it. Analyzing this quantitative comparison using Friedman tests revealed no statistically significant differences. Yet, our additional qualitative analysis revealed differences in participants' workflows that cannot directly be linked to one of the interaction techniques. We discuss these types of work behaviors from a qualitative perspective in the next subsection.

	MENU	TRAY	PICK'n'DROP
Sorting	16.96% (13.43)	20.33% (21.62)	16.38% (19.55)
Distributing	71.96% (15.29)	66.42% (20.89)	71.96% (22.63)
Exploring	3.00% (10.24)	4.13% (13.08)	3.25% (14.12)

Table 5.1: Mean values of relative duration of activities. The remaining relative duration (about 8%) was used to arrange workplaces. The standard deviation (SD) is shown in brackets.

Types of Work Behaviors

Analyzing the frequencies, sequences, and duration of the three different video codes helped us to get a rich understanding of participants' work behaviors. Although the quantitative analysis did not reveal any differences, we were able to identify three types (T1 – T3) of problem-solving approaches by analyzing all 72 sessions (24 participants×3 conditions) qualitatively: 1) *T1: One Tablet, Many Piles*, 2), *T2: Many Tablets, One Pile Each*, and 3) *T3: Mixed*.

T1: One Tablet, Many Piles – Here, sorting of sticky notes was the main activity. Participants piled up sticky notes with the same letter on the initial tablet. They creatively circumvent the spatial limitations of a single tablet (e.g., it was possible to arrange six sticky notes as a grid without overlaps, but they moved piles at/over the edge of the tablet to overcome spatial restrictions) and moved every "new" sticky note to its corresponding pile. Eventually, participants came up with eight piles, reflecting the eight different letters/classes of sticky notes. Then, they transferred each pile to other tablets.

T2: Many Tablets, One Pile Each – Here, participants skipped sorting and instantly transferred each top-layer sticky note (the sticky note that was shown on top of the pile of all sticky notes) to another tablet. They worked through the initial pile of sticky notes and finished the task when all sticky notes were distributed. Rather than sorting sticky notes on a single tablet, they used the affordances of physical devices as containers.

T3: Mixed – This type of working behavior combines attributes from the other types. Participants either explicitly or implicitly sorted sticky notes: When sorting explicitly, participants piled up sticky notes – similar to *T1* – but transferred them to another tablet either when there was no additional space for new piles on the initial tablet or as soon as a pile was complete (e.g., all five "*E*" sticky notes are piled up). Transferring them to another tablet freed space on the initial tablet, which was then used again to pile up other sticky notes. When sorting implicitly, participants were looking for a specific letter – exploring the sticky notes while spreading them.

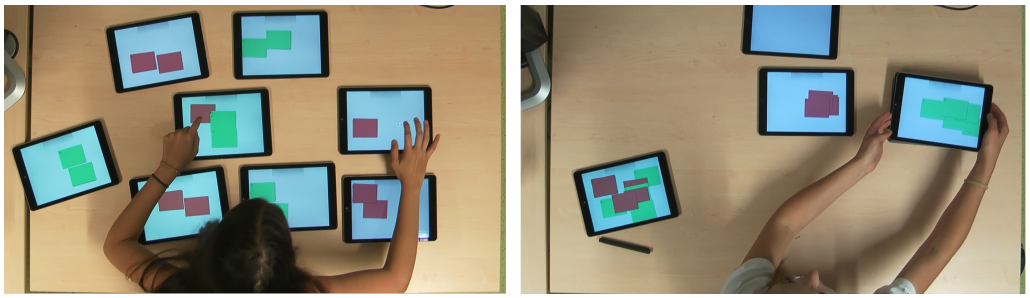


Figure 5.5: A participant uses both hands (MENU, left), and a participant stacks tablets (right).

In total, 14% of all sessions were classified as *T1*, 32% as *T2*, and 54% as *T3*. When linking the different types of work behavior to our three different interaction techniques, we found that 16 out of 24 participants acted as a *T3* using MENU. For TRAY, there were 10 *T3*, 9 *T2*, and 5 *T1*. Using PICK'n'DROP, there were 12 *T3*, 9 *T2* and 3 *T1*.

Additionally, our qualitative observations allowed us to gain insights into how participants utilized the large device ecology: 11 participants spread all tablets before they started with the actual task in all three conditions, 5 participants always added tablets incrementally, and 8 participants switched from an incremental workflow in their first condition to spreading all tablets initially in the remaining conditions. Further, participants used both hands to speed up their activity and stacked 'finished' tablets (see Fig. 5.5).

RQ3.2 – Findings: *From a quantitative perspective, there were no differences between the interaction techniques regarding the amount of time they explored, sorted, and distributed sticky notes. Our qualitative analysis revealed not only problem-solving approaches that participants used to solve the task across all techniques, but also various subtleties of interaction with a large ecology of devices.*

5.3.3 RQ3.3 – Interrelations | The Murder Mystery

Here, we differentiate between frequencies of using the interaction techniques, participants' tablet utilization, and participants' rationales for incorporating the interaction techniques into their workflow for the second task. After completing the murder mystery task, participants were asked five questions about the plot. For analysis of the results, we scored participants' correct answers. On average, participants answered 3.96 (SD=1.17) out of 5 questions correctly, indicating that the task was challeng-

ing enough, yet not too demanding. The mean task completion time was 911,08s (SD=469,44), and participants worked with 7.38 (SD=0.97) tablets on average to solve the task.

Number of Interactions

We analyzed the number of interactions that were used to solve the task, the times participants used one of the three interaction techniques for it, and how often one of the techniques was chosen by participants. On average, participants transferred sticky notes 48.35 (SD=9.70) times to solve the task.

As we did not restrict participants in their usage of the interaction techniques, we analyzed the number of interactions per technique. Also, it was possible to combine the interaction techniques (e.g., picking up a sticky note using PICK'n'DROP and placing it on another tablet using MENU) – thus, we divided this analysis accordingly. On average, participants used MENU 3.29 (SD=9.42) times to cut and 3.63 (SD=9.41) times to paste sticky notes. They used TRAY 21.19 (SD=22.59) times, and PICK'n'DROP was used 23.88 (SD=23.56) times to pick and 23.46 (SD=23.25) times to drop sticky notes. 16 out of 24 participants combined the different interaction techniques and incorporated them into their workflow. Six participants used MENU, 20 participants used TRAY, and 17 participants used PICK'n'DROP.

Tablet Utilization

We analyzed participants' utilization of the provided tablets using the time-weighted tablet utilization score (TUS) [316]. Our analysis revealed a mean TUS of 6.50 (SD=0.99).

Work Behavior

We did a qualitative analysis of the video data to see if the codes used for the abstract task and, thus, the different types of work behaviors also fit to the sensemaking task. Here, we rarely saw sorting activities – participants focused on distributing sticky notes to available tablets and clustering them semantically e.g., by persons or places. This work behavior is similar to the *T2* behavior, as participants used the physical affordances of the tablets as containers for different topics. This is further reflected in participants' TUS: Here, the TUS for the sensemaking task (6.50, SD=0.99) indicates an early utilization of a large number of tablets and was even higher than in the abstract classification task.

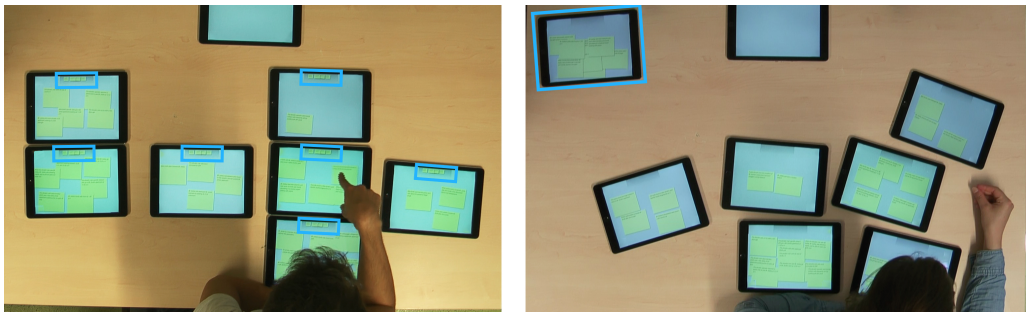


Figure 5.6: Participants dealing with non-important pieces of information: Repurposing TRAY (left) and a dedicated tablet as "trash bin" (both highlighted in blue).

Subjective Rationales for Using the Interaction Techniques

As a final step, we asked participants about their usage of the three different interaction techniques. Participants had to provide reasons for (not) using each of the interaction techniques.

Participants liked MENU because it is "*fun*", "*easy*", and it was the "*fastest technique to move sticky notes from one tablet to another*". Participants used it to "*select sticky notes that are related to persons and collect them on a single tablet.*" Participants chose MENU instead of PICK'n'DROP in case of "*overlapping sticky notes*" – indicating spatial limitations of PICK'n'DROP. Participants valued the *utilization of both hands*, as it was "*more natural to use both hands and thus, more fun and faster*". Yet, participants criticized the *integration into the workflow* and did not use MENU because of "*too many steps*", which made it "*inconvenient*". Participants "*didn't like it*" because it "*was easier to stick to one technique*" and "*sticky notes disappeared*" when cut – highlighting the importance of the *visibility of system status* [285].

Participants liked TRAY because it was "*handy*", they used it to "*sort sticky notes on different tablets*", and found it "*easy to distribute information on different displays.*" They used it when they were "*aware where a sticky note fits*", but also to "*get needless sticky notes out of view*" and called it a "*trash bin*" (see Figure 5.6). On the other hand, they did not use TRAY because sticky notes had to be moved "*sequentially*" (*integration into the workflow*); they thought they would "*have forgotten content while being in the TRAY*" (*visibility of system status*) because the task "*was too complex*" for using TRAY.

Participants liked PICK'n'DROP because they wanted to "*cherish information*", "*the technique fits best to the task*", and it "*feels natural*". They also described it as "*fun and you can use both hands*", as "*a fast way to move information*", and it was "*useful to move multiple sticky notes at the same time.*" Interestingly, they mentioned that "*it*

feels like picking up a real object" and that "*having this feeling of holding a sticky note in your right hand hinders you to drop it with the left hand, because it's in the other hand!*" (*utilization of both hands*). Participants did not use PICK'n'DROP because using "*TRAY was easier*" and "*more practical*", and they chose to use "*TRAY, because I wanted to focus on the murder mystery*" (*integration into the workflow*).

***RQ3.3 – Findings**: Our analysis showed that participants utilized a large number of tablets to solve the task at hand. The TUS for task 2 was even higher than in task 1. This high score might be based on extensive training during the first task. Participants' choice of interaction techniques for the sensemaking task reflects their subjective preferences for the abstract classification task. We identified quality features for interaction with this Complementary Interface based on participants' rationales for using the techniques: Visibility of status, utilization of both hands, and smooth integration into the workflow.*

5.4 Discussion

We discuss our findings based on the topics of legacy bias and work behaviors, quality features for successful interaction, and reflect on our holistic evaluation approach using different types of tasks and both, qualitative insights and quantitative measurements. We additionally provide implications for future research and design.

5.4.1 Legacy Bias and Work Behavior

Plank et al.'s evaluation of a collaborative multi-tablet system revealed that participants worked with one or two tablets to solve an analytic task [316]. The authors link this finding to a legacy bias, indicating that participants are influenced by their pre-existing knowledge of prior interfaces and technologies. They question if an extensive training in using multiple devices (similar to Morris et al.'s *priming* technique (2014)) can facilitate users in utilizing more devices and, thus, reduce the legacy bias (one of the key challenges for future cross-device interaction [50]). Based on our results, we identified two ways that might help to reduce this legacy bias: 1) ***Guidance*** and 2) ***Training***. We guided participants in using a large number of tablets with our abstract classification task. Participants had to utilize all available tablets to solve the task. This showed, on the one hand, the apparent fact that participants do utilize larger numbers of devices if they have to, yet, on the other hand, this can lead to dif-

ferent work behaviors – characterized by utilizing multiple devices at different points in time. We believe that guiding users is a potential way to reduce the legacy bias by learning the benefits of utilizing multiple devices, each adding value and increasing the quality of interaction to solve the task at hand.

Our qualitative analysis revealed three types of work behaviors: 1) *T1: One Tablet, Many Piles*, 2) *T2: Many Tablets, One Pile Each*, and 3) *T3: Mixed*. While working on the abstract classification task, participants were aware that they would not be surprised by the contents of sticky notes on the initial pile. It was clear that there are sticky notes with eight different letters that could be *explored*, *sorted*, and *distributed*. This can explain individual differences in work behavior. The different types of work behavior can also be at least loosely linked to the legacy bias and the transaction costs [153] of utilizing multiple tablets in the workflow. Participants had to involve all available devices at some point in time to solve task 1. The *T1* types were more tentative in utilizing multiple tablets at an earlier stage, compared to the *T2* and *T3* types that both involved other tablets after a short period of time. This differentiation in three types of work behavior was not present in the sensemaking task: Here, all participants focused on distributing activities (i.e., utilizing a larger number of devices early), e.g., to gain an overview, find suspicious activities, and finally come up with a solution. This behavior is connected to the lack of knowledge of the content of the piled sticky notes but might also be related to the previous training phase of working three times with a large device ecology during the abstract classification task.

5.4.2 Quality Features for Interaction with Complementary Interfaces

The importance of the *visibility of system status* and transparency of interaction was emphasized: Participants complained that "*sticky notes disappeared*" in MENU and in PICK'n'DROP. Using those techniques, participants had to remember which sticky notes were currently in a non-visual clipboard state – a state that was always visible in TRAY. We favored efficiency [361] over habits as participants were able to transfer multiple sticky notes at the same time – a common practice in WIMP interfaces is to only allow one element (e.g., a text fragment) in the clipboard state. Participants liked this feature and described that "*it fits to the task*". Although this additional cognitive effort was not reflected in terms of mental demand measured via a NASA-TLX, we believe that the number of elements in the clipboard state should be balanced for the task at hand. In our case, participants were perfectly capable of dealing with this feature as it fits with the abstract classification task – participants only had to remember single letters of a limited set. Remembering additional contextual information in the murder mystery or other tasks requires more cognitive capacities and thus should be supported in terms of visibility and transparency either via a visual clipboard (as

in TRAY), an additional overview device [49], or by limiting the number of elements in clipboard state.

Participants successfully made use of *bi-manual approaches* to solve both tasks. They emphasized the "*naturalness of using both hands*", which was "*more fun and faster*". Although this is a clear contrast to transferring objects in the real world and also described for PICK'n'DROP ("*It's in my other hand!*"), some participants valued this possibility, and thus, it shows a *reality vs. expressive power tradeoff* [172]. We argue that such interaction techniques should be carefully designed to be performed by both dominant and non-dominant hands.

Performing interaction techniques requires effort. Participants rated PICK'n'DROP with higher physical demand and mentioned that it "*was easier to stick to one technique*" (MENU), the murder mystery "*was too complex*" for using TRAY, and they did not use PICK'n'DROP because they "*wanted to focus*" on the task at hand. Although there was no difference in participants' subjective task load when comparing the interaction techniques, we argue for carefully designing interaction techniques that *interweave with the workflow* and to allow focusing on the task at hand.

We observed participants not only creatively *repurposing* TRAY as a trash bin, but they also used dedicated tablets as trash containers – focusing on the essential pieces of information. Additionally, they mentioned that TRAY was particularly helpful when "*aware where a sticky note fits*". In this experiment, all available tablets looked the same – with no differences in color, shape, or size. Although Conductor's color association was not perceived as beneficial [130], we think future research could investigate the influences of different form factors e.g., as design alternatives of Complementary Interfaces to see if e.g., differently sized (mobile) devices lead to the association of different purposes (e.g., a trash bin).

5.4.3 Towards Complementary Evaluation Approaches

Brudy et al. [50] discuss strategies to evaluate cross-device interactions and emphasize the need for more work on evaluation methods to "support human activities in a cross-device ecology or to meaningfully compare cross-device interaction techniques [50]." We addressed this issue with our methodological approach of using the experimental task as an independent variable in this Complementary Interface. With this, we employed different but complementary angles of evaluation that allowed us to study aspects of usability extended by insights into work behaviors.

User studies in HCI research usually involve specific tasks – depending on the research focus – ranging from rather artificial workflows to complex sensemaking activities. However, the task itself, besides the apparatus, is typically considered as

one of the limitations of a study. We tried to loosen this limitation by including the task as an independent variable: While the abstract classification task allowed us to independently investigate the different interaction techniques and their effect on participants' performance and work behavior, the sensemaking task helped us to explore how trained techniques are adopted and how a large ecology of devices is utilized. With this holistic evaluation approach, we did not only combine different strategies but also made **a first step to extend the ideas from Complementary Interfaces towards Complementary Evaluation Approaches**.

We believe that the methodological combination of different types of tasks in a user study is an effective way to gain different perspectives on a research problem. Future research should further investigate the effects and influences of suitable task combinations, allowing complementary insights into participants' usage of cross-device approaches. This might eventually result in components of a universal testbed or evaluation framework that enables to compare techniques or systems and creates a frame of reference for future research.

5.5 Limitations and Future Work

One of the limitations of our study is the focus on tablet-sized devices. The two previous chapters have shown the influences of device sizes on interaction, similarly, mobile devices with a smaller size (e.g., mobile phones) or larger tablets could lead to differences in our identified working behaviors – smaller devices could favor distributing activities due to spatial limitations; sorting activities could potentially benefit from larger devices. Also, the desk environment limits the generalizability of our results: Participants were free to choose spatial configurations of devices that suited their needs, yet they were still bound to the size of the desk, the seating arrangement, and the desk itself as a surface to arrange devices. Interesting work behavior could evolve when participants are also free to move around the desk or can arrange devices vertically (as seen on murder boards on TV), including handheld devices and additional pens for annotations, especially for sensemaking activities. Furthermore, allowing participants to work with such a large ecology of devices not only in a controlled lab environment but also in their familiar surroundings (i.e., in the wild) might lead to additional, complementary quality features. Although we integrated two task types in our user study that allowed us to see the bigger picture, different tasks types or task duration might lead to different results: The utilization of devices might differ for a) tasks with dedicated or pre-defined devices for various activities (e.g., creative tasks), b) the quantity of information (e.g., visual analytics), and c) longer task duration might extend behaviors (e.g., different purposes for devices).

5.6 Chapter Conclusion

We built a Complementary Interface to study the interplay of interaction techniques, device utilization, and task-specific work behavior. We integrated a methodological approach of using the task as an additional independent variable that allowed us to control the interdependence of components in the first task and study participants' assignment e.g., of device purposes in the second task. We conducted an experiment with different but complementary angles of evaluation with an abstract classification task [240], a sensemaking task [374], and three different interaction techniques. We studied 24 participants working with our multi-tablet application using a large ecology of devices of eight tablets (9.7″) in total. We evaluated participants' performance (RQ3.1), their work behavior (RQ3.2), and the interrelations of interaction techniques, device utilization, and task-specific activities (RQ3.3). We found that different interaction techniques have a surprisingly lower influence than expected, that work behaviors and device utilization heavily depend on the task at hand, and that participants value certain quality features of Complementary Interfaces. Our complementary evaluation approach with different task types and measurements allowed us to gain different yet complementary insights into otherwise hidden subtleties of Complementary Interfaces.

Part III

Measuring Cognitive Workload

Parts of the following Chapter 6 have been published as:

Thomas Kosch, Jakob Karolus*, **Johannes Zagermann***, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. ACM Computing Surveys 55, 13s, Article 283 (December 2023), 39 pages. DOI: https://doi.org/10.1145/3582272 *contributed equally to this research.*

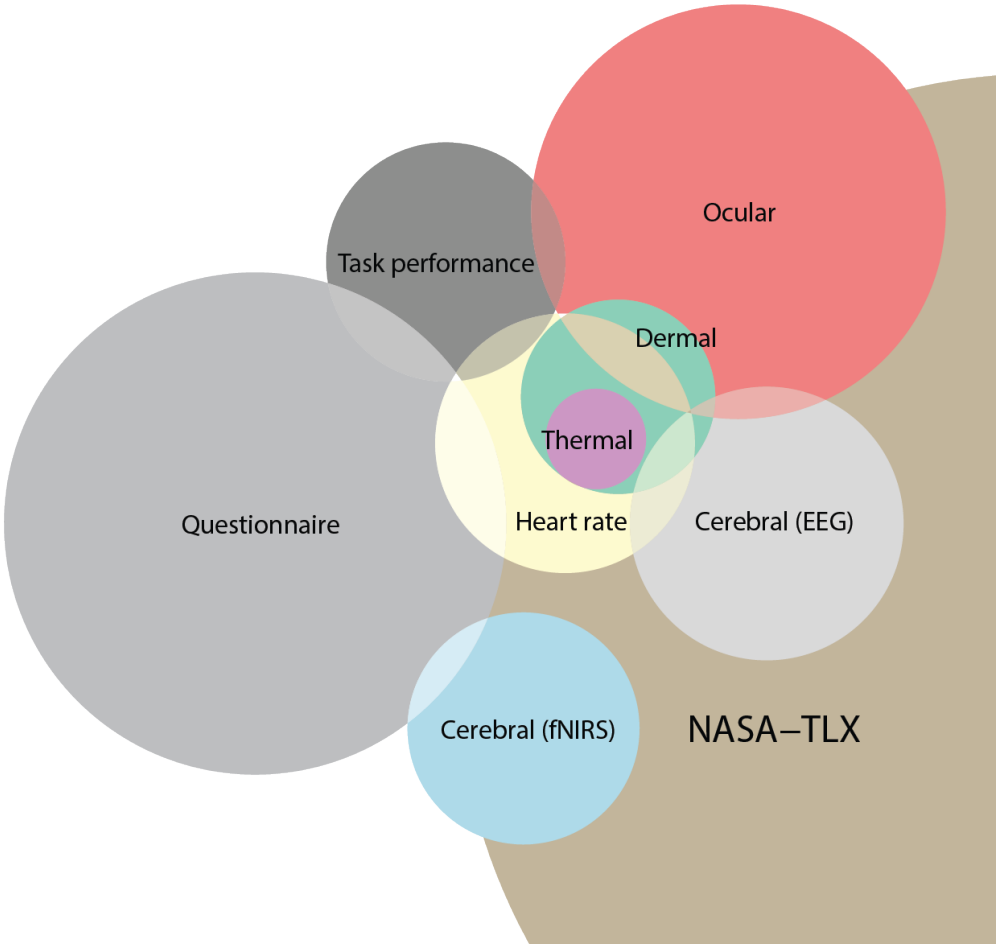
The responsibilities of this joint publication were divided as follows: Thomas Kosch, Jakob Karolus, and I equally contributed to the classification of papers and writing of the paper. Harald Reiterer and Albrecht Schmidt provided structural and procedural guidance. Paweł W. Woźniak contributed to the classification of papers, the writing of the paper, and supervised the project.

Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

Paper Library



The previous chapters presented use cases of Complementary Interfaces, each with an individual experimental design. Chapters 4 and 5 purposefully included the measurement of subjective task load using the NASA TLX questionnaire [135, 136] – showing the influences of handling multiple interactive components simultaneously on participants' task load. Similarly, prior research has identified both, negative and positive influences on cognitive workload of handling multiple interactive components [325, 439]. However, using a questionnaire to measure highly dynamic constructs such as cognitive workload might not be the best choice. Therefore, this chapter takes a step back from the focus on Complementary Interfaces and presents a literature review of measurements of cognitive workload in HCI in general – a potential dependent variable in experimental settings for Complementary Interfaces.

The chapter will address the following general research question:

RQ4: *How to measure cognitive workload?*

6

Measuring Cognitive Workload

Contents

Recent research has shown the influences of handling multiple interactive components on cognitive workload [325, 439] – similarly, the previously presented Complementary Interfaces have shown that different meaningful combinations of input and output modalities individually influence participants' load. In general, the increasing number and complexity of novel computing systems have an impact on cognitive workload during interaction. Therefore, cognitive workload analysis for the design (e.g., as an input modality) and evaluation (i.e., as a dependent variable of experimental settings) of interactive systems is an ongoing research area. While researchers use the term "cognitive workload" extensively, there is, to date, no consensus on a definition of the concept, nor is there a gold standard for measuring it. At the same time, the assessment of cognitive workload remains a primary objective in research. Cognitive workload is viewed as a factor that must be reduced or kept at an engaging level to achieve a satisfying user experience. By considering cognitive workload aspects, the research community strives to build improved user interfaces. Yet, the detailed history of how cognitive workload was adapted to HCI from the field of psychology [133, 134, 234] via HCI's Human Factors legacy remains uncharted. The HCI community used cognitive workload measures as early as 1994 [45], but the concept's generic applicability to computer systems remained a largely unexplored topic.

Parallel notions of the concepts of cognitive workload emerged in literature, resulting in an epistemic problem: On the one hand, cognitive workload is widely recognized as an important metric. On the other hand, no universal measures or definitions are recognized by the community. Despite this, research and practice use cognitive workload extensively to benchmark user interfaces where different human cognitive systems, such as memory, motor, visual, and auditory perception, are used. Traditionally, research employed questionnaires [135, 136] or think-aloud protocols [306] as cognitive workload measures. The NASA-TLX [135, 136] questionnaire has become a household name in HCI research, and it is widely taught in HCI education. With the advent of Ubiquitous Computing, sensing technologies enabled building systems that use physiological sensing data and user behavior as an objective indicator [55] of a user's mental workload. Sensing technologies provided researchers with a potent tool to objectively quantify the evoked workload of their user interface design in real-time and, going beyond, allow cognitive workload to be treated as a real-time input for user interface adaptations [103, 427]. This enables researchers to design for specific mental states of their users, providing a more fine-grained view on workload analysis or adapting interfaces to individual cognitive workload states.

Although well-researched literature on the evaluation of human work regarding cognitive and ergonomic factors exist [416], the missing consensus within the field still leads to misuse of metrics and, ultimately, misunderstanding the concept of cognitive workload. Misinterpretation of research findings, lack of reproducibility, and unsuitable measurement modalities are possible threats if cognitive workload is not fully understood. Furthermore, researchers and practitioners are challenged by an abundance of cognitive workload assessments, potentially hindering a proper evaluation of computing systems. Consequently, there is a need to categorize cognitive workload metrics to support researchers and practitioners in choosing evaluation metrics and allow a consistent interpretation of their findings.

This chapter presents a systematic review of the concept and measurements of cognitive workload in HCI research. The focus is intentionally broadened to the general field of HCI research to understand the bigger picture. Yet, cognitive workload is potentially a highly relevant dependent variable in experimental settings for Complementary Interfaces and a possible input modality for them.

Through charting the various metrics used for cognitive workload, we identify opportunities and challenges for the future use of cognitive workload metrics in Human-Computer Interaction (HCI). Further, we present a categorization providing insights into choosing appropriate cognitive workload metrics. Our work focuses explicitly on HCI's implicit understanding of cognitive workload rather than drawing from a theoretical background, as suggested by past work [145]. Consequently, we seek to chart

how HCI understands and measures cognitive workload. We contribute results that show that the field has primarily measured cognitive workload using post-hoc questionnaires, most predominantly the NASA-TLX. Later work used physiological sensing to use workload as an input or conduct in-depth analyses of how workload varies during a task. Finally, our review shows opportunities for educating and guiding researchers to understand the role of cognitive workload, allowing them to conduct focused and reproducible research through an improved understanding of cognitive workload metrics.

6.1 Review Synopsis

The review provides an overview of the theoretical background, metrics, and uses for cognitive workload. In contrast to previous research reviewing work on cognitive workload [18, 84, 88, 89, 206, 297, 436], our survey focuses on cognitive workload assessments in the field of HCI to support researchers in choosing cognitive workload measurements for their experiments. We summarize metrics used to assess cognitive workload for interface optimizations and adaptations. In the remainder of this chapter, we contextualize our inquiry to introduce cognitive workload theory first. Then, we report on our review methodology and describe frequently used cognitive workload metrics. After that, we categorize cognitive workload metrics to support the selection of appropriate metrics in future experiments. This includes a step-by-step procedure to select appropriate cognitive workload measurements (see Figure 6.4) and an interactive paper library that can be extended by future research contributions from the community[1], hence providing researchers and practitioners with an up-to-date tool for an efficient selection of suitable cognitive workload metrics beyond this chapter. Finally, we present three research gaps to advance the field and discuss opportunities and possible pitfalls for future research that uses cognitive workload e.g., as a dependent variable of an experimental setting or as an input modality.

6.2 Background

Previous research has presented different theories or models originating from the field of educational [378, 381] and cognitive psychology [133, 134, 411–413]. A majority of these concepts originate from the field of Human Factors, where HCI was adapting the concepts for their workload assessments. Although the two fields, Human Factors and HCI, have deviated during the past years [124], several cognitive workload assessments from Human Factors were adopted for HCI research. In this context, the HCI community uses the general terms “cognitive workload”, “cognitive load”, or “mental workload” in most of their experiments to describe the cognitive

[1] https://www.thomaskosch.com/cl-paper-library/—last access – last accessed 01.11.2023

demand of a task. Although these theories lay the foundations for assessing cognitive workload in Human Factors, cognitive psychology, and pedagogy, they were not designed for the HCI field per se. Subsequently, mental workload assessments were used to evaluate the cognitive demand of interacting with user interfaces, although they were not originally designed for the design and evaluation of user interfaces. For example, the NASA-TLX questionnaire [136], a method to assess pilot workload, was used as a mental workload measure in 1994 [45] in an HCI context for the first time.

The term workload characterizes the required mental effort to complete a task (*work*) with several constraints (*load*), such as fixed periods or large batches of tasks [267]. Extensive cognitive workload influences an individual's task performance and ability to cope with concurrent tasks. Whereas cognitive workload is often associated with a negative experience, it can be essential to elicit positive effects during daily activities. Nevertheless, a permanent high cognitive workload harms attention and focus when there is an imbalance between the individuals' cognitive abilities and the task difficulty, potentially distracting the person and eliciting mistakes. Cognitive workload occurs when events demand the mental resources of individuals in a specific context. These events can occur within an external context, such as interruptions caused by notifications, or an internal context, where the demand is inherent to the task. Multiple events that simultaneously impact a single processing channel impact cognitive processing performance [411–413]. How well a person copes with these events depends highly on their cognitive prowess and the type of events. For example, users interrupted by notifications revert faster to their previous mental state when multiple disruptions during long-term tasks demand higher cognitive abilities [231]. During times of high cognitive workload, the body allocates resources needed for dealing with the load. This includes changes in cortical activity, blood flow, an increase in respiration rate and electrodermal activity, or differences in eye behavior [79].

6.2.1 Understanding Cognitive Workload

Before we explore different workload theories and how they are employed in HCI research, we will clarify the understanding and term usage between them. Specifically, we will clarify the terms "cognitive load" and "mental workload". The term "cognitive load" was coined by Sweller [378] to describe the instructional design of pedagogical methods. Later, the three terms *intrinsic load*, *germane load*, and *extraneous load* were introduced to separate cognitive load into different modalities [381]. Although these concepts do not originate from the HCI field, we decided to include them in this survey since an increasing number of HCI papers started to mention the cognitive load theory to justify their workload assessment (cf. [6, 145]).

More recent work investigated the additive interaction between the two different concepts "cognitive load" (i.e., *intrinsic load*, *germane load*, and *extraneous load*) and "mental workload" [107]. Their results revealed an impact between task difficulty and time pressure, modulating the alertness, subjective, and psychophysiological workload measures of users. Thus, the definition and interpretation of cognitive workload, including its underlying theories and models, is a subject of controversy in psychology [80]. In the following, we describe concepts that theorize cognitive workload from an HCI perspective. These initial concepts led to a categorization of workload modalities later adopted for studies in the field of HCI. We begin by exploring theories, measurements, and assessments adopted in past HCI studies.

Since it originated in the 1980s, Sweller's cognitive load theory has become acknowledged within the field of instructions and learning design [88]. Sweller developed the cognitive load theory [378–380], a theoretical framework that explains how learning is limited through instructional design. The theory divides the relationship between instructional design and problem-solving into three components: *intrinsic load*, *germane load*, and *extraneous load* [381]. Initially, Sweller's cognitive load theory originated from educational psychology and was unrelated to how interfaces impact users' workload in HCI research. However, it provides an initial understandable separation between the different allocated mental resource fragments. Sweller assumes that each component cumulatively allocates and demands mental resources. *Intrinsic load* describes the inherent complexity of a task. Thus, the efficacy of a person to solve a given task heavily depends on their skill and ability to understand patterns or sequences. *Intrinsic load* is not trivial to manipulate since the intrinsic task complexity is associated with a person's necessary individual cognitive resources and proficiency. Keeping *intrinsic load* at a *sweet spot*[2] is considered to foster task engagement while reducing frustration [434]. *Germane load* represents the effort to process patterns within a task. Realizing new schemes or patterns that help solve a task may increase task engagement and foster learning. Thus, the maximization of *germane load* is emphasized as a crucial factor when designing engaging user interfaces. However, this is highly disputed among psychologists, where *intrinsic load* and *germane load* are considered to be identical [185, 357]. *Extraneous load* is manipulated by the task representation that can be perceived via human perception. For example, in the HCI context, a well-designed interface visualization allows for an easier interpretation of data. Relative to the other workload types, *extraneous load* is theoretically trivial to manipulate since the task representation can be exchanged through design alternatives. *Extraneous load* can be minimized to avoid unnecessary allocations of cognitive resources to resolve the intrinsic task complexity. Hence, *extraneous load* is directly

[2] A *sweet spot* is the optimal proximal zone, where a user is not affected by frustration or boredom that originates from cognitive over or underload.

related to the interface's representation and user interpretation. Work in HCI exists using *extraneous load* as the relationship between the visual representation of a task and measured workload [6, 145].

In contrast, "mental workload" is related to the cognitive demand of a task itself [267]. In contrast to the aforementioned cognitive load theory, mental workload refers directly to cognitive resources allocated by a task. The attention and task performance are more severely impacted the more mental workload is required. In this context, Wickens suggested a multiple resource theory, proposing that human operators do not have one information processing channel [411, 412], but several processing channels that can be addressed simultaneously. If one information processing channel is overly demanded, for example, through the sequential processing of a task, the more the performance is expected to decrease. Thus, users need more information processing resources, especially when performing multiple tasks on the same information channel, leading to errors and lowered task performance. However, simultaneously addressing different information channels, such as auditory or visual, does not necessarily cause mental overload.

The concepts of Sweller's cognitive load theory and Wickens' multiple resource theory are different, yet, they are used by the HCI community in a mixed way (cf. [6, 145]). The aim of this work is not to unify these theories. Instead, we aim to show how cognitive workload is measured by the community and provide an instructional guide to choosing suitable workload measures. This chapter systematically collects and summarizes workload assessment to guide researchers in selecting appropriate workload measurements. To avoid confusion regarding the terminology, we use the terms "cognitive workload", "cognitive load", and "mental workload" synonymously to describe workload imposed through the instructional system design of user interface visualizations (e.g., *extraneous load*) or cognitive demand of users who process information.

6.2.2 Cognitive Workload: Designing and Evaluating User Interfaces

Well-designed user interfaces are meant to support their users in their operation, preventing permanent high and low workload while fostering appropriate cognitive engagement. Therefore, cognitive workload has emerged as a frequently used metric for designing and evaluating interactive systems. The assessment of cognitive workload helps interface designers in two ways: Using a constructive or formative approach (e.g., during early design phases), where cognitive workload is measured to improve the design of novel interactive systems, or using a summative approach (e.g., using existing systems) that includes *cognitive workload as a dependent variable of an experimental setting* to study e.g., interaction techniques. Additionally, cognitive workload can be used as an input to design engaging adaptive interfaces that,

for example, adapt visualizations in real time. This concept is often referred to as *Cognition-Aware Computing* [51] and *Workload-Aware Computing* [208].

While most users can cope with changes in cognitive workload for a short time, the long-term management of high cognitive workload can be challenging. Over- or under-challenged user states are the result, leading to frustration or boredom [334]. In addition, health issues, such as burnout, mental exhaustion, or physical tiredness, may arise under prolonged exposure to cognitive workload [265]. To further elaborate on the role of cognitive workload in research, we provide an example of a user interface inducing high workload and then present how to use cognitive workload assessments to improve this interface: For example, interfaces that require users to search for items or memorize lengthy menu structures can increase cognitive workload by straining their memory [34]. Interacting with a lengthy menu list with many nestings for a short time might only impact cognitive resources. However, fewer cognitive resources are available to solve the actual task if users are repeatedly forced to navigate the same complex menu structure at varying times. Here, the visual complexity of the representation of the menu structure might strain the user's cognitive resources. The visual complexity can be reduced by (1) providing, for example, shortcuts for repetitive actions directly on the interface, (2) integrating context information (e.g., , only showing relevant items), or (3) sensing the cognitive workload through physiological sensing to predict following actions of a user. A typical user study would compare different user interface alternatives (e.g., different visual representations of a menu) and measure the user's cognitive workload using questionnaires (e.g., NASA-TLX [136]) or physiological sensing. These assessments are tools for researchers to design interfaces matching the cognitive resources of their users through the visual complexity of the representation of the user interface, hence manipulating extraneous load. Over the past 30 years, several metrics for cognitive workload have been explored by HCI research. Physiological sensing has gained attention as a measure of cognitive workload while, at the same time, new questionnaires have emerged to measure cognitive workload in various contexts. Hence, it becomes increasingly more challenging to maintain an overview of the most appropriate cognitive workload measurement modalities.

In this review, we investigate how this definition was applied in the design and evaluation of interactive systems. To this end, we summarize how extraneous load was evaluated in previous user studies and encapsulate the measures used to foster a shared understanding of cognitive workload assessments. This review focuses on the field of HCI in a computing sense, i.e., the interdisciplinary research domain and community historically built around the ACM Conference on Human Factors in Computing Systems (CHI) [124], usually classified by ACM as Human-Centered Comput-

ing. Thus, our inquiry addresses conceptualizations of cognitive workload within studies of discretionary use of computers [124]. We recognize that our analysis focuses solely on this academic tradition and that other communities, e.g., Human Factors Engineering (HFE) or Ergonomics, have built a more thorough understanding of workload (i.e., in the domain of human-machine interaction). We aim to show current practices, execution, and interpretation of cognitive workload assessments within the HCI community.

6.3 Review Methodology

Our review follows an iterative process of identifying relevant venues, filtering, and analyzing publications. This section describes our approach to the systematic review. We outline the inclusion criteria of our literature search and the phases of the refinement process (exclusion criteria). We followed the PRISMA procedures [199, 269] for systematic reviews and meta-analyses. The complete PRISMA is depicted in Section 6.3.3. We built our literature corpus by systematically identifying relevant papers through a keyword search in all proceedings and journals related to HCI.

6.3.1 Systematic Search

The first phase of our review was a systematic keyword search in the digital library databases relevant to the HCI field.

Inclusion Criteria for Venues

We included all proceedings and journals in the research area of HCI in our search. To get an objective overview, we used Microsoft Academic[3] to obtain conference and journal rankings in HCI. For conferences, we decided to include the top 20 venues plus six conferences in the top 40, all known as featuring highly relevant papers for this review. For journals, we included all top 10 journals. We additionally confirmed the relevance of the venues via CORE[4] rankings. We accumulated all matching papers within all the digital libraries, which contained at least one of the conferences or journals listed[5]. We adjusted the search syntax to work with the search APIs provided by the publishers. This process resulted in an initial corpus of 6414 papers. The cut-off date for all venues was the 30th of April 2020.

Inclusion Criteria for Keywords

We introduced an interpretation of the term “cognitive workload” and its implication for interaction in Section 6.2. We discovered that the term is used with different

[3]Service discontinued, previously available via https://www.academic.microsoft.com
[4]https://www.core.edu.au - last access – last accessed 01.11.2023
[5]Publishers changed for some venues.

semantics during this survey. To this end, we iteratively refined our keywords and evaluated them on a base corpus of top-rated venues. Starting from the initial keyword "cognitive workload", we identified synonyms used in literature and extended our set of keywords. The regular expression used in the review is presented in Equation (6.1). This expression matches all commonly used terms and synonyms related to cognitive workload. Using the expression in the relevant databases resulted in our initial set of papers:

$$((cognitive \mid mental)\,(load \mid workload)) \mid (working\ memory) \tag{6.1}$$

6.3.2 Review Process

Our systematic search yielded an initial corpus of 6414 papers in total; 4408 from conference proceedings and 2006 from journals. This section outlines how we filtered the papers to reach our final corpus. We employed two exclusion criteria and a final categorization of our paper selection as a meta-analysis step.

Exclusion Criteria

As detailed below, we reviewed the initial corpus concerning paper eligibility and relevance.

Paper Eligibility We only considered archival, peer-reviewed papers, i.e., journal and conference papers, full and short papers, and book chapters. This approach eliminates workshop summaries, posters, works in progress, and other non-archival publications. We excluded 904 papers in this step.

Relevance We used a double decision scheme to filter the initial corpus of eligible papers (5510 records). Four researchers participated in the filtering process. Per publication, two researchers independently assessed if the publication was relevant to the survey. Using researcher triangulation, we ensured that at least two researchers had read every paper in the final corpus. No author had assessed the relevance of their own work. Papers were considered relevant if they exhibited one of the following characteristics:

- **Metrics paper:** These papers discuss a system or application, including a form of cognitive workload that was measured, either using it as an evaluation metric or as an input channel.
- **Meta paper:** These papers discuss a concept of cognitive workload on a meta-level, such as definitions, application concepts, and possible frameworks for HCI research.

Each publication that was rated as relevant by two researchers was included in the final corpus. In addition, we excluded another 4931 papers during this step. This final corpus (579 records) was then subject to our analysis, extracting employed modalities in the categorization step.

Categorization

We aimed to understand the content structure within the corpus through this categorization, especially the employed modalities to measure cognitive workload. Thus, we additionally categorized metrics papers into works that used subjective metrics and those that used objective metrics. Finally, we directly filtered by employed modalities, such as questionnaires or gaze, in a more refined categorization. During this step, we did not exclude any publications.

To systematically code all papers and build a taxonomy of all cognitive workload measures used in HCI, we used a qualitative analysis approach where we listed all cognitive load measures used in all papers in the corpus. In a coding meeting, four researchers then agreed on a naming scheme for the methods to achieve a uniform matrix of measures in papers. We then used affinity diagramming to iteratively group papers and methods to reach a final taxonomy of methods. Applying this coding method does not guarantee that the proposed taxonomy is the only correct one (no qualitative analysis can offer such a result), yet our approach ensures that the structure is comprehensive.

6.3.3 PRISMA Flow Graph

In the following, we present the complete PRISMA flow graph for our systematic review, as detailed in this section. It depicts the individual stages of paper inclusion and exclusion (right-hand side).

6.3.4 Literature Corpus

Our final corpus includes a total of 579 records: 482 records from 26 conferences[6] and 97 records from 11 journals. The complete list is available on our website[7]. Among the irrelevant papers, most were excluded due to the ambiguity of the term "working memory". Especially in computer science venues, this is a technical term used in processor architecture. Moreover, papers often used cognitive workload to motivate research or explain results. Here, cognitive workload is merely a hypothetical factor contributing to results. These works did not actively engage with metrics or concepts related to cognitive workload and were subsequently excluded.

[6] Including conferences that switched to a journal format.
[7] https://www.thomaskosch.com/cl-paper-library/ - last access – last accessed 01.11.2023

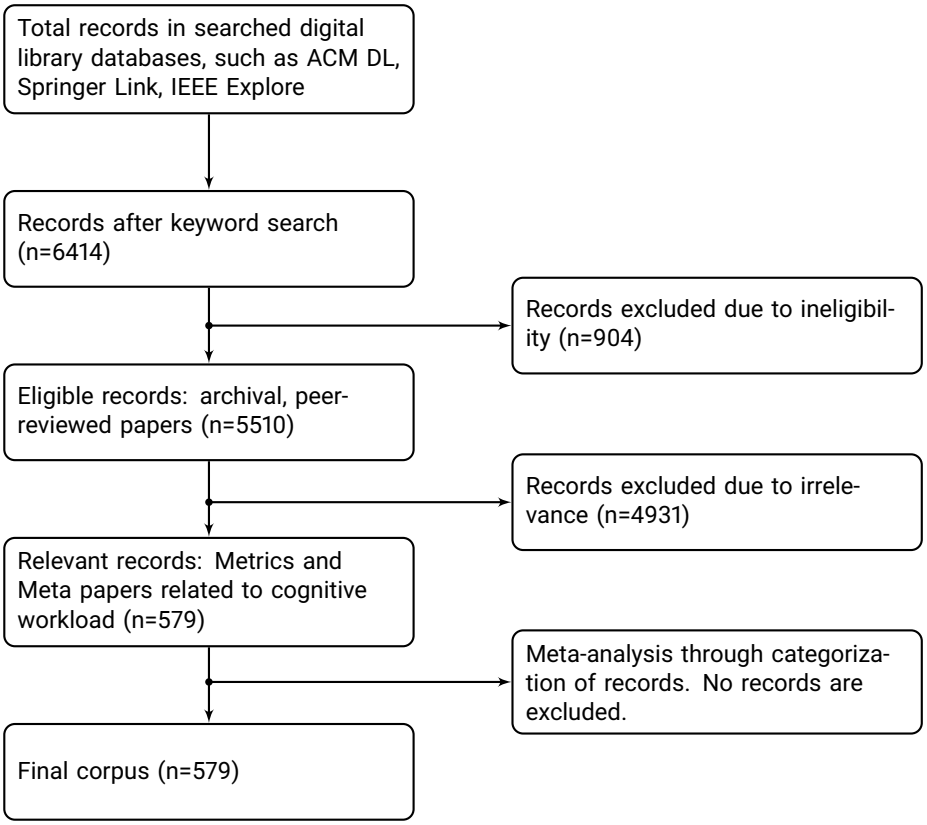
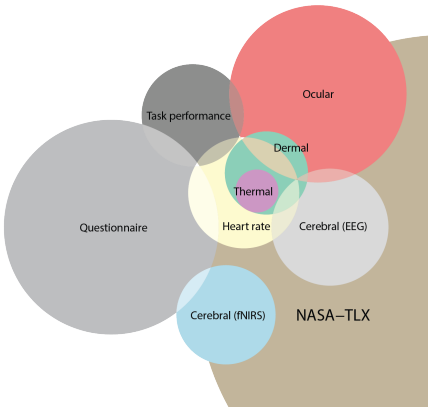


Figure 6.1: PRISMA flow chart highlighting the stages of our systematic search as detailed in Sections 6.3.1 and 6.3.2.

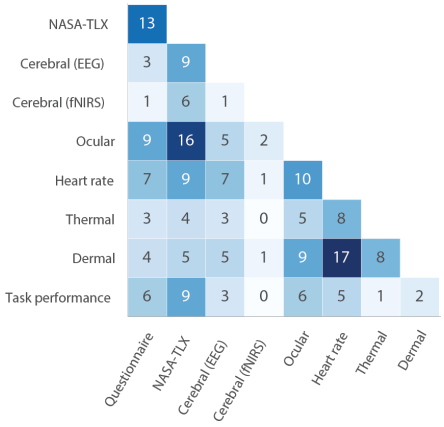
6.4 Cognitive Workload Metrics

Informed by the records of our literature corpus, we describe current practices for cognitive workload measurements in HCI. This section aims to provide a comprehensive overview of individual metrics, their respective characteristics, and their relation to each other to inform researchers about the functionality of the measurement. Please refer to Section 6.5 on how these metrics are employed in HCI studies, including a step-by-step guide on choosing an appropriate metric. Additionally, our website provides an interactive search-and-filter interface supporting researchers and practitioners to find relevant literature – and with this suitable cognitive workload measurements – for their own work. We present the whole spectrum from subjective (e.g., questionnaires) and objective (e.g., physiological) measurements. If not stated differently, details on the analyses of the measures can be found in the referenced literature. As an overview, we provide Figure 6.2 and Figure 6.3 depicting numbers for selected modalities[8] and the relationships among them.

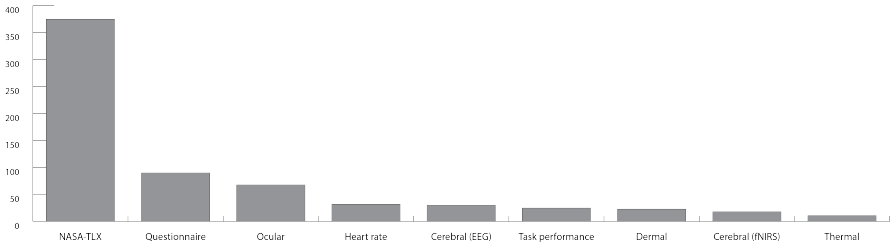
[8]We decided to omit less frequent modalities for visual clarity.



(a) An approximate graphical representation of which measurement modalities were used simultaneously in the corpus. Note how physiological measures were often used together.



(b) The table contains counts of all papers which concurrently used two or more measurement modalities.

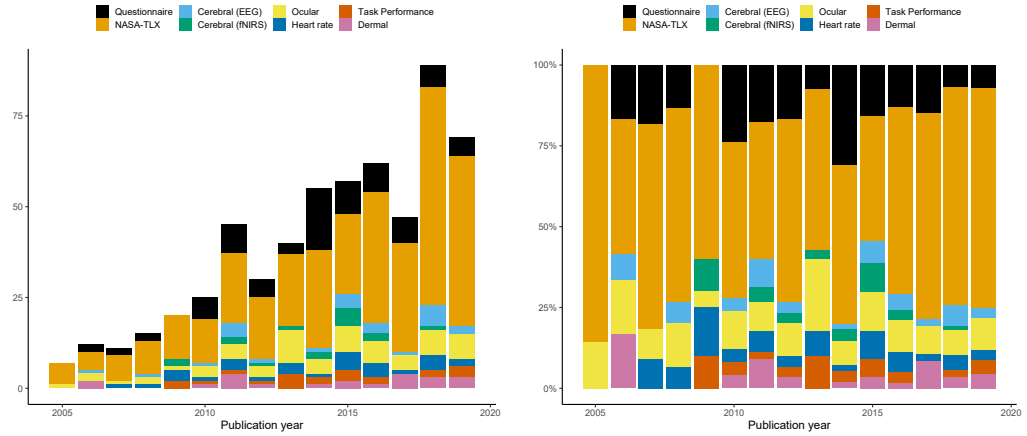


(c) The graph visualizes the total number of times each modality was employed in the corpus. Questionnaire-based methods were most prevalent, with a big emphasis on using the NASA-TLX.

Figure 6.2: An overview of the different modalities used to estimate cognitive load in the corpus.

6.4.1 Questionnaires

Questionnaires incorporate questions to infer the experienced cognitive workload of a single trial or whole experiment. This is often indicated by a score that is assigned by participants. The magnitude of the score (i.e., lowest and highest score) depends on the design of the questionnaire itself. Questionnaires have two key advantages, which make them popular: First, they can be easily deployed by asking a participant to fill in questionnaires after an experimental condition. This removes the inconvenience of rather complex workload measures that require a prior rigorous setup. Secondly, they provide a predefined and straightforward procedure to analyze the workload score. The cognitive demand of the presented stimuli is quickly quantified through automatic or manual analysis of workload scores. However, cognitive workload measures based on questionnaires are prone to biases. Questionnaires lack real-time capabili-



(a) Absolute number of papers given publication year for selected modalities.

(b) Individual share of selected modalities normalized per publication year.

Figure 6.3: An overview of the eight most prominent modalities broken down by publication year. Note how the overall number of modalities increases over time, yet the distribution of individual shares stays rather constant over time. The metrics are highlighted in bold font in Table 6.1 (excluding questionnaires).

ties since they are employed *after* an experimental condition. Hence, these measures aggregate their workload measure of an experimental condition into a single score, making it challenging to measure fluctuations of cognitive workload *during* the experimental condition or user interface interaction. Furthermore, questionnaires are highly susceptible to subjective perception: Users assess questionnaires depending on their perception, so it becomes difficult for the experimenter to interpret the results if high variations of workload scores exist within the same experimental condition. Despite the disadvantages above, questionnaires resemble the most used metric for cognitive workload among HCI researchers and practitioners (see Figure 6.2). We describe the design of the questionnaires that emerged as relevant throughout the literature review in the following subsections.

NASA Task Load Index

The NASA Task Load Index (NASA-TLX) [135, 136] is a multi-scale questionnaire to assess participants' perceived subjective workload. Its origins date back to the 1980s when NASA researchers introduced it to study task load of pilots. Although it was not originally designed to evaluate cognitive workload based on the interaction with a computer, it is the most frequently used questionnaire to measure cognitive workload in our literature corpus (cf. Figures 6.2 and 6.3). For the questionnaire, the overall workload is split into six sub-scales: 1) Mental Demand, 2) Physical Demand, 3) Temporal Demand, 4) Performance, 5) Effort, and 6) Frustration. Each sub-scale

has an additional description that participants need to read before rating the specific sub-scale to minimize misunderstandings (e.g., "How mentally demanding was the task?" for the *Mental Demand* sub-scale). Sub-scales are horizontal lines divided into 20 intervals with bipolar descriptions on the left and right (i.e., Low/High). Tick marks with increments of 5 allow participants to have ratings between 0 and 100. Researchers can then analyze the overall perceived workload in two ways: a) with an individual sub-scale weighting, or b) without an individual sub-scale weighting, often referred to as "Raw NASA-TLX".

Dundee Stress State Questionnaire

The Dundee Stress State Questionnaire (DSSQ) is a questionnaire measuring task-induced stress [261, 262]. Past research used this questionnaire to measure cognitive workload, assuming that stressful situations are associated with negative emotions that lead to higher cognitive demand. The DSSQ employs eleven state factors, which are associated with the three dimensions of *Task Engagement* (factors: Energetic Arousal, Tense Arousal, Hedonic Tone), *Distress* (Intrinsic Motivation), and *Worry* (Self-focus of Attention, Self-Esteem, Confidence, and Control, Concentration, Task-related Cognitive Interference, Task-irrelevant Cognitive Interference) of which each scale consists of eight items.

Instantaneous Self-Assessment

The Instantaneous Self-Assessment (ISA) [388] is a uni-dimensional subjective workload assessment technique. Participants rate their perceived workload in predefined intervals during the execution of a task – allowing for real-time assessment of cognitive workload. Participants are continually asked to self-rate their perceived workload during a task on a scale of 1 ('underutilized') to 5 ('excessive'). While this can be done using pen and paper, traditionally, an additional keypad consisting of 5 buttons (each button representing a workload level) and a flashlight (flashing when a rating is requested) is required.

Bedford Workload Scale

The Bedford Workload Scale [337] is a uni-dimensional rating scale to determine free mental capacity. It consists of a hierarchical decision tree, and participants are asked whether 1) it was possible to complete the task, 2) the workload was tolerable, and 3) the workload was satisfactory without reduction. Depending on the decision, participants then have to rank their perceived workload on a rating scale with end-points from *workload insignificant* (1) to *task abandoned* (10). This value represents the perceived workload of a participant.

Workload Profile

The "Workload Profile" questionnaire [393] combines a *Psychophysical Scale*, the *Bedford Workload Scale*, and a *Workload Profile* to measure cognitive workload. The psychophysical scaling procedure assigns an arbitrary workload score to a reference task. After completing the reference task, the participants are asked to rate the cognitive demand of the other conditions relative to the reference task. There are no numeric restrictions in the workload rating. Hence, each rating is subjectively proportional regarding the difficulty assigned to the reference task. Conceptually similar to the Bedford Workload Scale, the *Workload Profile* collects the proportion of remaining attentional resources after the subject has experienced all tasks. Eight workload dimensions are rated, summed up, and compared to the other two scales.

System Usability Scale

The System Usability Scale (SUS) [47] is a ten-item questionnaire for the subjective assessment of usability. Each item is represented as a statement (e.g., "I found the system unnecessarily complex"), and participants use five-point Likert scales to indicate the extent to which they agree with the specific statement (i.e., strongly disagree (1) to strongly agree (5)). The SUS has a predefined procedure to calculate the overall SUS score between 0 and 100. Although the SUS represents a classical subjective assessment method for usability, previous research has argued that cognitive workload can be "calculated indirectly through some of the questions in the SUS" [336] (e.g., "I thought the system was easy to use").

Rating Scale Mental Effort

The Rating Scale Mental Effort (RSME) [438] is a single-scale questionnaire with nine workload levels ranging from "not at all hard to do" to "tremendously hard to do." The RSME scale consists of a 150-millimeter vertical line with tick marks from 0 to 150 on the left-hand side and nine workload levels as labels on the right-hand side. Participants rate their workload on this scale. It is noteworthy that the lowest workload classification ("Not at all hard to do") is not represented as value 0, and neither is the highest workload classification ("Tremendously hard to do") represented as value 150. This allows participants to rate their workload beyond the predefined classification.

Driving Activity Load Index

The Driving Activity Load Index (DALI) [306] is a multi-scale questionnaire to assess participants' perceived subjective workload. The DALI is an adapted version of the NASA-TLX, yet tailored to the driving task. Similar to the NASA-TLX, the DALI con-

sists of six scales: 1) Effort of Attention, 2) Visual Demand, 3) Auditory Demand, 4) Temporal Demand, 5) Interference, and 6) Situational Stress. Participants rate each sub-scale on a 100-point range with 5-point steps. In line with the NASA-TLX, an individual sub-scale weighting is needed to compose a global workload score.

Cognitive Task Load Model Questionnaire

The "Cognitive Task Load Model" resembles a theoretical model to analyze cognitive workload [281], which was developed explicitly for naval ship control centers. Cognitive workload is formalized as a metric to refine allocated mental resources, which a user interface can utilize to provide adaptive in-situ support. Users are observed and classified into an abstract model that monitors workload utilization rather than assigning scores to conditions. In contrast to the other workload-related questionnaires, the cognitive task load model does not result in a quantitative score but a classification into the framework.

Customized Questionnaires

In addition to the established questionnaires, previous research has incorporated various custom questionnaires to subjectively measure cognitive workload – tailored to their specific use cases or tasks. In addition, they often ask informally about the perceived mental effort (e.g., using Likert scales).

6.4.2 Cerebral System

The human brain is the information and control unit of a human being. Thereby, approximately 100 billion neurons underlie human cognition [139]. Electrical activity in several brain regions can be measured by placing electrodes on the scalp. Neurons communicate by exchanging electrical activity using neurotransmitters, a chemical substance transferred between neurons. Measuring cerebral activity has a rich history, in which scientists experimented with different modalities to capture cortical activity. In 1929, Hans Berger published his first report about cortical electrical measurements that can be captured by placing electrodes on the outer scalp of the brain [225]. Since then, Berger became recognized as a pioneer in the quantification of cerebral activity. Other modalities, such as Functional Near-Infrared Spectroscopy (fNIRS) that investigate the oxygen levels in the blood, or Functional Magnetic Resonance Imaging (fMRI) that measures the electromagnetic responses of the brain, were introduced in the 70s [99] and 90s [376] respectively to obtain more accurate brain imaging measurements at the cost of temporal resolution. We present the functionality of the brain measurements and their features in the following.

Electroencephalography (EEG)

EEG is commonly leveraged in clinical applications and yields a non-invasive method to estimate cortical activity [260, 284, 418]. Electrical potentials between $1\mu v$ and $100\mu v$ (microvolts) are measured by placing conductive electrodes on a scalp. An additional electrode serves as a reference electrode, which can be placed on the earlobe or scalp [255]. The measured EEG potentials allow the processing and extraction of different features. Machine learning has been used on these features to retrieve insights into cognitive processes [243, 244]. EEG has been used to discriminate between cognitive states [102, 120, 232], as a measurement for user experience evaluation [103], and as an input method [16, 389]. For example, changes in electrical potentials are observed by analyzing frequency bands. Previous work found a drop in frequencies of alpha (8 - 12 Hertz (Hz)) and an increase of theta (4 - 8 Hz) [200, 209, 211, 216] when subjects had to raise mental capacities. An alternative approach is the assessment of Event-Related Potentials (ERPs) to infer mental workload [48, 338]. To complement this, real-time brain visualizations enable deeper insights into sequences of neuronal activity [210].

However, EEG measurements are prone to noise. Head movements, muscle contractions, eye movements, or even eye blinks cause changes in the electrical field on the scalp. Researchers are concerned about this and invest significant effort to reduce the number of measurement artifacts [76, 105, 207, 287]. Since artifacts cannot be completely avoided, EEG often requires a controlled environment comprising minimal body movements by the subject. Such conditions are often impractical for end-users due to their high experimental control. However, recent technical advances ameliorate these disadvantages [273], which allows artifact corrections for EEG during mobile settings [40]. The barrier to using EEG in the real world has been lowered by making EEG headsets accessible to the consumer market. These are usually priced between $249 and $1600. More affordable open source solutions can be acquired within the OpenEEG Project[9].

Frequencies Different oscillations in EEG signals are attributed to several cognitive states. The six bandwidths delta (1 Hz - 3 Hz), theta (4 Hz - 7 Hz), alpha (8 Hz - 12 Hz), lower beta (13 Hz - 20 Hz), upper beta (21 Hz - 30 Hz), and gamma (31 Hz - 100 Hz) power are predominantly employed when analyzing frequencies. Delta power is associated with varying levels of attention, salience, detection, and subliminal perception [205]. According to our literature survey, delta power has not been widely used to measure cognitive workload of computing systems. In contrast, changes in alpha and theta frequencies are correlated with the mental demand placed

[9] https://openeeg.sourceforge.net - last access – last accessed 01.11.2023

on working memory and increase or decrease in task engagement [111]. Theta oscillations correlate with the frontal anterior cingulate cortex regarding task engagement the users are experiencing [346]. Increased theta power is associated with higher task engagement, while lower theta power indicates low task engagement. The alpha power is associated with changes in brain resting states. Modulating alpha waves is achieved if the brain is resting, decreasing when users perform tasks requiring their memory. Since it is unknown if correlations between the demand of working memory (i.e., alpha power) and task engagement (i.e., theta power) might elicit a flow state, researchers use the theta-alpha ratio as a metric for cognitive demand [346]. Similarly, beta power is widely used as a metric for mental workload, including alertness, arousal, frustration, engagement, and workload states [222]. In contrast, gamma power indicates high mental activity at the somatosensory cortex [222]. Overall, theta, alpha, and beta power are dominantly used for workload assessments.

Event-Related Potentials Event-Related Potentials (ERP) are specific amplitudes that characteristically appear after users have perceived a stimulus [247]. Different amplitudes may be measured depending on the type of provided stimulus. For example, if a stimulus does not match a user's expectation, a negative peak in the EEG signal is measured after 400ms. In contrast, the processing of stimuli is visible after 300ms in the EEG signal. The increase in amplitude is known as P300, a specific ERP component that is suspected to correlate with cognitive workload [48]. Therefore, smaller ERP amplitudes are expected when users perceive stimuli that diversely demand their working memory. Tasks that require a large amount of working memory are more difficult to process, thus resulting in a smaller P300 amplitude. This provides an indicator for working memory placed on users when potentially interacting with a user interface.

Functional Near-Infrared Spectroscopy (fNIRS)

Another non-invasive modality to measure brain activity is fNIRS [99]. fNIRS utilizes near-infrared light within a range of 650 nm and 1000 nm to measure changes in the concentration of Oxygenated Hemoglobin (HBO) and Deoxygenated Hemoglobin (HbR) in the human brain. The light emitters are placed on the human scalp and measure the outcoming light to affect the oxygen used. While fNIRS is famous for brain sensing in HCI research due to its easy setup [279], latencies exist between stimulus onset and the exact measure. This can make fNIRS challenging to be used with applications that require immediate feedback.

Functional Magnetic Resonance Imaging (fMRI)

fMRI is another non-invasive measure for cortical activity. In contrast to measures of electrical activity, fMRI uses magnetic fields to measure the level of blood oxygen in the brain [92]. fMRI uses electromagnetic waves instead of infrared light to find changes in blood oxygenation. Similar to fNIRS, it takes several seconds to measure changes in cortical activity after stimulus onset when using fMRI. While providing very accurate results, fMRI is a stationary and expensive measurement modality, making it rather unsuited to evaluate interactive systems.

6.4.3 Ocular System

The eyes are part of the human visual system. The eye collects light from the environment, focuses and tracks visual stimuli, and transports these signals to the human brain for further interpretation. In addition, the eye constantly moves – voluntarily or involuntarily – to adjust the light intensity by dilating, contracting the pupil, or focusing and following objects of interest. These eye movements can be recorded in real-time using eye-tracking devices. While stationary eye trackers provide high precision and are usually attached to a display, mobile eye trackers (e.g., as a standalone device or integrated into a head-mounted display) are worn like glasses, providing the possibility to move in an environment freely. Both eye-tracking devices allow for (e.g., gaze-based) interaction with a system or evaluation purposes. For example, websites or interfaces are potential use cases for evaluation with eye trackers. Here, areas of interest and focus sequences can be visualized as, e.g., heat maps or scan paths. However, previous research has shown that various eye movements can also indicate increased cognitive demand [86, 212, 213, 429, 431] or derive the user proficiency during interaction [236].

Pupil Diameter

Changes in pupil diameter are referred to as pupil dilation: This is an involuntary eye movement (i.e., a reflex), and the diameter can range from 1.5mm up to 8mm. Previous research has shown that pupils dilate with increasing task difficulty [66] across tasks like reading, problem-solving, or visual search activities [340], where the cognitive demand can be estimated in real-time using machine learning [212]. However, a significant influence on pupil diameter is the environment. In a darker environment, pupils dilate to acquire more light – pupils contract to reduce the amount of light in a brighter environment. Thus, different methods and models (e.g., [311]) were developed to ensure a valid measurement of cognitive workload using pupil dilation: Here, the total pupil dilation is seen as the sum of dilation due to light intensity and cognitive workload. Knowing the pupil diameter for a given light intensity allows

us to calculate the effect due to cognitive workload. The Index of Cognitive Activity (ICA) [258] and Index of Pupillary Activity (IPA) [86, 87] follow a similar principle to separate light-invoked and cognitive workload-related changes in the pupil size.

Saccades

The eye movement that allows for shifting between two fixations (i.e., areas of interest) is called saccades. Saccades are voluntary eye movements that take 30 to 80 ms to complete and are usually visualized as scan paths to, e.g., show a sequence of areas of interest. Previous research has shown that the size and speed of saccades are highly discriminatory parameters for an increased cognitive workload: Here, an increased cognitive workload is indicated by larger or faster saccades [66].

Eye Blinks

Eye blinks indicate the perceived workload of a user, whereas repeated involuntary eye blinks are a sign of cognitive fatigue. Previous research has shown that the rate and latency of blinks are related to cognitive workload: Lower blink rates and higher blink latencies indicate an increased cognitive workload [66].

Fixations

Fixations are voluntary eye movements that focus on an area of interest. Previous research has shown that the fixation rate and duration can indicate an increased cognitive workload. The fixation rate of an area of interest describes the number of times an object was looked at – this can be interpreted as a repeated interest in an area and relevant for the current cognitive activity [340]. The fixation duration was found to indicate an increased strain on cognitive workload. Previous research shows that the rate and duration of fixation are indicators of an attention shift due to increasing task complexity [66].

Smooth Pursuit and Ocular Movements

Smooth pursuit eye movements are necessary to follow a moving target. Smooth pursuits are voluntary and can be considered a conscious decision to track a target. Previous research investigated how these movements can be used as an input technique (e.g., target selection) and real-time cognitive workload measurement. Previous work has also shown that smooth pursuit eye movements deviate from a given trajectory in situations with a higher cognitive demand [213] while being less prone to changes in light intensity (cf. pupil diameter).

6.4.4 Cardiovascular System

Within our autonomic nervous systems (ANS), two[10] opposing systems – the sympathetic (activating) and the parasympathetic (inhibiting) nervous system [226] – control bodily functions, such as heart rate (HR). These changes are reflected in the cardiac cycle, and related metrics, e.g., heart rate variability (HRV) [347]. Among others, external influences such as stress, emotion, and work impact these metrics and can be used to infer those [254]. This section highlights cardiovascular responses evoked by the ANS. Other related responses, such as dermal activity or pupillary response, are covered in their respective sections.

Heart Rate

Techniques such as electrocardiography (ECG) and photoplethysmogram (PPG) are used to monitor a subject's heart rate and derived metrics in the time and frequency domains unobtrusively. Scenarios include, for example, office-work [70], elementary cognitive tasks [129] and automotive environments [399].

While heart rate-related metrics have been shown to indicate a user's cognitive workload, the abundance of external factors that additionally influence our ANS can make this modality ambiguous and warrant close control of confounding variables. Normalization techniques, such as establishing baselines for the sympathetic and parasympathetic nervous system responses [70], as well as relying on multiple modalities, may allow building more robust models [177].

Respiration

Research showed various links between respiratory activities and an increased cognitive workload: Here, respiration belts or thermistor-based flow sensors underneath the participant's nose were used to measure the respiration rate, its variance, and depth [177]. Increased cognitive workload can lead to higher and lower respiration rate variability [160].

Temperature

Here, recording facial and skin temperature via thermal imaging and temperature sensors has been investigated by previous research. A challenge is that temperature measurements significantly lag between stimulus onset and physiological response. Likewise, our temperature is influenced by the ANS. Research using thermal imaging found promising results in distinguishing cognitive workload levels using standardized cognitive workload tests [1]. However, in lower constraint scenarios, e.g., in

[10]Depending on literature, the enteric nervous system is also included.

a driving context, classification is more difficult [9]. Furthermore, using skin temperature sensors is often accompanied by other on-skin sensors (e.g., ECG, PPG). The availability of intelligent wearables makes these measurements readily available. Hence, models for cognitive workload estimation include multiple metrics derived from ANS responses. Here, temperature-related measures such as heat flux (heat transfer rate) are more robust [129, 356], most likely due to the less noise-prone method to measure them.

We classify temperature measurement as related to the cardiovascular systems as most works in our survey assumed the correlation between increased heart activity, skin temperature, arousal, and cognitive workload. This reflects assumptions and partial evidence present in other fields. In physiological computing, Ikehara and Crosby [162] deemed temperature relevant for cognitive workload measurement. In the Human Factors engineering field, temperature was linked to stress [421] and heart rate variability [106]. Thus, our review shows that HCI operates under the premise that skin temperature is correlated with cardiovascular processes induced by cognitive workload.

6.4.5 (Electro) Dermal Activity

Another physiological response linked to the autonomic nervous systems is electrodermal activity (EDA), covering all electrical phenomena in skin [43]. There are various recording techniques, such as applying currents to measure the conductance or resistance of the skin. A significant use case in engineering psychology includes detecting different levels of arousal and stress [43].

Measurements of EDA and derived metrics have been successfully applied, including, e.g., user experience evaluation tools [109], detecting office workload [351] and driver-related tasks [332, 356, 370]. Similar to metrics related to heart rate, EDA suffers from ambiguity. External factors, such as body position and emotional stress [43] impact accuracy. Reported results in research fluctuate in terms of the discriminative power of EDA. While some work showed feasible accuracy [289, 332], other work argues that other modalities are superior [129, 330].

6.4.6 Task Performance

Traditional usability measurements for efficiency or effectiveness, such as time and error, could increase cognitive workload. Here, various studies report on task completion time (e.g., [65, 85]) and reaction time (e.g., [68, 119]) or the number of errors (e.g., [15, 85]) and task accuracy (e.g., [65, 238]) – all highly related to changes in cognitive workload. Additionally, keystroke dynamics and linguistic

markers of typed text (i.e., language production) were highly related to cognitive workload [46, 401].

6.4.7 Haptic Interaction

Different modalities, such as input via pen or touch, proved to be valid discriminatory indicators of changes in cognitive workload. During handwriting with a stylus, the velocity [341, 425], frequency [343], trajectory duration [342], and pressure of strokes [425] were studied as indicators for cognitive workload. Similarly, touch input [57] and finger trajectories [268] were used to measure cognitive workload.

6.4.8 Speech

Think-aloud protocols allow researchers, e.g., to analyze reasons for participants' behavior, mood, or expectations. Additionally, previous research has shown that participants' spoken words during an evaluation can be used to analyze the cognitive workload. Speech features such as the flow of words (e.g., tempo [62] or pauses [193, 194]), but also more nuanced features such as the lexical density [195], the spectral domain [424] or the pitch contour [57] while talking were shown to be indicators of changes in participants' cognitive workload. Additionally, linguistic features were analyzed to study participants' cognitive workload [196, 403].

6.4.9 Body Movements

So far, previous research has investigated 1) postural behavior and gestures to assess phases of cognitive underload or overload [141], 2) mouse movements such as traveled distances [14] or pause/break activities [197], and 3) differences in glance duration at input modalities and output devices to solve a given task [85]. In addition, specialized measurement techniques, such as EMG, which records the electrical activity of muscles [187, 188, 266], can assess co-occurring physical demand to decompose cognitive load factors, such as the impact of motor memory [220].

6.4.10 Biomarkers

Biomarkers (or biological markers) are measurable indicators of a medical state [377]. More generally, biomarkers can be used to indicate the physiological state. These markers play a significant role in medicine, e.g., stress biomarkers in behavioral therapy. Typical markers include cortisol, alpha-amylase, and pro-inflammatory cytokines for different biological stress systems[11] [280]. However, despite their distinctiveness, biomarkers have not been widely researched within the area of HCI. One

[11] Hypothalamic-pituitary-adrenal (HPA) axis, autonomic nervous system (ANS) and immune systems, respectively.

primary reason is the intricate process of obtaining and analyzing biomarkers, often requiring lab work and physicians to draw samples. There is potential for accurate ground-truth labeling. A careful evaluation of the actual measurement is inevitable, as external stressors, e.g., the social-evaluative threat of performing well in a study, may impact biomarkers [70].

Salivary cortisol, usually used as a biological marker for stress [192], has recently been utilized for cognitive workload measurements [70]. Cortisol levels are measured from salivary probes that require laboratory analysis. While the study of salivary cortisol is a reliable indicator for stress and workload [198], the need for a laboratory analysis restricts the real-time capabilities of interactive systems. However, cortisol is a reliable measure of cognitive workload and represents a well-grounded measure in clinical research.

6.5 Using Evaluation Methods for Cognitive Workload

While our review demonstrates current challenges in understanding cognitive workload aspects in HCI, it also surfaced examples where the concept of cognitive workload was used effectively. In our review, we found 579 papers that used cognitive workload measurements. Most importantly, past work can guide if, how, and when to use cognitive workload measures. This section aims to provide a compiled step-by-step procedure (cf. Figure 6.4) on choosing suitable metrics, complemented through Table 6.1 categorizing the papers in the corpus in terms of measurement modalities and metrics used.[12] We suggest that researchers and practitioners deciding on the cognitive workload metrics for their studies first consult the procedure (cf. Figure 6.4) in Section 6.5.1 to get an overview of specific methods from related work in Table 6.1, and details on the particular applicability, advantages, and limitations of metrics in Section 6.4. The referenced papers in Table 6.1 and our interactive paper library can then be investigated to get more details on a selected metric.

[12] We do not list papers that exclusively use questionnaires to assess cognitive workload in their studies. For a complete list of papers, including those that use questionnaires, consult our interactive paper library: https://www.thomaskosch.com/cl-paper-library/ - last access – last accessed 01.11.2023.

Table 6.1: Papers in the review corpus classified in terms of the cognitive workload metrics used in the reported studies. The bold values refer to the most used and prominent cognitive workload measurement modalities highlighted in Figures 6.2 and 6.3. We did not include questionnaires in the table due to the large number of entries.

Type	Metric	References
Cerebral	**EEG** Frequencies	[19, 20, 36, 75, 94, 103, 120, 126, 129, 161, 171, 177, 204, 211, 216, 219, 232, 235, 291, 318, 320, 339, 382, 397, 399, 440]
	EEG ERPs	[33, 36, 83, 118, 291, 349, 373, 397]
	Hemodynamic Response (fNIRS)	[3, 5, 44, 116, 142, 143, 170, 248, 253, 307, 308, 312, 368, 369, 373, 391, 426, 427]
	Electromagnetic Changes (fMRI)	[364]
Ocular	Pupil Diameter	[4, 5, 10, 11, 13, 22–24, 26, 30, 38, 39, 57, 64–67, 78, 81, 86, 87, 104, 129, 138, 140, 160, 166, 181, 182, 189, 223, 224, 238, 239, 245, 259, 272, 302, 303, 305, 311, 328–330, 339, 350, 373, 383, 390, 407, 408, 420, 422, 437]
	Saccades	[11, 26, 39, 64–67, 109, 129, 160, 259, 291, 309, 375, 383, 423]
	Eye Blinks	[4, 5, 39, 63–67, 109, 129, 160, 272, 289, 291, 383, 440]
	Fixations	[2, 11, 77, 95, 160, 221, 238, 291, 362, 375]
	Smooth Pursuits	[213]
Cardiovascular	**Heart Rate**	[4, 11, 35, 39, 58, 59, 70, 72, 109, 112, 119, 125, 129, 160, 177, 227, 237, 250, 263, 290, 291, 310, 330, 332, 349, 351, 356, 370, 371, 373, 392, 399, 435, 435, 437, 440]
	Respiration	[129, 160, 177, 290, 291, 310]
	Temperature	[1, 2, 9, 129, 290, 291, 310, 351, 356, 440]

Dermal	Skin Conductance (EDA)	[35, 72, 73, 109, 112, 129, 160, 177, 214, 288–291, 310, 330, 332, 350, 351, 356, 370, 371, 373, 440]
Task Performance	Number of Errors	[15, 41, 85, 191, 437]
	Task Completion Time	[8, 65, 85, 95, 238, 291, 396]
	Task Accuracy	[65, 238]
	Reaction Time	[68, 115, 119]
	Keystroke Dynamics	[46, 373, 401]
	Language Production	[46, 401]
	Real-Time Self-Report	[372]
Haptic	Pen Input	[341–343, 425]
	Touch Input	[57, 165, 268, 406]
Speech	Lexical Density	[195]
	Tempo	[62]
	Pauses	[193, 194]
	Spectral Domain	[424]
	Pitch Contour	[57]
	Linguistic Features	[196, 403]
Body Movement	Muscle Contractions (EMG)	[59, 330]
	Glance Duration	[85]
	Mouse Movement	[14, 197]
	Postural Behavior and Gestures	[141, 373]
Biomarkers	Salivary Cortisol	[70]

6.5.1 Choosing Suitable Metrics for Experiments

Based on our understanding of the corpus of work included in this review, we propose a four-step procedure (illustrated in Figure 6.4), which can help researchers to choose ways to operationalize and measure the dependent variable cognitive workload in experimental settings. Our procedure should only be used to select candidate modalities for user studies of interfaces and concerns primarily the design goals of the system to be studied. The procedure serves primarily as a thinking tool for understanding the possibilities for measuring cognitive workload induced by an interactive system. The choice of modality should be based on good research practice and the requirements for validity in conducting studies using a particular modality.

Step 1: *Are you actually evaluating cognitive workload?*

When considering the measurement of cognitive workload in an experiment, the first step is verifying if cognitive workload is the right concept. We observed that many papers equated cognitive workload to usability, flow, or design quality throughout the corpus. Thus, we propose double-checking if the expected difference between the experimental conditions in cognitive workload would not be a product of a different underlying concept. For instance, if the interface is frustrating due to low usability in one of the conditions, measuring cognitive workload may obscure the underlying cause of the difference. In other words, when researchers decide to measure cognitive workload, they must ensure that possible sources of cognitive workload can be reliably identified.

Step 2: *Do you expect cognitive workload to vary during system usage?*

If a researcher decides to define cognitive workload as their dependent variable, the next step is to choose the metrics and measurement instruments (i.e., the operationalization of the dependent variable). The corpus in this review shows that researchers most likely prefer questionnaires due to their practical benefits – they are quick to administer and require a single interaction. In addition, the high number of past papers using questionnaires implies that using a questionnaire offers the advantage of comparing one's work with a large body of past research. There are, however, established drawbacks to this method. Questionnaires usually provide only one measurement point per condition and rely on retrospection. Further, they can distract the user from interacting with the system. However, the wide use of questionnaires shows that, in many cases, the benefits outweigh the disadvantages. Thus, if a single-point, summative assessment of cognitive workload is enough to verify the hypotheses or answer the research questions in an experiment, a questionnaire is an appropriate choice. If richer, formative data is needed, researchers should use other methods. The procedure ends with this step if a questionnaire is chosen.

Step 3: *Does your system need to adapt to changes in user's cognitive workload?*

Having decided that the study will use more granular measures than questionnaires for cognitive workload, we suggest that researchers inquire if the system will use cognitive workload to learn about the implications and improve the interface design or as user input itself, i.e., whether the study will involve elements of the studied systems changing their parameters based on measured cognitive workload. Our corpus showed that such systems are increasingly present in literature. In cognitive workload-based adaptation, the metric considerations may primarily stem from how the system is implemented. In such systems, the recommendation is to **use modali-**

ties that offer robust data that can be processed quickly. For example, cognitive workload data may need to be processed rapidly for the system to adapt interactively. The input should also be easily interpreted so the system can adapt algorithmically. In contrast, if the system does not need to react to the current cognitive workload, researchers can choose complex, data-intensive metrics that can be analyzed and interpreted after completing the study. Thus, when cognitive workload is not used for adaptation, we recommend to **use modalities that offer rich data for detailed analysis**. This may offer additional insight into the workload induced by the system.

Step 4: *Choose a modality based on the considerations*

The final step in the procedure is choosing a metric of the required richness and complexity. The selected metric must provide a volume of data that can be effectively processed and offer the necessary richness to understand a given system. Researchers can also employ multiple metrics to build hybrid assessments of cognitive workload. This, however, appears to be complex as the fraction of papers using such procedures in our corpus is low (see Figure 6.2). To facilitate this difficult choice, we provide a classification for the most common measurement modalities in terms of increasing complexity (i.e., data volume) and decreasing speed for the last step of this procedure (see Figure 6.4). An informed choice of modalities can best be made by studying the examples provided for each modality in Table 6.1.

6.5.2 A Note on Selecting Questionnaires

If a researcher decides to end the procedure in step two and use a questionnaire, they can choose which questionnaire to use. Our corpus shows that the NASA-TLX is significantly more prevalent in HCI research and may often appear to be the optimal choice. However, the reasons for the scale's success are unclear. The primary reason behind the use of NASA-TLX appears to be community convention (cf. Section 6.6.2). Consequently, we would like to encourage researchers to consider alternative scales that may offer advantages over the NASA-TLX. For example, Zijlstra's [438] RSME is a single-item rating and may enable rapid evaluations or be used in experience sampling; the ISA scale [388] can provide near real-time assessment; while the Bedford Workload Scale [337] can provide insights into potential mental overload situations for the user—all roles for which the NASA-TLX is not well suited. Further, one should not forget that the use of certain metrics places strict requirements on the experimental design, e.g., block designs are usually recommended for cerebral modalities. Thus, the final design of an experiment that uses cognitive workload measurement is almost always a trade-off between the ecological validity of the task and the requirements posed by the cognitive workload measurement modality.

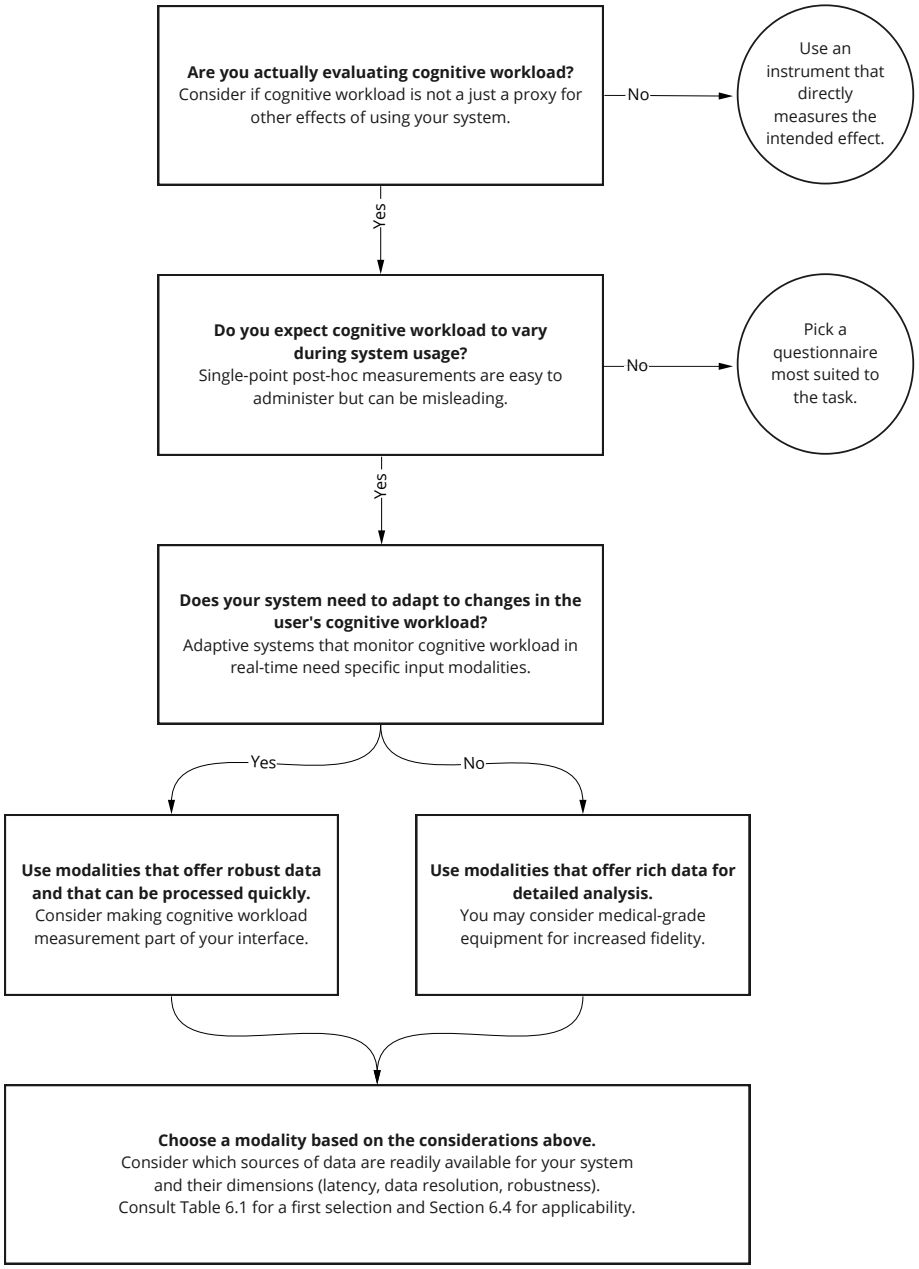


Figure 6.4: A step-by-step procedure to select appropriate cognitive workload measurements. HCI researchers can use this procedure as a reference for future experiments.

6.6 Future Directions and Challenges

Research in HCI has a broad and, at times, vague understanding of the concept of cognitive workload, as evident from the results of our review. We presented current practices, related measurement modalities, and metrics often used in contemporary HCI research (cf. Section 6.4). Additionally, our step-by-step guide on choosing suitable metrics (cf. Section 6.5) provides a functional categorization for researchers to choose a suitable cognitive workload measurement modality. However, our review has also showcased open research gaps, where choosing the correct metric can still be suboptimal depending on the research question. Choosing the right assessment for the right task has been recently discussed in various research publications [e.g. 18, 84, 88, 89, 206, 297, 436], but especially HCI research needs to deepen and conceptualize the understanding of cognitive workload in interacting with computer systems.

6.6.1 Defining and Interpreting Cognitive Workload in HCI

The results of our review show that cognitive workload is perceived as a qualitative component of an interactive system. The presented interfaces are designed to handle the user's cognitive resources economically. This is communicated as an explicit design goal or, more commonly, implied while evaluating user interfaces. In most reviewed papers, cognitive workload is used as an implicit concept. It is assumed that the reader is familiar with it. Our review showed that our search terms initially resulted in many papers that mention cognitive workload as a justification for specific system design choices (see Section 6.3.4). Consequently, low cognitive workload is often treated as equivalent to high system quality without necessarily determining what features or properties of the system's design lead to a perception of quality. The concept of cognitive workload in HCI is generalized without taking the individual, temporally changing, cognitive resources of users into account. As there is no definition widely recognized by the community, researchers should define their understanding of cognitive workload, if the concept is used to explain study results:

Research Gap 1: *HCI implicitly uses cognitive workload without discussing which workload components are measured (e.g., working memory, visual imagery). Research should engage with the underlying theories of workload to improve how empirical results are interpreted.*

We suggest that researchers define their understanding of cognitive workload in cases where cognitive workload levels are reported. Consequently, researchers must elab-

orate on how they expect different interfaces or interface adaptations to affect cognitive workload. **Thus, it remains a challenge to study how different usability components and usability-related qualities of interactive systems are linked to different types of cognitive workload measured by various methods.** There is a need to evaluate novel systems in terms of both cognitive workload and usability using robust, multi-dimensional measures. In that case, **the HCI field can gradually build a meta-understanding of the relation between usability and cognitive workload**.

6.6.2 The Hidden Cost of the NASA-TLX—A Legacy Issue?

A key finding of our literature review is that the HCI field prominently relies on using questionnaires and particularly on the NASA-TLX questionnaire for assessing cognitive workload. Previous work suggests combining the NASA-TLX questionnaire with other measures to increase the meaningfulness while improving the interpretation of cognitive workload measures [88]. One could risk stating that the NASA-TLX questionnaire has become a local standard for the HCI research community. This is likely due to historical reasons since many HCI pioneers had roots in Human Factors engineering. The NASA-TLX was prominently used in Human Factors engineering in the nascent days of HCI. Since then, the HCI community has applied the scale in various modalities and contexts, leading to individual deviations in the subjective interpretation of the NASA-TLX. Although the NASA-TLX questionnaire is affected by several drawbacks, it has remained a frequently used workload measure throughout the past decades. Potential reasons for the continued widespread use of the NASA-TLX include its simple usage and analysis of the questionnaire data. In addition, the fact that NASA-TLX results can be rapidly obtained facilitates its use in comparative studies. Despite these benefits, reliability and replicability are limited when sampling individuals. Our results show that using the NASA-TLX could be interpreted as an academic tradition, despite being introduced to the HCI field without an extensive explanation [45]. However, given the ever-growing number of research contributions in the HCI field, it is worthwhile to critically reflect on how the HCI field uses and interprets the results of the NASA-TLX questionnaire.

Our review shows that convenience and speed are key advantages of the NASA-TLX and questionnaires in general. In particular, the unweighted variant of the NASA-TLX, commonly known as Raw NASA-TLX, provides a quick assessment of perceived task load, including cognitive workload. Researchers often use it in repeated-measures experimental designs where participants must complete the questionnaire multiple times. If practical considerations are the primary motivation for choosing a cognitive workload evaluation method, one might wonder if convenience does not negatively impact how the evaluation is conceptualized; mainly since, in most stud-

ies in our corpus, only the unweighted scores (raw NASA-TLX) are analyzed. Such a choice offers the convenience of not hypothesizing which components examined in the NASA-TLX (cf. Section 6.4.1) are crucial in a given system. Instead, **the scale is employed as an accurate universal measure of cognitive workload *per se***. Furthermore, the NASA-TLX questionnaire is susceptible to individual biases through user expectations, leading to improved assessments of user interfaces by belief and placebos [218]. It remains a challenge for the HCI community to further reflect on *if* and *how* the workload induced by a novel interactive system can be accurately measured with a scale that is more than 30 years old and not designed for HCI research:

Research Gap 2: *HCI still relies on scales for measuring cognitive workload adapted from another field in the early days of HCI. The validity of these scales should be reassessed. We must develop new measurement instruments for cognitive workload specific to human-computer interaction.*

Consequently, future researchers should be wary that the NASA-TLX is an inherited rather than an efficient tool. Our review found no empirical evidence to determine the range of artifacts or experiences to which the NASA-TLX would be applicable. This implies that **we currently do not know how correct the NASA-TLX is in estimating cognitive workload in HCI**. Hence, **future researchers should be cautious in ascribing explanatory properties to the scale's dimensions** Interestingly, the NASA-TLX often serves as a *fallback* or baseline measurement when assessing cognitive workload with other metrics (e.g., pupil diameter) to see how reliable a novel measurement is. Instead, future research should revert this methodology by first studying cognitive workload via robust physiological measurements to (1) validate the dimensions of the NASA-TLX and (2) identify how user studies have to be designed to minimize the biases of the NASA-TLX and questionnaires in general. We envision this direction ultimately leading to an adapted variant of the NASA-TLX that better fits the HCI domain as a pacing and evolving research area.

6.6.3 HCI Research as a Catalyst for Workload-Aware Systems

The literature review reveals that the concept of cognitive workload is inherently relevant to HCI research. However, HCI researchers still primarily rely on suboptimal evaluation methods adapted from adjacent research domains, often failing to capture the multi-faceted nature of interactive systems. We argue that HCI researchers are equipped with the tools to explore further the role of cognitive workload in interacting with computers beyond measuring performance and usability. Integrating

cognitive workload as an input parameter for workload-aware systems should be at the heart of HCI's expertise. Understanding and using cognitive workload in interactive systems can effectively support users in real time when performing their tasks. How the design of such support (e.g., through adaptive workload-aware mixed reality interfaces [203, 215]) would look like remains a question for future research. Thus, HCI should not be content with a static evaluation of interfaces but look at how we can create workload-aware systems that make best use of the user's current cognitive resources:

Research Gap 3: *Research in HCI needs to capitalize on its opportunities and create interactive systems utilizing cognitive workload as an input parameter for workload-aware systems. The field should innovate by combining established methods (i.e., questionnaires) with novel measures (i.e., physiological sensing) to gain detailed insights into the user's cognitive demand.*

Our literature review showed that **physiological measures can provide a detailed, real-time understanding of the user's cognitive processing**. Consequently, HCI should lead in creating interactive systems capable of supporting an individual's optimal point for cognitive engagement. We strongly believe that real-time cognitive workload sensing capabilities will be essential to achieve this goal.

6.7 Chapter Conclusion

This chapter presented a literature review of current practices for measuring cognitive workload – a potential dependent variable in experimental settings studying Complementary Interfaces or interactive systems in general. Our survey found 579 relevant papers that assess cognitive workload using questionnaires, physiological sensing, and human behavior. Our results show that the contextualization and selection of suitable cognitive workload measures is a challenging subject in HCI research. We present a step-by-step procedure for selecting a suitable cognitive workload measurement modality for their studies to guide researchers. Our survey reveals three research gaps in the current landscape of HCI research on cognitive workload: improving the definition, broadening the choice of measures, and developing more workload-aware systems. The gaps intend to stimulate future research on cognitive workload assessments in HCI. Future researchers will determine how cognitive workload measurements can be integrated into their interaction paradigms and evaluation settings. Our review shows that questionnaires are a popular method to assess

cognitive workload due to their easy deployment, use, and analysis. However, the insights into cognitive processes obtained using questionnaires could be more extensive. In contrast, more advanced cognitive workload metrics (e.g., through physiological sensing or user behavior) require user-friendly methods, interfaces, and deployment strategies to be adopted in future research. Our work provides a structured overview of current cognitive workload measurements and the means for future researchers and designers to choose an appropriate metric for their particular study or design. Yet, this literature survey makes evident that HCI lacks a proper understanding of the concepts underlying cognitive workload. Thus, research towards tailored frameworks and theories is necessary to consolidate the understanding of cognitive workload in HCI.

Part IV
Conclusion

Parts of the following Chapter 7 have been published as:

***Johannes Zagermann**, Sebastian Hubenschmid, Daniel Immanuel Fink, Jonathan Wieland, Harald Reiterer, and Tiare Feuchtner. 2023. Challenges and Opportunities for Collaborative Immersive Analytics with Hybrid User Interfaces. In: "HybridUI: 1st Workshop on Hybrid User Interfaces: Complementary Interfaces for Mixed Reality Interaction" (IEEE ISMAR 2023 adjunct).*

The responsibilities of this joint publication were divided as follows: I spearheaded the writing. Sebastian Hubenschmid, Daniel Immanuel Fink, and Jonathan Wieland supported the writing process. Harald Reiterer and Tiare Feuchtner supervised the work.

Johannes Zagermann, Sebastian Hubenschmid*, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. 2022. Complementary Interfaces for Visual Computing. it - Information Technology 64, no. 4-5, 145-154. DOI: https://doi.org/10.1515/itit-2022-0031 *contributed equally to this research.*

The responsibilities of this joint publication were divided as follows: Sebastian Hubenschmid and I equally spearheaded the writing of the paper. Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, and Albrecht Schmidt contributed to the writing of the paper. Harald Reiterer contributed to the writing and supervised the work.

Sebastian Hubenschmid, **Johannes Zagermann***, Daniel Immanuel Fink*, Jonathan Wieland*, Tiare Feuchtner*, and Harald Reiterer*. 2021. Towards Asynchronous Hybrid User Interfaces for Cross-Reality Interaction. In: ACM 2021 Workshop "Transitional Interfaces in Mixed and Cross-Reality: A new frontier?" (TI 2021) (ACM ISS '21 adjunct). *contributed equally to this research.*

All authors contributed equally to this research.

Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Collaborative Hybrid UIs

Complementary Interfaces

Asynchronous Hybrid UIs

7

Conclusion

"Don't adventures ever have an end? I suppose not. Someone [...] always has to carry on the story."

Bilbo Baggins, The Fellowship of the Ring

Contents

In this thesis, we investigated the concept of Complementary Interfaces. For three use cases of Complementary Interfaces, we conducted individual experiments that served as design-informing activities for future Complementary Interfaces and Ubiquitous Computing environments in general. Additionally, a spotlight was put on the measurement of cognitive workload in HCI as a potential dependent variable in experimental situations involving Complementary Interfaces and beyond. In this chapter, we provide a summary of the three use cases of Complementary Interfaces and the measurement of cognitive workload. Additionally, the contributions and results of this thesis are presented, and an outlook on future activities with closing remarks concludes this thesis.

7.1 Summary

An increasing number of interactive devices is finding their way into our everyday lives. People own several devices [82] and find their own ways to combine them [295]. Prior research has found interest in these UbiComp environments and pushed the boundaries of technological possibilities by sophistically blending devices together and following a "the more the merrier" approach – a common theme in HCI research that is often "much better at proposing new technologies than at validating

them" [149] – leading to a fragmented research landscape and a need for evaluation [50].

This thesis set out to address this need for evaluation by studying Complementary Interfaces as representatives for UbiComp environments using controlled lab-based experiments that served as design-informing activities – based on the guiding research question of how to design and evaluate Complementary Interfaces.

Chapter 2 provided the foundations for Complementary Interfaces. Here, the historical developments of computing devices that led to Mark Weiser's famous vision of the computer for the 21st century [409] were described. This vision was aligned with current research streams such as Cross-Device Interaction, Hybrid User Interfaces, and Multimodal Interaction. By elaborating on the essence of what makes these novel environments worthwhile, a point was made that the complementarity of interactive components is the key to successful interaction (cf. [91]). Based on this, Complementary Interfaces were defined as meaningful combinations of interfaces that support users in their current task at hand. Experiments were chosen as an empirical way to study these Complementary Interfaces as they guarantee for a high validity of measurements but also as experiments can serve as design-informing activities [296]. Typical components of experiments were described that framed the research in this thesis, leading to three experiments studying a variety of Complementary Interfaces and a survey on the measurement of a possible dependent variable for experimental settings.

The first Complementary Interface was presented in Chapter 3. Here, the research was guided by the following question: "*What are the influences of the size of a shared interactive surface on collaboration in a Complementary Interface?*" (RQ1). To address this question, a Complementary Interface that involved individual tablets for personal activities and a complementary shared interactive surface for collaborative activities was studied. Here, the size of the shared interactive surface was determined as the independent variable. Dyads were working with one of the three conditions – mapping common sizes for personal and shared devices – and were tasked to solve an authentic sensemaking task in a between-subjects experimental design. The three conditions resembled valid design alternatives for this Complementary Interface, showing that the choice for the size of the shared interactive surface should depend on the task at hand and that bigger is not necessarily better: While all conditions led to comparable results in the quality of task completion, a larger shared surface led to more playful interaction and a tablet-sized shared surface resulted in more face-to-face communication.

Chapter 4 investigated the question "*What are the influences of embodiment (represented as meaningful combinations of input and output modalities) on spatial memory in a Complementary Interface?*" (RQ2). Therefore, a Complementary Interface was studied that involved a variety of meaningful combinations of input and output modalities. The combinations led to different manifestations of kinesthetic cues and feedback. Here, three input modalities served as conditions for the within-subjects aspect, and two display sizes served as conditions for the between-subjects characteristics – resulting in a 2×3 factor split-plot counterbalanced experimental design. The choice for this experimental setting was grounded in prior work that either investigated aspects of input or output, but not their combination – a key for Complementary Interfaces. The results show that participants engage in distinct interaction strategies when working with different combinations, which is shown in two trade-offs: The *Efficiency vs. Spatial Memory* trade-off shows that navigation speed negatively correlates with the navigation and spatial memory performance, and the *User Satisfaction vs. Spatial Memory* trade-off shows that a subjective user satisfaction score negatively correlates with the performance of spatial memory. Thus, the interrelation of user satisfaction, efficiency, and spatial memory leads to different usage patterns that participants take on to fulfill the task at hand.

The task at hand was one of the key elements in the last Complementary Interface presented in this work in Chapter 5. Here, the guiding question was "*What are the influences of interaction techniques on users' working behaviors and utilization of components (e.g., devices) in a Complementary Interface?*" (RQ3). The experimental setting included a methodological approach of using the study task as an independent variable, which allowed to study the interplay of tablet utilization, interaction techniques, and task-related aspects. Therefore, a Complementary Interface was studied in an experimental setting that included two independent variables: Interaction techniques and tasks. This allowed us to gain usability-related insights concerning the studied interaction techniques for the first task. We pre-defined the order of tasks to uncover intended learning effects. Therefore, we could observe how participants create purposes and meanings for the included devices, interaction techniques, and their interplay. Results showed that interaction techniques have less influence than expected, that work behaviors and device utilization depend on the task at hand, and revealed quality features of Complementary Interfaces. The key to uncovering this was our methodological approach, resulting in a Complementary Evaluation – revealing another asset of Complementary Interfaces.

Methodological aspects were also central aspects of Chapter 6. While the previous chapters focussed on the empiricism of Complementary Interfaces, this chapter put a spotlight on the question "*How to measure cognitive workload?*" (RQ4). Prior work

has shown the influence of handling multiple devices and modalities on cognitive workload [325, 439], therefore, cognitive workload can be seen as a potential dependent variable in experimental settings for Complementary Interfaces. Thus, this chapter presents a literature survey on the measurement of cognitive workload in HCI in general – presenting assessment methods such as questionnaires, physiological sensing, and human behavior. Results show that the selection of suitable cognitive workload measures is a challenging subject. Based on our experiences, we, therefore, present a step-by-step procedure for selecting suitable measurement modalities and reveal three research gaps: Defining and interpreting cognitive workload in HCI, the hidden cost of the NASA-TLX, and HCI research as a catalyst for workload-aware systems, highlighting that research has to deepen and conceptualize the understanding of cognitive workload (e.g., in Complementary Interfaces and beyond).

7.2 Contributions and Results

Research in this thesis can be classified into five types of contributions that are relevant to the HCI community [417]: 1) empirical, 2) artifact, 3) theoretical, 4) methodological, and 5) survey.

The concept of Complementary Interfaces is one of the foundations of this work. This theoretical contribution stems from the knowledge of designing and evaluating a variety of interactive systems in the area of Ubiquitous Computing (e.g., Cross-Device Interaction, Multimodal Interaction, or Hybrid User Interfaces). Complementary Interfaces draw their strengths from meaningful combinations across and within input and output modalities; and they are dependent on a given context. The concept of Complementary Interfaces can serve as a framework for designing and evaluating interactive systems that include multiple interactive components with the goal of supporting users in their task at hand.

To empirically study the concept of Complementary Interfaces, three artifacts were created that represent the individual experimental apparatus. These three artifacts are purposefully targeted to answer specific research questions.

The first experiment has shown the influence of the size of a shared interactive surface on collaborative activities. Here, findings show that a larger shared interactive surface does not necessarily improve collaboration or sensemaking results, as they can divert the attention away from collaborators. This has implications on the interaction with the shared interactive surface and the communication style of dyads – showing that a smaller shared interactive surface can lead to more face-to-face communication.

The second experiment investigated a variety of meaningful combinations of input and output modalities and their effect on spatial memory, efficiency, and user satisfaction. The results show that participants engage in different interaction strategies depending on the input modality. These strategies are characterized by trade-offs regarding spatial memory vs. efficiency as well as spatial memory vs. user satisfaction.

The third experiment investigated interaction techniques and their relation to device utilization with tablets. Here, findings show that the influence of interaction techniques is lower than expected, and participants favor unfamiliar yet faster techniques. This can lead to different problem-solving strategies facilitated by a high utilization of devices and different quality features of Complementary Interfaces. The high utilization of devices for a task with a high dependency of devices can be seen as a training and guidance in using a similar high utilization of devices for a task with lower dependency and restriction.

This third experiment included a methodological contribution. While each experiment required adapting existing methods and creating new methods, the last experiment included a methodological approach that led to a unique experimental setting. Here, integrating the study task as an independent variable in the experimental setting allowed us to gain different yet complementary insights into a research problem. While the abstract classification task allowed us to gain insights into usability-related aspects of the usage of interaction techniques and devices, the secondary sensemaking task – with fewer restrictions regarding the usage of techniques and devices – allowed us to observe how participants created distinct purposes for them. This was coined as Complementary Evaluation techniques and can be seen as a facet of Complementary Interfaces.

Lastly, the meta-analysis and literature survey of past and present measurements for cognitive workload in HCI is a survey contribution. The systematic search yielded an initial corpus of 6414 papers, which was reduced to the final literature corpus of 579 papers. Based on this, we could classify measurements for cognitive workload in HCI. This classification is further extended by a step-by-step approach that can guide researchers in choosing a suitable measurement modality for studying cognitive workload in their own experimental settings. Additionally, we provide research gaps that can serve as a research agenda for future work.

7.3 Future Work

This thesis addressed the overall research goal of designing and evaluating Complementary Interfaces. Therefore, three use cases of Complementary Interfaces were empirically studied, focusing on homogeneous types of devices, e.g., tablets and in-

teractive surfaces. Experimental comparisons of valid design alternatives allowed us to understand causes and their effects; serving as design-informing activities. The spotlight on measuring cognitive workload as a potential dependent variable for such experimental settings complemented the described work. However, the used concepts, methods, tasks, and measurements are not limited to the studied use cases and can be transferred and expanded to more dimensions: As described in Chapters 1 and 2, I was already involved in the design and evaluation of Complementary Interfaces based on meaningful combinations of e.g., tablets and smartphones for precise input and head-worn devices for close-to-unlimited output (e.g., STREAM [155] and IDIAR [402]). The evaluation of these two Complementary Interfaces was rather focused on the feasibility of the concepts than an actual comparative user study involving different independent variables or valid baseline conditions (also known as point studies or single-shot studies [296]).

One of our recent works shows how the ideas of Complementary Interfaces that build on heterogeneous device combinations and experimental situations can be combined: With "ARound the Smartphone" [159], we studied the effects and influences of virtually-extended mobile devices on spatial memory (see Figure 7.1).

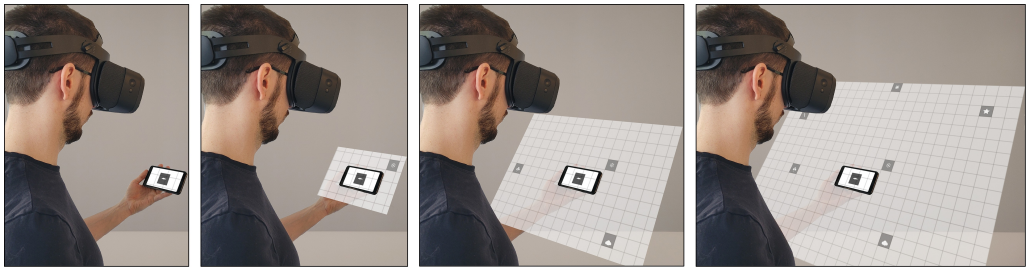


Figure 7.1: Studying virtually-extended displays by comparing four display sizes: Featuring no AR extension (baseline), a small AR extension (cf. a tablet), a medium AR extension (cf. a desktop display), and a large AR extension (cf. a TV) (from left to right).

With this Complementary Interface, we transferred the experimental situation (including e.g., tasks, measurements, and procedures) that was described in Chapter4 into an apparatus that allowed us to closely study the influence of design alternatives for a virtually-extended smartphone on spatial memory. The report of the experimental comparison and findings is not part of this thesis; however, it showed new possible directions for studying Complementary Interfaces and made evident that more advanced tools are necessary for the analysis and evaluation of Complementary Interfaces that involve heterogeneous devices, components, and modalities.

For the analysis of such experimental settings, we designed, developed, and evaluated an initial prototype of an evaluation framework (ReLive [157]). ReLive is a

Complementary Interface itself, combines immersive and non-immersive visualizations (i.e., in-situ and ex-situ) for the holistic analysis and exploration of user study data, and can be used by individual analysts asynchronously [156]. Potentially, ReLive can also be used by multiple analysts synchronously, resulting in a collaborative Complementary Interface across and within different levels of immersion. Interestingly, prior work has highlighted the benefits of collaboration for analyzing, interpreting, and understanding visualizations [167] and defined collaborative analytics as well as the establishment of an evaluation framework as grand challenges for immersive analytics [93]. Complementary Interfaces are not necessarily anchored within immersive analytics, yet, we have shown that Complementary Interfaces such as STREAM [155] or ReLive [157] are valid settings to support analysts within visual and immersive analytics.

Figure 7.2: An illustrative example of a prospective collaborative Complementary Interface based on ReLive for the joint analysis and exploration of Mixed Reality user study data. Here, one analyst studies aggregated interaction data using a 2D representation in a non-immersive desktop view (left), and another analyst is looking at the same time at the same data, visualized as 3D trajectories in an immersive environment. Their views are synchronized and linked, allowing for tightly and loosely coupled collaborative activities and switches and transitions between and across environments. This figure is based on previous work on ReLive [157] and asynchronous Hybrid User Interfaces [155] and combined using Adobe Photoshop (beta)'s generative fill function.

A prospective collaborative variant of ReLive (see Figure 7.2) is a perfect playground for the concepts presented and studied in this thesis: ReLive is a Complementary Interface that allows for the evaluation and analysis of Mixed Reality user study data. In a collaborative scenario, this might lead to dynamic meaningful combinations of modalities, interaction techniques, and devices – within and across space, time, and levels of immersion: The combination of synchronized immersive and non-immersive components that allow for dynamic and fluid transitions within and across all levels of immersion comes with future research opportunities regarding the task allocation (exploring and analyzing Mixed Reality user study data includes several subtasks),

and the interplay of components (e.g., transitions and coordination between immersive and non-immersive environments). From our own work, we gained experiences in designing, developing, and evaluating collaborative scenarios for co-located (e.g., [414, 428]) and remote collaboration (e.g., [100, 277]). Traditionally, collaboration can be classified according to Johansen's Space-Time Matrix [183]: Here, collaboration either happens at the same or different place (i.e., co-located or remote) and at the same or different time (i.e., synchronous or asynchronous). However, having a mixed-immersion environment, we have to add another dimension to this classification: level of immersion. This additional dimension comes with research opportunities regarding transitional user representations (e.g., from blinking cursors in shared documents to highly realistic 3D user representations) and collaborative coupling styles and symmetry of collaboration (e.g., different views might lead to differences in roles).

The following subsections will further elaborate on the mentioned research opportunities for collaborative Complementary Interfaces to evaluate and analyze interaction in e.g., Complementary Interfaces. Adressing these research opportunities will likewise inform the design [296] and evaluation methods [149] of Complementary Interfaces. The research opportunities are combined from our previous efforts on ReLive [157] and the general challenges for Complementary Interfaces as described in Section 2.6. We intentionally do not limit the following sections to a specific type of collaboration (e.g., by specifying time, place, level of immersion, or roles) as a prospective collaborative variant of ReLive might allow for highly dynamic transitions between these attributes.

7.3.1 Research Opportunity 1 – Task Allocation

The use case of exploring and analyzing user study data includes several subtasks (e.g., gaining overview, reasoning, or comparing data across sessions). However, it needs to be clarified, which subtask is more suitable to be performed in an immersive or non-immersive environment. Ideally, both environments provide equal analysis opportunities and synchronized data access. However, at least two aspects might shape the way analysts will use the framework: (1) the device-specific affordances and constraints of each individual input and output modality can influence interaction and workflows (e.g., as described in Chapter 4); and (2) the view-specific capabilities of e.g., ego-centric navigation in 3D visualizations of spatial recordings (i.e., immersive view) and aggregated 2D visualizations (i.e., non-immersive view).

The immersive view might be beneficial to gain an overview of the situation and the non-immersive view, with its 2D visualizations, might be more suited for an in-detail analysis. However, it might also be the other way around: Analysts might be inter-

ested in gaining an overview regarding e.g., task completion times and continue their analysis in the immersive view to find reasons for differences in the data. Additionally, transitions between these two assumed approaches might be the case.

Similarly, tasks and activities that require, e.g., text input via a keyboard, will presumably be performed in the non-immersive view, whereas activities that require the 3D environment of the actual user study will instead be performed in the immersive view. However, the analysis workflow of Mixed Reality user study data can differ heavily across different user studies [32]. Therefore, future research should study analysis and exploration workflows regarding device-specific and view-specific efficiency, effectiveness, and preferences.

7.3.2 Research Opportunity 2 – Interplay of Components

Depending on the workflow, analysts might transition or switch between the immersive and non-immersive environments. Here, a discrete switch between the environments might be an efficient way to move from one view to the other [359]. However, this switch might come at the cost of a significant overhead [331] that might lead to losing context. Alternatively, a continuous transition could prevent analysts from reorienting in the data representations. Yet, as the input modalities of the immersive and non-immersive views might differ (e.g., mouse vs. controller), analysts might run into trouble when the handling of input devices requires more cognitive capacities than the current task at hand.

The loss of context due to discrete switches and the increased cognitive load of handling view-specific input modalities can be addressed via (1) the adaptation of established techniques for analyzing multiple coordinated views and (2) input devices that can be used across environments.

Multiple coordinated views can traditionally be analyzed with linking and brushing techniques: Here, selecting a data point in one visualization results in a visual highlight of the same data point in a coordinated visualization. Similarly, having synchronized immersive and non-immersive views, the potential for a cross-reality linking and brushing technique becomes promising. This might support analysts in exploring the visualized data and reduce the loss of context, as they might directly start where they left off in the previous step. Future research should study the effects and influences of cross-reality linking and brushing for analyzing Mixed Reality user study data. The envisioned cross-reality linking and brushing method could support the coordination within different data representations. In line with this, shared points of reference (i.e., shared across levels of immersion and with co-workers, e.g., virtual landmarks [276]) hold the potential to coordinate individual and collaborative activities. Additionally, a hybrid input device could be used across environments to reduce

the cognitive load of handling multiple modalities. In the non-immersive view, it can be used like a mouse and be placed on a desk environment. When transitioning to the immersive view, it can be used as a controller. However, this might affect the consistency of the involved components: Complementary Interfaces aim for a symbiosis of interfaces, where each component purposefully increases the quality of interaction and further supports users in their current activity. For the workflow of analyzing user study data, each component might bring in its individual strengths, but also affordances and constraints. For example, a desktop interface profits from the familiarity and precision of a WIMP interface, and a VR environment is more suited for 3D user interfaces. Future research should explore how relevant design factors such as consistency, affordances, and constraints in different levels of immersion influence user experiences. A high user experience is a key quality of Complementary Interfaces as they aim to support their users in their current activities. If they do not integrate well with workflows, they might not be used as intended [316].

So far, we have mainly looked at explicit interaction with Complementary Interfaces. However, a combination of implicit interaction [352] and explicit components might further increase the user experience of Complementary Interfaces. Integrating the findings of Chapter 6, different components of a Complementary Interface could assess users' current cognitive workload or visual focus (e.g., using built-in eye trackers of head-worn devices). This information can then be used to adapt and optimize the interfaces [69]: e.g., by switching to familiar representations [186] or reducing the amount of visual presentations [216], which could further lead to mutual adaptation scenarios [25]. Building on this, Complementary Interfaces could not only react to implicit user input, but also – e.g., using artificial intelligence or machine learning – explicitly suggest analyses (e.g., based on an automatic outlier detection), data visualizations (e.g., task completion times might be easier to understand in 2D than in 3D), or even optimal interactive components (e.g., for input or output) to view, understand, or interact with data – a potential further angle on addressing the legacy bias [316].

7.3.3 Research Opportunity 3 – Transitional User Representations

User representations (e.g., virtual avatars) can create awareness of a collaborator's activities [100]. In immersive environments, remote collaboration can be enriched with highly realistic avatars that provide qualities usually attributed to co-located collaboration, such as facial expressions, deictic referencing, or the sensation of presence. In common non-immersive interfaces (e.g., Overleaf, Google Docs, or Microsoft Office), collaborators are often represented as a colored cursor or representative icon to communicate e.g., the current position in a document. Having a Complementary Interface with immersive and non-immersive views, a user representation is not only

beneficial to support remote collaboration, but also to create awareness of co-located activities, especially when each collaborator is working within a different level of immersion. The two named options define the extremes of a user representation continuum: On the immersive end, there are highly realistic avatars resembling all qualities of a co-located collaborator, and on the non-immersive end, there is a cursor that shows the location within a document.

For remote collaboration, this leads to at least two research opportunities: (1) the user representation across different levels of immersion and (2) the transition of the user representation between different levels of immersion.

There is extensive research on user representations in Mixed Reality – from viewing frustums [277] over simple avatars [100] to highly realistic yet hardware-heavy representations (e.g., Holoportation [294]). However, combining different levels of immersion leads to new challenges: While we can use established techniques to represent others within the same level of immersion (e.g., avatars in virtual reality), it is unclear how to visualize the other person across different levels of immersion. Here, the user representation can also be further differentiated between being true to the physical environment (e.g., an avatar sitting at an office table working in the non-immersive view) and being true to the visualized data that is currently analyzed (e.g., positioning a user representation in close proximity to a 3D visualization in the immersive view).

Here, each view comes with unique opportunities: When being in the immersive view and the other person analyzes data within the non-immersive environment, this allows to represent this user, e.g., as an avatar sitting at an office table, re-arranged to match the local environment (cf. [100]); as a see-through option of the actual person and their environment (i.e., augmented virtuality); or as a representative icon that is bound to a data visualization (e.g., being true to the data that is currently analyzed). Additionally, the user representation could not only be used, e.g., to create awareness, but also to communicate the current capabilities of the person within their own local environment (e.g., interacting with a mouse). Also, this could further lead to multiple user representations of the same user, such as an avatar being true to the physical environment, sitting at an office table, and additionally, a representative icon being bound to a data visualization to further indicate the current activity (cf. cross-reality linking and brushing as described above).

When working in the non-immersive view and the other person currently analyzes data in the immersive environment, this allows representing the other user true to the data, e.g., as a representative icon bound to a data visualization. However, imagining a scenario where the user in the non-immersive view still wears a mixed-reality

headset with a see-through option, further allows to represent the other user as an avatar in augmented reality, allowing the analyst to proceed working within their local physical environment.

As the non-immersive and immersive views are ideally synchronized and allow analysts to fluidly switch or transition between the environments, their user representation should also match these transitions. Here, complementing the visual representation of users with additional auditory output can create awareness: On the one hand, symbolic sounds could indicate entering or exiting an environment. On the other hand, spatial sound can increase awareness of the other person's location and influence the sensation of presence [100].

Further, it is necessary to understand user representations for co-located activities. Being in the same physical space already brings some benefits compared to the remote counterpart (e.g., verbal communication). However, the opportunity to work within different levels of immersion leads to the need for some form of user representation. Similar to remote collaboration, a user representation can support the collaboration by increasing the awareness of each other's activities by representing the other person true to the physical environment or visualized data.

7.3.4 Research Opportunity 4 – Collaborative Coupling Styles and Symmetry of Collaboration

The combination of synchronized immersive and non-immersive views allows multiple analysts to gain a broader perspective on the visualized data. Here, the immersive view might invite analysts to explore 3D simulations of the actual study environment, while the non-immersive view can support analysts in detailed analyses, and vice versa. This can also lead to a mixed-focus collaboration [128, 387], where peers transition between loosely-coupled parallel work (e.g., individually analyzing data visualizations) and closely-coupled activities (e.g., active discussions on the data). Previous work identified coupling styles for a variety of collaborative scenarios (e.g., multi-device collaboration [49] or hybrid scenarios [282]), and we were successfully able to classify collaborative activities during remote collaboration using these well-established coupling styles [100]. Although coupling styles have been proven useful to classify and characterize collaboration in co-located and remote settings, it remains unclear to what extent they can be re-used to classify mixed-immersion collaboration. The characterization of collaborative activities (co-located and remote) in these mixed-immersion environments is crucial to describe the quality of interaction.

An additional way to characterize collaborative behavior is to describe its symmetry: We studied symmetrical collaborative activities (i.e., equal opportunities and

features) for remote settings [100]. However, for remote settings, there is often the need for asymmetrical settings [345], such as remote assistance scenarios, where an expert can guide a novice in solving a task (i.e., representing roles). Having the combination of immersive and non-immersive views can likewise lead to an asymmetry of collaboration. Here, one analyst might guide the other person in their exploration of the visualized data to identify, e.g., outliers in the data. This might also allow for a spectator view e.g., using handheld devices to observe a person's activity (e.g., similar to our user study setups for IDIAR [402] and STREAM [155]). This might be done within the same level of immersion or across different levels of immersion. Additionally, this might also be highly dynamic – leading to frequent switches of roles.

7.4 Closing Remarks

Complementary Interfaces aim to support users in their current task at hand using meaningful combinations of interactive input and output components, where each component can be seen as a piece of the puzzle. This leads to combinations of homogeneous (e.g., Cross-Device Interaction) and heterogeneous (e.g., Hybrid User Interfaces) device classes, but also input (e.g., interaction techniques) and output (e.g., visually or auditory) modalities. Ultimately, each component purposefully increases the quality of interaction. This dissertation's overall research goal was to design and evaluate Complementary Interfaces. Therefore, three use cases for Complementary Interfaces were evaluated using *experiments as design-informing activities* – addressing the need to evaluate such UbiComp environments. An additional meta-analysis and literature survey of current and past methods to measure cognitive workload complemented the research. With the research presented in this thesis, we started to scratch the surface of the potential of the concept of Complementary Interfaces and ways to evaluate them. In the upcoming years, we will see increasing numbers and even more diverse types of interactive devices in our everyday lives – potentially fueled by advances in artificial intelligence. This will emphasize the need to study the utility, relevance, and role of each piece of the puzzle to eventually create a *symbiosis of interfaces*.

Listings and Bibliography

List of Figures

List of Tables

Bibliography

[1] Abdelrahman, Y., Velloso, E., Dingler, T., Schmidt, A., and Vetere, F.: Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 1, doi: 10.1145/3130898, 2017.

[2] Abdelrahman, Y., Khan, A. A., Newn, J., Velloso, E., Safwat, S. A., Bailey, J., Bulling, A., Vetere, F., and Schmidt, A.: Classifying Attention Types with Thermal Imaging and Eye Tracking, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 3, doi: 10.1145/3351227, 2019.

[3] Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., Chang, R., and Jacob, R. J.: Dynamic Difficulty Using Brain Metrics of Workload, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, pp. 3797–3806, Association for Computing Machinery, Toronto, Ontario, Canada, doi: 10.1145/2556288.2557230, 2014.

[4] Ahmad, M. I., Keller, I., Robb, D. A., and Lohan, K. S.: A framework to estimate cognitive load using physiological data, Personal and Ubiquitous Computing, pp. 1–15, doi: 10.1007/s00779-020-01455-7, 2020.

[5] Amft, O., Wahl, F., Ishimaru, S., and Kunze, K.: Making Regular Eyeglasses Smart, IEEE Pervasive Computing, 14, 32–43, doi: 10.1109/MPRV.2015.60, 2015.

[6] Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., and Silva, C. T.: A User Study of Visualization Effectiveness Using EEG and Cognitive Load, Computer Graphics Forum, 30, 791–800, doi: https://doi.org/10.1111/j.1467-8659.2011.01928.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2011.01928.x, 2011.

[7] Andrews, C., Endert, A., and North, C.: Space to Think: Large High-resolution Displays for Sensemaking, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, pp. 55–64, ACM, New York, NY, USA, 2010.

[8] Antrobus, V., Burnett, G., and Skrypchuk, L.: "Turn Left At The Fairham Pub" Using Navigational Guidance to Reconnect Drivers With Their Environment, in: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Automotive'UI 16, pp. 35–42, ACM, Ann Arbor, MI, USA, doi: 10.1145/3003715.3005392, 2016.

[9] Anzengruber, B. and Riener, A.: "FaceLight": Potentials and Drawbacks of Thermal Imaging to Infer Driver Stress, in: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '12, pp. 209–216, ACM, Portsmouth, New Hampshire, doi: 10.1145/2390256.2390292, 2012.

[10] Appel, T., Scharinger, C., Gerjets, P., and Kasneci, E.: Cross-subject Workload Classification Using Pupil-related Measures, in: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, ETRA '18, pp. 4:1–4:8, ACM, New York, NY, USA, doi: 10.1145/3204493.3204531, URL http://doi.acm.org/10.1145/3204493.3204531, 2018.

[11] Appel, T., Sevcenko, N., Wortha, F., Tsarava, K., Moeller, K., Ninaus, M., Kasneci, E., and Gerjets, P.: Predicting Cognitive Load in an Emergency Simulation Based on Behavioral and Physiological Measures, in: 2019 International Conference on Multimodal Interaction, ICMI '19, pp. 154–163, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3340555.3353735, 2019.

[12] Apple Vision Pro: Introducing Apple Vision Pro: Apple's First Spatial Computer, https://www.apple.com/newsroom/2023/06/introducing-apple-vision-pro/.

[13] Arevalo, S., Miller, S., Janka, M., and Gerken, J.: What's behind a Choice? Understanding Modality Choices under Changing Environmental Conditions, in: 2019 International Conference on Multimodal Interaction, ICMI '19, pp. 291–301, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3340555.3353717, 2019.

[14] Arshad, S., Wang, Y., and Chen, F.: Analysing Mouse Activity for Cognitive Load Detection, in: Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, OzCHI '13, pp. 115–118, ACM, Adelaide, Australia, doi: 10.1145/2541016.2541083, 2013.

[15] Asif, A. and Boll, S.: Where to Turn My Car?: Comparison of a Tactile Display and a Conventional Car Navigation System Under High Load Condition, in: Proceedings of the 2Nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '10, pp. 64–71, ACM, Pittsburgh, Pennsylvania, doi: 10.1145/1969773.1969786, 2010.

[16] Auda, J., Gruenefeld, U., Kosch, T., and Schneegass, S.: The Butterfly Effect: Novel Opportunities for Steady-State Visually-Evoked Potential Stimuli in Virtual Reality, in: Proceedings of the 3rd Augmented Humans International Conference, AHs '22, ACM, New York, NY, USA, doi: 10.1145/3519391.3519397, 2022.

[17] Avellino, I., Fleury, C., Mackay, W. E., and Beaudouin-Lafon, M.: CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 6718–6729, ACM, New York, NY, USA, doi: 10.1145/3025453.3025604, URL http://doi.acm.org/10.1145/3025453.3025604, 2017.

[18] Ayres, P., Lee, J. Y., Paas, F., and van Merriënboer, J. J.: The validity of physiological measures to identify differences in intrinsic cognitive load, Frontiers in Psychology, 12, doi: 10.3389/fpsyg.2021.702538, 2021.

[19] Babu, A. R., Rajavenkatanarayanan, A., Brady, J. R., and Makedon, F.: Multimodal Approach for Cognitive Task Performance Prediction from Body Postures, Facial Expressions and EEG Signal, in: Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data, MCPMD '18, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3279810.3279849, 2018.

[20] Baceviciute, S., Mottelson, A., Terkildsen, T., and Makransky, G.: Investigating Representation of Text and Audio in Educational VR Using Learning Outcomes and EEG, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, pp. 1–13, Association for Computing Machinery, Honolulu, HI, USA, doi: 10.1145/3313831.3376872, 2020.

[21] Badam, S. K., Mathisen, A., Rädle, R., Klokmose, C. N., and Elmqvist, N.: Vistrates: A component model for ubiquitous analytics, IEEE transactions on visualization and computer graphics, 25, 586–596, 2018.

[22] Bailey, B. P. and Busbey, C. W.: TAPRAV: A Tool for Exploring Workload Aligned to Models of Task Execution, in: Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '06, pp. 467–470, ACM, Venezia, Italy, doi: 10.1145/1133265.1133360, 2006.

[23] Bailey, B. P. and Iqbal, S. T.: Understanding Changes in Mental Workload During Execution of Goal-Directed Tasks and Its Application for Interruption Management, ACM Trans. Comput.-Hum. Interact., 14, 21:1–21:28, doi: 10.1145/1314683.1314689, 2008.

[24] Bailey, B. P., Busbey, C. W., and Iqbal, S. T.: TAPRAV: An interactive analysis tool for exploring workload aligned to models of task execution, Interacting with Computers, 19, 314–329, doi: 10.1016/j.intcom.2007.01.004, URL https://doi.org/10.1016/j.intcom.2007.01.004, 2007.

[25] Balestrucci, P. and Ernst, M. O.: Visuo-motor adaptation during interaction with a user-adaptive system, Journal of Vision, 19, 187a, doi: 10.1167/19.10.187a, URL https://jov.arvojournals.org/article.aspx?articleid=2750667, 2019.

[26] Bartels, M. and Marshall, S. P.: Eye Tracking Insights into Cognitive Modeling, in: Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications, ETRA '06, pp. 141–147, ACM, New York, NY, USA, doi: 10.1145/1117309.1117358, 2006.

[27] Batch, A., Butcher, P. W., Ritsos, P. D., and Elmqvist, N.: Wizualization: A hard magic visualization system for immersive and ubiquitous analytics, IEEE Transactions on Visualization and Computer Graphics, 30, 2024.

[28] Baudisch, P. and Rosenholtz, R.: Halo: A Technique for Visualizing Off-screen Objects, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03, pp. 481–488, ACM, New York, NY, USA, doi: 10.1145/642611.642695, URL http://doi.acm.org/10.1145/642611.642695, 2003.

[29] Beaudouin-Lafon, M., Huot, S., Nancel, M., Mackay, W., Pietriga, E., Primet, R., Wagner, J., Chapuis, O., Pillias, C., Eagan, J., Gjerlufsen, T., and Klokmose, C.: Multisurface Interaction in the WILD Room, Computer, 45, 48–56, 2012.

[30] Bednarik, R., Bartczak, P., Vrzakova, H., Koskinen, J., Elomaa, A.-P., Huotarinen, A., de Gómez Pérez, D. G., and von und zu Fraunberg, M.: Pupil Size As an Indicator of Visual-motor Workload and Expertise in Microsurgical Training Tasks, in: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, ETRA '18, pp. 60:1–60:5, ACM, New York, NY, USA, doi: 10.1145/3204493.3204577, URL http://doi.acm.org/10.1145/3204493.3204577, 2018.

[31] Benko, H., Ishak, E., and Feiner, S.: Collaborative mixed reality visualization of an archaeological excavation, in: Third IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 132–140, doi: 10.1109/ISMAR.2004.23, 2004.

[32] Bergström, J., Dalsgaard, T.-S., Alexander, J., and Hornbæk, K.: How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems,

CHI '21, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3411764.3445193, URL https://doi.org/10.1145/3411764.3445193, 2021.

[33] Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., and Olmstead, R.: Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset, International Journal of Human-Computer Interaction, 17, 151–170, doi: 10.1207/s15327590ijhc1702_3, URL http://www.tandfonline.com/doi/abs/10.1207/s15327590ijhc1702_3, 2004.

[34] Bevan, N. and Macleod, M.: Usability measurement in context, Behaviour & Information Technology, 13, 132–145, doi: 10.1080/01449299408914592, URL https://doi.org/10.1080/01449299408914592, 1994.

[35] Bhandari, R., Parnandi, A., Shipp, E., Ahmed, B., and Gutierrez-Osuna, R.: Music-Based Respiratory Biofeedback in Visually-Demanding Tasks, in: Proceedings of the International Conference on New Interfaces for Musical Expression, edited by Berdahl, E. and Allison, J., pp. 78–82, Louisiana State University, Baton Rouge, Louisiana, USA, 2015.

[36] Bilalpur, M., Kankanhalli, M., Winkler, S., and Subramanian, R.: EEG-based Evaluation of Cognitive Workload Induced by Acoustic Parameters for Data Sonification, in: Proceedings of the 20th ACM International Conference on Multimodal Interaction, ICMI '18, pp. 315–323, ACM, New York, NY, USA, doi: 10.1145/3242969.3243016, URL http://doi.acm.org/10.1145/3242969.3243016, 2018.

[37] Bishop, F., Zagermann, J., Pfeil, U., Sanderson, G., Reiterer, H., and Hinrichs, U.: Construct-A-Vis: Exploring the Free-Form Visualization Processes of Children, IEEE Transactions on Visualization and Computer Graphics, 26, 451–460, doi: 10.1109/TVCG.2019.2934804, 2020.

[38] Biswas, P. and Langdon, P.: Multimodal Intelligent Eye-Gaze Tracking System, International Journal of Human-Computer Interaction, 31, 277–294, doi: 10.1080/10447318.2014.1001301, URL http://www.tandfonline.com/doi/full/10.1080/10447318.2014.1001301, 2015.

[39] Biswas, P., Dutt, V., and Langdon, P.: Comparing Ocular Parameters for Cognitive Load Measurement in Eye-Gaze-Controlled Interfaces for Automotive and Desktop Computing Environments, International Journal of Human-Computer Interaction, 32, 23–38, doi: 10.1080/10447318.2015.1084112, URL http://www.tandfonline.com/doi/full/10.1080/10447318.2015.1084112, 2016.

[40] Blum, S., Debener, S., Emkes, R., Volkening, N., Fudickar, S., and Bleichner, M. G.: EEG Recording and Online Signal Processing on Android: A Multiapp Framework for Brain-Computer Interfaces on Smartphone, doi: 10.1155/2017/3072870, URL https://www.hindawi.com/journals/bmri/2017/3072870/abs/, 2017.

[41] Boll, S., Asif, A., and Heuten, W.: Feel Your Route: A Tactile Display for Car Navigation, IEEE Pervasive Computing, 10, 35–42, doi: 10.1109/MPRV.2011.39, 2011.

[42] Borowski, M., Zagermann, J., Klokmose, C. N., Reiterer, H., and Rädle, R.: Exploring the Benefits and Barriers of Using Computational Notebooks for Collaborative Programming Assignments, in: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, pp. 468–474, 2020.

[43] Boucsein, W.: Electrodermal activity, 2nd ed, Springer Science + Business Media, New York, NY, US, doi: 10.1007/978-1-4614-1126-0, pages: xviii, 618, 2012.

[44] Boyer, M., Cummings, M., Spence, L. B., and Solovey, E. T.: Investigating Mental Workload Changes in a Long Duration Supervisory Control Task, Interacting with Computers, 27, 512–520, doi: 10.1093/iwc/iwv012, URL https://doi.org/10.1093/iwc/iwv012, 2015.

[45] Brewster, S. A., Wright, P. C., and Edwards, A. D. N.: The Design and Evaluation of an Auditory-Enhanced Scrollbar, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '94, pp. 173–179, Association for Computing Machinery, Boston, Massachusetts, USA, doi: 10.1145/191666.191733, 1994.

[46] Brizan, D. G., Goodkind, A., Koch, P., Balagani, K., Phoha, V. V., and Rosenberg, A.: Utilizing Linguistically Enhanced Keystroke Dynamics to Predict Typist Cognition and Demographics, International Journal of Human-Computer Studies, 82, 57–68, doi: https://doi.org/10.1016/j.ijhcs.2015.04.005, 2015.

[47] Brooke, J. et al.: SUS-A quick and dirty usability scale, Usability evaluation in industry, 189, 4–7, 1996.

[48] Brouwer, A.-M., Hogervorst, M. A., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R.: Estimating workload using EEG spectral power and ERPs in the n-back task, doi: 10.1088/1741-2560/9/4/045008, URL https://doi.org/10.1088%2F1741-2560%2F9%2F4%2F045008, 2012.

[49] Brudy, F., Budiman, J. K., Houben, S., and Marquardt, N.: Investigating the Role of an Overview Device in Multi-Device Collaboration, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 300:1–300:13, ACM, New York, NY, USA, doi: 10.1145/3173574.3173874, URL http://doi.acm.org/10.1145/3173574.3173874, 2018.

[50] Brudy, F., Holz, C., Rädle, R., Wu, C.-J., Houben, S., Klokmose, C. N., and Marquardt, N.: Cross-Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, pp. 562:1–562:28, ACM, New York, NY, USA, doi: 10.1145/3290605.3300792, URL http://doi.acm.org/10.1145/3290605.3300792, 2019.

[51] Bulling, A. and Zander, T. O.: Cognition-Aware Computing, IEEE Pervasive Computing, 13, 80–83, doi: 10.1109/MPRV.2014.42, 2014.

[52] Butscher, S. and Reiterer, H.: Applying Guidelines for the Design of Distortions on Focus+Context Interfaces, in: Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '16, pp. 244–247, ACM, New York, NY, USA, doi: 10.1145/2909132.2909284, URL http://doi.acm.org/10.1145/2909132.2909284, 2016.

[53] Butscher, S., Hubenschmid, S., Müller, J., Fuchs, J., and Reiterer, H.: Clusters, Trends, and Outliers: How Immersive Technologies Can Facilitate the Collaborative Analysis of Multidimensional Data, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, pp. 1–12, ACM Press, New York, New York, USA, doi: 10.1145/3173574.3173664, 2018.

[54] Butz, A., Höllerer, T., Feiner, S., MacIntyre, B., and Beshers, C.: Enveloping users and computers in a collaborative 3D augmented reality, in: Proceedings 2nd IEEE

and ACM International Workshop on Augmented Reality (IWAR'99), pp. 35–44, doi: 10.1109/IWAR.1999.803804, 1999.

[55] Cacioppo, J. T., Tassinary, L. G., and Berntson, G.: Handbook of psychophysiology, Cambridge university press, 2007.

[56] Caine, K.: Local Standards for Sample Size at CHI, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 981–992, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/2858036.2858498, URL https://doi.org/10.1145/2858036.2858498, 2016.

[57] Calandra, D. M., Caso, A., Cutugno, F., Origlia, A., and Rossi, S.: CoWME: A General Framework to Evaluate Cognitive Workload During Multimodal Interaction, in: Proceedings of the 15th ACM on International Conference on Multimodal Interaction, ICMI '13, pp. 111–118, ACM, Sydney, Australia, doi: 10.1145/2522848.2522867, 2013.

[58] Cao, Y., Theune, M., and Nijholt, A.: Modality Effects on Cognitive Load and Performance in High-Load Information Presentation, in: Proceedings of the 14th International Conference on Intelligent User Interfaces, IUI '09, pp. 335–344, ACM, New York, NY, USA, doi: 10.1145/1502650.1502697, 2009.

[59] Chen, D., Hart, J., and Vertegaal, R.: Towards a Physiological Model of User Interruptability, in: Human-Computer Interaction - INTERACT 2007, edited by Baranauskas, C., Palanque, P., Abascal, J., and Barbosa, S. D. J., vol. 4663, pp. 439–451, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-540-74800-7_39, URL http://link.springer.com/10.1007/978-3-540-74800-7_39, 2007.

[60] Chen, N., Guimbretière, F., and Sellen, A.: Designing a Multi-slate Reading Environment to Support Active Reading Activities, ACM Trans. Comput.-Hum. Interact., 19, 18:1–18:35, doi: 10.1145/2362364.2362366, URL http://doi.acm.org/10.1145/2362364.2362366, 2012.

[61] Chen, N., Guimbretière, F., and Sellen, A.: Graduate Student Use of a Multi-slate Reading System, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pp. 1799–1808, ACM, New York, NY, USA, doi: 10.1145/2470654.2466237, URL http://doi.acm.org/10.1145/2470654.2466237, 2013.

[62] Chen, R., Xie, T., Xie, Y., Lin, T., and Tang, N.: Do Speech Features for Detecting Cognitive Load Depend on Specific Languages?, in: Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI 2016, pp. 76–83, ACM, Tokyo, Japan, doi: 10.1145/2993148.2993149, 2016.

[63] Chen, S. and Epps, J.: Blinking: Toward Wearable Computing That Understands Your Current Task, IEEE Pervasive Computing, 12, 56–65, doi: 10.1109/MPRV.2013.45, 2013.

[64] Chen, S. and Epps, J.: Using Task-Induced Pupil Diameter and Blink Rate to Infer Cognitive Load, Human-Computer Interaction, 29, 390–413, doi: 10.1080/07370024.2014.892428, URL https://doi.org/10.1080/07370024.2014.892428, 2014.

[65] Chen, S., Epps, J., and Chen, F.: A Comparison of Four Methods for Cognitive Load Measurement, in: Proceedings of the 23rd Australian Computer-Human Interaction Conference, OzCHI '11, pp. 76–79, ACM, Canberra, Australia, doi: 10.1145/2071536.2071547, 2011.

[66] Chen, S., Epps, J., Ruiz, N., and Chen, F.: Eye Activity As a Measure of Human Mental Effort in HCI, in: Proceedings of the 16th International Conference on Intelligent User Interfaces, IUI '11, pp. 315–318, ACM, New York, NY, USA, doi: 10.1145/1943403.1943454, 2011.

[67] Chen, S., Epps, J., and Chen, F.: Automatic and Continuous User Task Analysis via Eye Activity, in: Proceedings of the 2013 International Conference on Intelligent User Interfaces, IUI '13, pp. 57–66, ACM, New York, NY, USA, doi: 10.1145/2449396.2449406, 2013.

[68] Chevalier, A. and Kicka, M.: Web Designers and Web Users: Influence of the Ergonomic Quality of the Web Site on the Information Search, International Journal of Human-Computer Studies, 64, 1031–1048, doi: https://doi.org/10.1016/j.ijhcs.2006.06.002, 2006.

[69] Chiossi, F., Zagermann, J., Karolus, J., Rodrigues, N., Balestrucci, P., Weiskopf, D., Ehinger, B., Feuchtner, T., Reiterer, H., Chuang, L. L., Ernst, M., Bulling, A., Mayer, S., and Schmidt, A.: Adapting visualizations and interfaces to the user, it - Information Technology, 64, 133–143, doi: doi:10.1515/itit-2022-0035, URL https://doi.org/10.1515/itit-2022-0035, 2022.

[70] Cinaz, B., Arnrich, B., La Marca, R., and Tröster, G.: Monitoring of mental workload levels during an everyday life office-work scenario, Personal and Ubiquitous Computing, 17, 229–239, doi: 10.1007/s00779-011-0466-1, URL https://doi.org/10.1007/s00779-011-0466-1, 2013.

[71] Cockburn, A., Karlson, A., and Bederson, B. B.: A Review of Overview+Detail, Zooming, and Focus+Context Interfaces, ACM Comput. Surv., 41, 2:1–2:31, doi: 10.1145/1456650.1456652, URL http://doi.acm.org/10.1145/1456650.1456652, 2009.

[72] Collins, J., Regenbrecht, H., Langlotz, T., Said Can, Y., Ersoy, C., and Butson, R.: Measuring Cognitive Load and Insight: A Methodology Exemplified in a Virtual Reality Learning Context, in: 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 351–362, 2019.

[73] Conway, D., Dick, I., Li, Z., Wang, Y., and Chen, F.: The Effect of Stress on Cognitive Load Measurement, in: Human-Computer Interaction - INTERACT 2013, edited by Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., and Winckler, M., vol. 8120, pp. 659–666, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-642-40498-6_58, URL http://link.springer.com/10.1007/978-3-642-40498-6_58, 2013.

[74] Cook, T. D., Campbell, D. T., and Shadish, W.: Experimental and quasi-experimental designs for generalized causal inference, vol. 1195, Houghton Mifflin Boston, MA, 2002.

[75] Crk, I., Kluthe, T., and Stefik, A.: Understanding Programming Expertise: An Empirical Study of Phasic Brain Wave Changes, ACM Trans. Comput.-Hum. Interact., 23, doi: 10.1145/2829945, 2016.

[76] Cutmore, T. R. H. and James, D. A.: Identifying and reducing noise in psychophysiological recordings, International Journal of Psychophysiology, 32, 129 – 150, doi:

https://doi.org/10.1016/S0167-8760(99)00014-8, URL http://www.sciencedirect.com/science/article/pii/S0167876099000148, 1999.

[77] Dabbish, L. and Kraut, R. E.: Controlling Interruptions: Awareness Displays and Social Motivation for Coordination, in: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, CSCW '04, pp. 182–191, ACM, New York, NY, USA, doi: 10.1145/1031607.1031638, 2004.

[78] de Santana, V. F., Ferreira, J. J., de Paula, R. A., and de Gusmão Cerqueira, R. F.: An Eye Gaze Model for Seismic Interpretation Support, in: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, ETRA '18, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3204493.3204554, 2018.

[79] Debie, E., Fernandez Rojas, R., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., and Abbass, H. A.: Multimodal Fusion for Objective Assessment of Cognitive Workload: A Review, IEEE Transactions on Cybernetics, 51, 1542–1555, doi: 10.1109/TCYB.2019.2939399, 2021.

[80] Debue, N. and van de Leemput, C.: What does germane load mean? An empirical contribution to the cognitive load theory, Frontiers in Psychology, 5, 1099, doi: 10.3389/fpsyg.2014.01099, URL https://www.frontiersin.org/article/10.3389/fpsyg.2014.01099, 2014.

[81] Demberg, V., Sayeed, A., Mahr, A., and Müller, C.: Measuring Linguistically-Induced Cognitive Load During Driving Using the ConTRe Task, in: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '13, pp. 176–183, ACM, Eindhoven, Netherlands, doi: 10.1145/2516540.2516546, 2013.

[82] Di Geronimo, L., Husmann, M., and Norrie, M. C.: Surveying Personal Device Ecosystems with Cross-device Applications in Mind, in: Proceedings of the 5th ACM International Symposium on Pervasive Displays, PerDis '16, pp. 220–227, ACM, New York, NY, USA, doi: 10.1145/2914920.2915028, URL http://doi.acm.org/10.1145/2914920.2915028, 2016.

[83] Ding, Y., Huynh, B., Xu, A., Bullock, T., Cecotti, H., Turk, M., Giesbrecht, B., and Höllerer, T.: Multimodal Classification of EEG during Physical Activity, in: 2019 International Conference on Multimodal Interaction, ICMI '19, pp. 185–194, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3340555.3353759, 2019.

[84] Dönmez, O., Akbulut, Y., Telli, E., Kaptan, M., Özdemir, İ. H., and Erdem, M.: In search of a measure to address different sources of cognitive load in computer-based learning environments, Education and Information Technologies, pp. 1–22, doi: 10.1007/s10639-022-11035-2, 2022.

[85] Doubé, W. and Beh, J.: Typing over Autocomplete: Cognitive Load in Website Use by Older Adults, in: Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI '12, pp. 97–106, ACM, Melbourne, Australia, doi: 10.1145/2414536.2414553, 2012.

[86] Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M., and Giannopoulos, I.: The Index of Pupillary Activity: Measuring Cognitive Load Vis-à-vis Task Difficulty with Pupil Oscillation, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 282:1–282:13, ACM, New

York, NY, USA, doi: 10.1145/3173574.3173856, URL http://doi.acm.org/10.1145/3173574.3173856, 2018.

[87] Duchowski, A. T., Krejtz, K., Gehrer, N. A., Bafna, T., and Bækgaard, P.: The Low/High Index of Pupillary Activity, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, pp. 1–12, Association for Computing Machinery, Honolulu, HI, USA, doi: 10.1145/3313831.3376394, 2020.

[88] Duran, R., Zavgorodniaia, A., and Sorva, J.: Cognitive Load Theory in Computing Education Research: A Review, ACM Trans. Comput. Educ., 22, doi: 10.1145/3483843, URL https://doi.org/10.1145/3483843, 2022.

[89] Ehrich, J. F., Howard, S. J., Bokosmaty, S., and Woodcock, S.: An Item Response Modeling Approach to Cognitive Load Measurement, Frontiers in Education, 6, doi: 10.3389/feduc.2021.648324, URL https://www.frontiersin.org/articles/10.3389/feduc.2021.648324, 2021.

[90] Elmqvist, N.: Distributed User Interfaces: State of the Art, pp. 1–12, Springer London, London, doi: 10.1007/978-1-4471-2271-5_1, URL https://doi.org/10.1007/978-1-4471-2271-5_1, 2011.

[91] Elmqvist, N.: Data Analytics Anywhere and Everywhere, Commun. ACM, 66, 52–63, doi: 10.1145/3584858, URL https://doi.org/10.1145/3584858, 2023.

[92] Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E.-J., and Shadlen, M. N.: fMRI of human visual cortex., Nature, 1994.

[93] Ens, B., Bach, B., Cordeil, M., Engelke, U., Serrano, M., Willett, W., Prouzeau, A., Anthes, C., Büschel, W., Dunne, C., Dwyer, T., Grubert, J., Haga, J. H., Kirshenbaum, N., Kobayashi, D., Lin, T., Olaosebikan, M., Pointecker, F., Saffo, D., Saquib, N., Schmalstieg, D., Szafir, D. A., Whitlock, M., and Yang, Y.: Grand Challenges in Immersive Analytics, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3411764.3446866, URL https://doi.org/10.1145/3411764.3446866, 2021.

[94] Evain, A., Argelaguet, F., Roussel, N., Casiez, G., and Lécuyer, A.: Can I Think of Something Else when Using a BCI?: Cognitive Demand of an SSVEP-based BCI, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 5120–5125, ACM, New York, NY, USA, doi: 10.1145/3025453.3026037, URL http://doi.acm.org/10.1145/3025453.3026037, 2017.

[95] Evans, D. C. and Fendley, M.: A Multi-Measure Approach for Connecting Cognitive Workload and Automation, International Journal of Human-Computer Studies, 97, 182–189, doi: https://doi.org/10.1016/j.ijhcs.2016.05.008, 2017.

[96] Fauconnier, G. and Turner, M.: The way we think: Conceptual blending and the mind's hidden complexities, Basic books, 2008.

[97] Feiner, S. and Shamash, A.: Hybrid User Interfaces: Breeding Virtually Bigger Interfaces for Physically Smaller Computers, in: Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology, UIST '91, pp. 9–17, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/120782.120783, URL https://doi.org/10.1145/120782.120783, 1991.

[98] Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A.: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment, Personal Technologies, 1, 208–217, 1997.

[99] Ferrari, M. and Quaresima, V.: A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application, NeuroImage, 63, 921–935, doi: https://doi.org/10.1016/j.neuroimage.2012.03.049, URL https://www.sciencedirect.com/science/article/pii/S1053811912003308, 2012.

[100] Fink, D. I., Zagermann, J., Reiterer, H., and Jetter, H.-C.: Re-locations: Augmenting Personal and Shared Workspaces to Support Remote Collaboration in Incongruent Spaces, Proc. ACM Hum.-Comput. Interact., 2022.

[101] Fitts, P. M.: The information capacity of the human motor system in controlling the amplitude of movement., Journal of experimental psychology, 47, 381, 1954.

[102] Frey, J., Mühl, C., Lotte, F., and Hachet, M.: Review of the Use of Electroencephalography as an Evaluation Method for Human-Computer Interaction, arXiv:1311.2222 [cs], URL http://arxiv.org/abs/1311.2222, arXiv: 1311.2222, 2013.

[103] Frey, J., Daniel, M., Castet, J., Hachet, M., and Lotte, F.: Framework for Electroencephalography-based Evaluation of User Experience, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 2283–2294, ACM, Santa Clara, California, USA, doi: 10.1145/2858036.2858525, URL http://doi.acm.org/10.1145/2858036.2858525, 2016.

[104] Fridman, L., Reimer, B., Mehler, B., and Freeman, W. T.: Cognitive Load Estimation in the Wild, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 652:1–652:9, ACM, New York, NY, USA, doi: 10.1145/3173574.3174226, URL http://doi.acm.org/10.1145/3173574.3174226, 2018.

[105] Friedman, B. H. and Thayer, J. F.: Facial muscle activity and EEG recordings: redundancy analysis, Electroencephalography and Clinical Neurophysiology, 79, 358 – 360, doi: https://doi.org/10.1016/0013-4694(91)90200-N, URL http://www.sciencedirect.com/science/article/pii/001346949190200N, 1991.

[106] Gabaude, C., Baracat, B., Jallais, C., Bonniaud, M., and Fort, A.: Cognitive load measurement while driving. In : Human Factors: a view from an integrative perspective, in: Cognitive load measurement while driving. In : Human Factors: a view from an integrative perspective, pp. pp. 67–80, HUMAN FACTORS AND ERGONOMICS SOCIETY, URL https://hal.archives-ouvertes.fr/hal-01027475, 2012.

[107] Galy, E., Cariou, M., and Melan, C.: What is the relationship between mental workload factors and cognitive load types?, International Journal of Psychophysiology, 83, 269–275, doi: https://doi.org/10.1016/j.ijpsycho.2011.09.023, URL https://www.sciencedirect.com/science/article/pii/S016787601100300X, 2012.

[108] Gaver, W. W.: The SonicFinder: An Interface That Uses Auditory Icons, Human-Computer Interaction, 4, 67–94, doi: 10.1207/s15327051hci0401_3, 1989.

[109] Georges, V., Courtemanche, F., Senecal, S., Baccino, T., Fredette, M., and Leger, P.-M.: UX Heatmaps: Mapping User Experience on Visual Interfaces, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 4850–4860, ACM, Santa Clara, California, USA, doi: 10.1145/2858036.2858271, URL http://doi.acm.org/10.1145/2858036.2858271, 2016.

[110] Gershenfeld, N., Krikorian, R., and Cohen, D.: The internet of things, Scientific American, 291, 76–81, 2004.

[111] Gevins, A., Smith, M. E., McEvoy, L., and Yu, D.: High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice, Cerebral Cortex, 7, 374–385, doi: 10.1093/cercor/7.4.374, 1997.

[112] Giakoumis, D., Tzovaras, D., and Hassapis, G.: Subject-Dependent Biosignal Features for Increased Accuracy in Psychological Stress Detection, International Journal of Human-Computer Studies, 71, 425–439, doi: https://doi.org/10.1016/j.ijhcs.2012.10.016, 2013.

[113] Gibbs Jr, R. W.: Embodiment and cognitive science, Cambridge University Press, 2005.

[114] Ginns, P.: Integrating Information: A Meta-Analysis of the Spatial Contiguity and Temporal Contiguity Effects, Learning and Instruction, 16, 511–525, doi: 10.1016/j.learninstruc.2006.10.001, 2006.

[115] Giraud, S., Thérouanne, P., and Steiner, D. D.: Web Accessibility: Filtering Redundant and Irrelevant Information Improves Website Usability for Blind Users, International Journal of Human-Computer Studies, 111, 23–35, doi: https://doi.org/10.1016/j.ijhcs.2017.10.011, 2018.

[116] Girouard, A., Solovey, E. T., Hirshfield, L. M., Chauncey, K., Sassaroli, A., Fantini, S., and Jacob, R. J. K.: Distinguishing Difficulty Levels with Non-invasive Brain Activity Measurements, in: Human-Computer Interaction - INTERACT 2009, edited by Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R. O., and Winckler, M., vol. 5726, pp. 440–452, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-642-03655-2_50, URL http://link.springer.com/10.1007/978-3-642-03655-2_50, 2009.

[117] Girouard, A., Shaer, O., Solovey, E. T., Poor, G. M., and Jacob, R. J.: The reality of Reality-Based interaction: understanding the impact of a framework as a research tool, ACM Transactions on Computer-Human Interaction (TOCHI), 26, 1–35, 2019.

[118] Glatz, C., Krupenia, S. S., Bülthoff, H. H., and Chuang, L. L.: Use the Right Sound for the Right Job: Verbal Commands and Auditory Icons for a Task-Management System Favor Different Information Processes in the Brain, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 472:1–472:13, ACM, New York, NY, USA, doi: 10.1145/3173574.3174046, URL http://doi.acm.org/10.1145/3173574.3174046, 2018.

[119] Griffiths, C., Bowen, J., and Hinze, A.: Investigating Wearable Technology for Fatigue Identification in the Workplace, in: Human-Computer Interaction - INTERACT 2017, edited by Bernhaupt, R., Dalvi, G., Joshi, A., K. Balkrishan, D., O'Neill, J., and Winckler, M., vol. 10514, pp. 370–380, Springer International Publishing, Cham, doi: 10.1007/978-3-319-67684-5_22, URL http://link.springer.com/10.1007/978-3-319-67684-5_22, 2017.

[120] Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P.: Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, pp. 835–844, ACM, Florence, Italy, doi: 10.1145/1357054.1357187, URL http://doi.acm.org/10.1145/1357054.1357187, 2008.

[121] Grinstein, G. G., O'Connell, T. A., Laskowski, S. J., Plaisant, C., Scholtz, J., and Whiting, M. A.: VAST 2006 Contest - A Tale of Alderwood, 2006 IEEE Symposium On Visual Analytics Science And Technology, pp. 215–216, 2006.

[122] Grubert, J. and Kranz, M.: HeadPhones: Ad Hoc Mobile Multi-Display Environments Through Head Tracking, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 3966–3971, ACM, New York, NY, USA, doi: 10.1145/3025453.3025533, URL http://doi.acm.org/10.1145/3025453.3025533, 2017.

[123] Grubert, J., Heinisch, M., Quigley, A., and Schmalstieg, D.: MultiFi: Multi Fidelity Interaction with Displays On and Around the Body, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 3933–3942, Association for Computing Machinery, Seoul, Republic of Korea, doi: 10.1145/2702123.2702331, 2015.

[124] Grudin, J.: From tool to partner: The evolution of human-computer interaction, Synthesis Lectures on Human-Centered Interaction, 10, i–183, 2017.

[125] Guo, W. and Wang, J.: Towards Attentive Speed Reading on Small Screen Wearable Devices, in: Proceedings of the 20th ACM International Conference on Multimodal Interaction, ICMI '18, pp. 278–287, ACM, New York, NY, USA, doi: 10.1145/3242969.3243009, URL http://doi.acm.org/10.1145/3242969.3243009, 2018.

[126] Gupta, S., Coles, T., Dumas, C., McBride, S. J., and Bradford, D.: Gamer Style: Performance Factors in Gamified Simulation, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 2014–2025, ACM, Santa Clara, California, USA, doi: 10.1145/2858036.2858461, URL http://doi.acm.org/10.1145/2858036.2858461, 2016.

[127] Gustafson, S., Baudisch, P., Gutwin, C., and Irani, P.: Wedge: Clutter-free Visualization of Off-screen Locations, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, pp. 787–796, ACM, New York, NY, USA, doi: 10.1145/1357054.1357179, URL http://doi.acm.org/10.1145/1357054.1357179, 2008.

[128] Gutwin, C. and Greenberg, S.: Design for Individuals, Design for Groups: Tradeoffs Between Power and Workspace Awareness, in: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, CSCW '98, pp. 207–216, ACM, New York, NY, USA, 1998.

[129] Haapalainen, E., Kim, S., Forlizzi, J. F., and Dey, A. K.: Psycho-Physiological Measures for Assessing Cognitive Load, in: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, UbiComp '10, pp. 301–310, ACM, Copenhagen, Denmark, doi: 10.1145/1864349.1864395, 2010.

[130] Hamilton, P. and Wigdor, D. J.: Conductor: Enabling and Understanding Cross-device Interaction, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, pp. 2773–2782, ACM, New York, NY, USA, 2014.

[131] Harper, R., Rodden, T., Rogers, Y., and Sellen, A.: Being Human: Human-Computer Interaction in the Year 2020, URL https://www.microsoft.com/en-us/research/publication/being-human-human-computer-interaction-in-the-year-2020/, 2008.

[132] Harrison, C., Wiese, J., and Dey, A. K.: Achieving ubiquity: The new third wave, IEEE MultiMedia, 17, 8–12, 2010.

[133] Hart, S., Childress, M., and Hauser, J.: Individual definitions of the term ”workload”, in: Eighth Symposium on Psychology in the Department of Defense, pp. 478–485, 1982.

[134] Hart, S. G.: Theory and measurement of human workload, Human productivity enhancement, 1, 396–456, 1986.

[135] Hart, S. G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50, 904–908, doi: 10.1177/154193120605000909, URL https://doi.org/10.1177/154193120605000909, 2006.

[136] Hart, S. G. and Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in: Human mental workload, Advances in psychology, 52, pp. 139–183, North-Holland, Oxford, England, doi: 10.1016/S0166-4115(08)62386-9, 1988.

[137] Hartson, R. and Pyla, P. S.: The UX book: Agile UX design for a quality user experience, Morgan Kaufmann, 2018.

[138] Heeman, P. A., Meshorer, T., Kun, A. L., Palinko, O., and Medenica, Z.: Estimating Cognitive Load Using Pupil Diameter During a Spoken Dialogue Task, in: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI ’13, pp. 242–245, ACM, Eindhoven, Netherlands, doi: 10.1145/2516540.2516570, 2013.

[139] Herculano-Houzel, S.: The human brain in numbers: a linearly scaled-up primate brain, Frontiers in Human Neuroscience, 3, doi: 10.3389/neuro.09.031.2009, URL https://www.frontiersin.org/articles/10.3389/neuro.09.031.2009/full, 2009.

[140] Hertzum, M. and Holmegaard, K. D.: Perceived Time as a Measure of Mental Workload: Effects of Time Constraints and Task Success, International Journal of Human-Computer Interaction, 29, 26–39, doi: 10.1080/10447318.2012.676538, URL http://www.tandfonline.com/doi/abs/10.1080/10447318.2012.676538, 2013.

[141] Hihn, H., Meudt, S., and Schwenker, F.: Inferring Mental Overload Based on Postural Behavior and Gestures, in: Proceedings of the 2nd Workshop on Emotion Representations and Modelling for Companion Systems, ERM4CT ’16, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3009960.3009961, 2016.

[142] Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Jacob, R. J., Sassaroli, A., and Fantini, S.: Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload with Functional Near Infrared Spectroscopy, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’09, pp. 2185–2194, ACM, Boston, MA, USA, doi: 10.1145/1518701.1519035, URL http://doi.acm.org/10.1145/1518701.1519035, 2009.

[143] Hirshfield, L. M., Gulotta, R., Hirshfield, S., Hincks, S., Russell, M., Ward, R., Williams, T., and Jacob, R.: This is Your Brain on Interfaces: Enhancing Usability Testing with Functional Near-infrared Spectroscopy, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’11, pp. 373–382, ACM, Vancouver, BC, Canada, doi: 10.1145/1978942.1978996, URL http://doi.acm.org/10.1145/1978942.1978996, 2011.

[144] Hollan, J., Hutchins, E., and Kirsh, D.: Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research, ACM Trans. Comput.-Hum. Interact.,

7, 174–196, doi: 10.1145/353485.353487, URL https://doi.org/10.1145/353485.353487, 2000.

[145] Hollender, N., Hofmann, C., Deneke, M., and Schmitz, B.: Integrating cognitive load theory and concepts of human-computer interaction, Computers in Human Behavior, 26, 1278–1288, doi: 10.1016/j.chb.2010.05.031, URL http://www.sciencedirect.com/science/article/pii/S0747563210001718, 2010.

[146] Homaeian, L., Goyal, N., Wallace, J. R., and Scott, S. D.: Group vs Individual: Impact of TOUCH and TILT Cross-Device Interactions on Mixed-Focus Collaboration, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 73:1–73:13, ACM, New York, NY, USA, doi: 10.1145/3173574.3173647, URL http://doi.acm.org/10.1145/3173574.3173647, 2018.

[147] Horak, T., Mathisen, A., Klokmose, C. N., Dachselt, R., and Elmqvist, N.: Vistribute: Distributing Interactive Visualizations in Dynamic Multi-Device Setups, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, pp. 616:1–616:13, ACM, New York, NY, USA, doi: 10.1145/3290605.3300846, URL http://doi.acm.org/10.1145/3290605.3300846, 2019.

[148] Hornbæk, K.: Current Practice in Measuring Usability: Challenges to Usability Studies and Research, Int. J. Hum.-Comput. Stud., 64, 79–102, 2006.

[149] Hornbæk, K.: Some whys and hows of experiments in human–computer interaction, Foundations and Trends® in Human–Computer Interaction, 5, 299–373, 2013.

[150] Hornbæk, K. and Law, E. L.-C.: Meta-analysis of Correlations Among Usability Measures, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07, pp. 617–626, ACM, New York, NY, USA, doi: 10.1145/1240624.1240722, URL http://doi.acm.org/10.1145/1240624.1240722, 2007.

[151] Hornbæk, K., Bederson, B. B., and Plaisant, C.: Navigation Patterns and Usability of Zoomable User Interfaces with and Without an Overview, ACM Trans. Comput.-Hum. Interact., 9, 362–389, doi: 10.1145/586081.586086, URL http://doi.acm.org/10.1145/586081.586086, 2002.

[152] Hornecker, E., Marshall, P., Dalton, N. S., and Rogers, Y.: Collaboration and Interference: Awareness with Mice or Touch Input, in: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW '08, pp. 167–176, ACM, New York, NY, USA, 2008.

[153] Houben, S., Marquardt, N., Vermeulen, J., Klokmose, C., Schöning, J., Reiterer, H., and Holz, C.: Opportunities and Challenges for Cross-device Interactions in the Wild, interactions, 24, 58–63, doi: 10.1145/3121348, URL http://doi.acm.org/10.1145/3121348, 2017.

[154] Hubenschmid, S., Zagermann, J., Butscher, S., and Reiterer, H.: Employing Tangible Visualisations in Augmented Reality with Mobile Devices, in: AVI'18: 2018 International Conference on Advanced Visual Interfaces, 2018.

[155] Hubenschmid, S., Zagermann, J., Butscher, S., and Reiterer, H.: STREAM: Exploring the Combination of Spatially-Aware Tablets with Augmented Reality Head-Mounted Displays for Immersive Analytics, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–14, 2021.

[156] Hubenschmid, S., Zagermann, J., Fink, D., Wieland, J., Feuchtner, T., and Reiterer, H.: Towards Asynchronous Hybrid User Interfaces for Cross-Reality Interaction, in: ISS'21 Workshop Proceedings: "Transitional Interfaces in Mixed and Cross-Reality: A new frontier?", edited by Jetter, H.-C., Schröder, J.-H., Gugenheimer, J., Billinghurst, M., Anthes, C., Khamis, M., and Feuchtner, T., 2021.

[157] Hubenschmid, S., Wieland, J., Fink, D. I., Batch, A., Zagermann, J., Elmqvist, N., and Reiterer, H.: ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3491102.3517550, URL https://doi.org/10.1145/3491102.3517550, 2022.

[158] Hubenschmid, S., Zagermann, J., Dachselt, R., Elmqvist, N., Feiner, S., Feuchtner, T., Lee, B., Reiterer, H., and Schmalstieg, D.: Hybrid User Interfaces: Complementary Interfaces for Mixed Reality Interaction, in: Workshop at IEEE ISMAR'23, 2023.

[159] Hubenschmid, S., Zagermann, J., Leicht, D., Reiterer, H., and Feuchtner, T.: ARound the Smartphone: Investigating the Effects of Virtually-Extended Display Size on Spatial Memory, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), ACM, New York, NY, USA, doi: https://doi.org/10.1145/3544548.3581438, 2023.

[160] Hussain, M. S., Calvo, R. A., and Chen, F.: Automatic Cognitive Load Detection from Face, Physiology, Task Performance and Fusion During Affective Interference, Interacting with Computers, 26, 256–268, doi: 10.1093/iwc/iwt032, URL https://doi.org/10.1093/iwc/iwt032, 2013.

[161] Hwang, Y., Lee, S., Jeon, H. S., Park, J. H. Y., Lee, K. W., and Lee, J.: "Eat What You Want and Be Healthy!": Comfort Food Effects: Human-Food Interaction in View of Celebratory Technology, in: Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction, MHFI'18, pp. 4:1–4:8, ACM, New York, NY, USA, doi: 10.1145/3279954.3279958, URL http://doi.acm.org/10.1145/3279954.3279958, 2018.

[162] Ikehara, C. S. and Crosby, M. E.: Assessing cognitive load with physiological sensors, in: Proceedings of the 38th annual hawaii international conference on system sciences, pp. 295a–295a, IEEE, 2005.

[163] Ikematsu, K. and Siio, I.: Memory Stones: An Intuitive Information Transfer Technique Between Multi-touch Computers, in: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, HotMobile '15, pp. 3–8, ACM, New York, NY, USA, doi: 10.1145/2699343.2699352, URL http://doi.acm.org/10.1145/2699343.2699352, 2015.

[164] Institution, B. S.: Test charts for clinical determination of distance visual acuity. 2003. London.

[165] Intarasirisawat, J., Ang, C. S., Efstratiou, C., Dickens, L. W. F., and Page, R.: Exploring the Touch and Motion Features in Game-Based Cognitive Assessments, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 3, doi: 10.1145/3351245, 2019.

[166] Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., and Bailey, B. P.: Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI

'05, pp. 311–320, ACM, Portland, Oregon, USA, doi: 10.1145/1054972.1055016, URL http://doi.acm.org/10.1145/1054972.1055016, 2005.

[167] Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K.-L., and Hagen, H.: Collaborative visualization: Definition, challenges, and research agenda, Information Visualization, 10, 310–326, doi: 10.1177/1473871611412817, URL https://doi.org/10.1177/1473871611412817, 2011.

[168] Isenberg, P., Fisher, D., Paul, S. A., Morris, M. R., Inkpen, K., and Czerwinski, M.: Co-Located Collaborative Visual Analytics around a Tabletop Display, IEEE Trans. Vis. Comput. Graph., 18, 689–702, 2012.

[169] Ishii, H.: Tangible Bits: Designing the Seamless Interface between People, Bits, and Atoms, in: Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI '03, p. 3, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/604045.604048, URL https://doi.org/10.1145/604045.604048, 2003.

[170] Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., and Chance, B.: Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks, International Journal of Human-Computer Interaction, 17, 211–227, doi: 10.1207/s15327590ijhc1702_6, URL http://www.tandfonline.com/doi/abs/10.1207/s15327590ijhc1702_6, 2004.

[171] Jacko, J. A., Barreto, A., Scott, I. U., Rosa, R. H., and Pappas, C. J.: Using Electroencephalogram to Investigate Stages of Visual Search in Visually Impaired Computer Users: Preattention and Focal Attention, International Journal of Human-Computer Interaction, 12, 135–150, doi: 10.1207/S15327590IJHC1201_6, URL http://www.tandfonline.com/doi/abs/10.1207/S15327590IJHC1201_6, 2000.

[172] Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J.: Reality-based interaction: a framework for post-WIMP interfaces, in: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 201–210, 2008.

[173] Jaeger, S., Klein, K., Joos, L., Zagermann, J., de Ridder, M., Kim, J., Yang, J., Pfeil, U., Reiterer, H., and Schreiber, F.: Challenges for Brain Data Analysis in VR Environments, in: 2019 IEEE Pacific Visualization Symposium (PacificVis), pp. 42–46, doi: 10.1109/PacificVis.2019.00013, 2019.

[174] Jakobsen, M. R. and Hornbæk, K.: Up Close and Personal: Collaborative Work on a High-resolution Multitouch Wall Display, ACM Trans. Comput.-Hum. Interact., 21, 11:1–11:34, 2014.

[175] Jakobsen, M. R. and Hornbæk, K.: Is Moving Improving?: Some Effects of Locomotion in Wall-Display Interaction, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 4169–4178, ACM, New York, NY, USA, doi: 10.1145/2702123.2702312, URL http://doi.acm.org/10.1145/2702123.2702312, 2015.

[176] Jakobsen, M. R., Jansen, Y., Boring, S., and Hornbæk, K.: Should I Stay or Should I Go? Selecting Between Touch and Mid-Air Gestures for Large-Display Interaction, pp. 455–473, Springer International Publishing, Cham, doi: 10.1007/978-3-319-22698-9_31, URL http://dx.doi.org/10.1007/978-3-319-22698-9_31, 2015.

[177] Jarvis, J., Putze, F., Heger, D., and Schultz, T.: Multimodal Person Independent Recognition of Workload Related Biosignal Patterns, in: Proceedings of the 13th Interna-

tional Conference on Multimodal Interfaces, ICMI '11, pp. 205–208, ACM, Alicante, Spain, doi: 10.1145/2070481.2070516, 2011.

[178] Jetter, H.-C., Gerken, J., Zöllner, M., Reiterer, H., and Milic-Frayling, N.: Materializing the Query with Facet-streams: A Hybrid Surface for Collaborative Search on Tabletops, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pp. 3013–3022, ACM, New York, NY, USA, 2011.

[179] Jetter, H.-C., Leifert, S., Gerken, J., Schubert, S., and Reiterer, H.: Does (Multi-)Touch Aid Users' Spatial Memory and Navigation in 'Panning' and in 'Zooming & Panning' UIs?, in: Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '12, pp. 83–90, ACM, New York, NY, USA, doi: 10.1145/2254556.2254575, URL http://doi.acm.org/10.1145/2254556.2254575, 2012.

[180] Jetter, H.-C., Reiterer, H., and Geyer, F.: Blended Interaction: understanding natural human–computer interaction in post-WIMP interactive spaces, Personal and Ubiquitous Computing, 18, 1139–1158, 2014.

[181] Jiang, X., Atkins, M. S., Tien, G., Bednarik, R., and Zheng, B.: Pupil Responses During Discrete Goal-directed Movements, in: Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pp. 2075–2084, ACM, Toronto, Ontario, Canada, doi: 10.1145/2556288.2557086, URL http://doi.acm.org/10.1145/2556288.2557086, 2014.

[182] Jiang, X., Zheng, B., Bednarik, R., and Atkins, M. S.: Pupil Responses to Continuous Aiming Movements, International Journal of Human-Computer Studies, 83, 1–11, doi: https://doi.org/10.1016/j.ijhcs.2015.05.006, 2015.

[183] Johansen, R.: Groupware: Computer support for business teams, The Free Press, 1988.

[184] Jokela, T., Ojala, J., Grassel, G., Piippo, P., and Olsson, T.: A Comparison of Methods to Move Visual Objects Between Personal Mobile Devices in Different Contexts of Use, in: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '15, pp. 172–181, ACM, New York, NY, USA, doi: 10.1145/2785830.2785841, URL http://doi.acm.org/10.1145/2785830.2785841, 2015.

[185] Kalyuga, S.: Cognitive load theory: How many types of load does it really need?, Educational Psychology Review, 23, 1–19, doi: 10.1007/s10648-010-9150-7, 2011.

[186] Karolus, J., Wozniak, P. W., Chuang, L. L., and Schmidt, A.: Robust Gaze Features for Enabling Language Proficiency Awareness, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 2998–3010, Association for Computing Machinery, Denver, Colorado, USA, doi: 10.1145/3025453.3025601, 2017.

[187] Karolus, J., Bachmann, F., Kosch, T., Schmidt, A., and Woźniak, P. W.: Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity, in: Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction, MobileHCI '21, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3447526.3472027, URL https://doi.org/10.1145/3447526.3472027, 2021.

[188] Karolus, J., Thanheiser, S., Peterson, D., Viot, N., Kosch, T., Schmidt, A., and Wozniak, P. W.: Imprecise but Fun: Playful Interaction Using Electromyography, Proc.

ACM Hum.-Comput. Interact., 6, doi: 10.1145/3546725, URL https://doi.org/10.1145/3546725, 2022.

[189] Katidioti, I.: Interrupted by Your Pupil: An Interruption Management System Based on Pupil Dilation, p. 12.

[190] Kaufmann, B. and Ahlström, D.: Studying Spatial Memory and Map Navigation Performance on Projector Phones with Peephole Interaction, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pp. 3173–3176, ACM, New York, NY, USA, doi: 10.1145/2470654.2466434, URL http://doi.acm.org/10.1145/2470654.2466434, 2013.

[191] Kelleher, C. and Hnin, W.: Predicting Cognitive Load in Future Code Puzzles, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, pp. 1–12, Association for Computing Machinery, Glasgow, Scotland Uk, doi: 10.1145/3290605.3300487, 2019.

[192] Khalfa, S., BELLA, S. D., Roy, M., Peretz, I., and Lupien, S. J.: Effects of relaxing music on salivary cortisol level after psychological stress, Annals of the New York Academy of Sciences, 999, 374–376, 2003.

[193] Khawaja, M. A., Ruiz, N., and Chen, F.: Potential Speech Features for Cognitive Load Measurement, in: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, OZCHI '07, pp. 57–60, ACM, Adelaide, Australia, doi: 10.1145/1324892.1324902, 2007.

[194] Khawaja, M. A., Ruiz, N., and Chen, F.: Think Before You Talk: An Empirical Study of Relationship Between Speech Pauses and Cognitive Load, in: Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat, OZCHI '08, pp. 335–338, ACM, Cairns, Australia, doi: 10.1145/1517744.1517814, 2008.

[195] Khawaja, M. A., Chen, F., and Marcus, N.: Using Language Complexity to Measure Cognitive Load for Adaptive Interaction Design, in: Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI '10, pp. 333–336, ACM, New York, NY, USA, doi: 10.1145/1719970.1720024, 2010.

[196] Khawaja, M. A., Chen, F., and Marcus, N.: Measuring Cognitive Load Using Linguistic Features: Implications for Usability Evaluation and Adaptive Interaction Design, International Journal of Human-Computer Interaction, 30, 343–368, doi: 10.1080/10447318.2013.860579, URL https://doi.org/10.1080/10447318.2013.860579, 2014.

[197] Khawaji, A., Chen, F., Zhou, J., and Marcus, N.: Trust and Cognitive Load in the Text-Chat Environment: The Role of Mouse Movement, in: Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design, OzCHI '14, pp. 324–327, ACM, Sydney, New South Wales, Australia, doi: 10.1145/2686612.2686661, 2014.

[198] Kirschbaum, C. and Hellhammer, D. H.: Salivary cortisol in psychobiological research: an overview, Neuropsychobiology, 22, 150–169, doi: 10.1159/000118611, 1989.

[199] Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., and Linkman, S.: Systematic literature reviews in software engineering - A systematic literature review, Information and Software Technology, 51, 7–15, doi: 10.1016/j.infsof.2008.09.009, URL http://www.sciencedirect.com/science/article/pii/S0950584908001390, 2009.

[200] Klimesch, W., Schimke, H., and Pfurtscheller, G.: Alpha frequency, cognitive load and memory performance, Brain topography, 5, 241–251, 1993.

[201] Klinkhammer, D., Tennie, J. O., Erdoes, P., and Reiterer, H.: Body Panning: A Movement-based Navigation Technique for Large Interactive Surfaces, in: Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, ITS '13, pp. 37–40, ACM, New York, NY, USA, doi: 10.1145/2512349.2512822, URL http://doi.acm.org/10.1145/2512349.2512822, 2013.

[202] Klokmose, C. N., Eagan, J. R., Baader, S., Mackay, W., and Beaudouin-Lafon, M.: Webstrates: Shareable Dynamic Media, in: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, UIST '15, pp. 280–290, ACM, New York, NY, USA, doi: 10.1145/2807442.2807446, URL http://doi.acm.org/10.1145/2807442.2807446, 2015.

[203] Knierim, P., Kosch, T., Groschopp, J., and Schmidt, A.: Opportunities and Challenges of Text Input in Portable Virtual Reality, in: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20, ACM, New York, NY, USA, doi: 10.1145/3334480.3382920, 2020.

[204] Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., and Zarjam, P.: Measuring Cognitive Workload with Low-Cost Electroencephalograph, in: Human-Computer Interaction - INTERACT 2011, edited by Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., and Winckler, M., vol. 6949, pp. 568–571, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-642-23768-3_84, URL http://link.springer.com/10.1007/978-3-642-23768-3_84, 2011.

[205] Knyazev, G. G.: EEG delta oscillations as a correlate of basic homeostatic and motivational processes, Neuroscience & Biobehavioral Reviews, 36, 677–695, doi: https://doi.org/10.1016/j.neubiorev.2011.10.002, URL https://www.sciencedirect.com/science/article/pii/S0149763411001849, 2012.

[206] Korbach, A., Brünken, R., and Park, B.: Differentiating different types of cognitive load: A comparison of different measures, Educational Psychology Review, 30, 503–529, doi: 10.1007/s10648-017-9404-8, 2018.

[207] Korshakov, A., Frolov, A., and Bobrov, P.: On-line automatic suppression of artifacts in multi-dimensional signals using ICA, in: Abstr. 10th Europ. Conf. on Non-Destructive Testing, p. 370, 2010.

[208] Kosch, T.: Workload-Aware Systems and Interfaces for Cognitive Augmentation, 2020.

[209] Kosch, T. and Chuang, L. L.: Investigating the Influence of RSVP Display Parameters on Working Memory Load using Electroencephalography, in: 2nd International Conference on Neuroadaptive Technology, 2019.

[210] Kosch, T., Hassib, M., and Schmidt, A.: The Brain Matters: A 3D Real-Time Visualization to Examine Brain Source Activation Leveraging Neurofeedback, in: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16, pp. 1570–1576, ACM, New York, NY, USA, doi: 10.1145/2851581.2892484, URL http://doi.acm.org/10.1145/2851581.2892484, 2016.

[211] Kosch, T., Funk, M., Schmidt, A., and Chuang, L. L.: Identifying Cognitive Assistance with Mobile Electroencephalography: A Case Study with In-Situ Projections

for Manual Assembly, Proceedings of the ACM on Human-Computer Interaction, 2, 1–20, doi: 10.1145/3229093, URL http://dl.acm.org/citation.cfm?doid=3233739.3229093, 2018.

[212] Kosch, T., Hassib, M., Buschek, D., and Schmidt, A.: Look into My Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation, in: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, CHI EA '18, pp. LBW617:1–LBW617:6, ACM, New York, NY, USA, doi: 10.1145/3170427.3188643, URL http://doi.acm.org/10.1145/3170427.3188643, 2018.

[213] Kosch, T., Hassib, M., Woźniak, P., Buschek, D., and Alt, F.: Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 436:1–436:13, ACM, New York, NY, USA, doi: 10.1145/3173574.3174010, URL http://doi.acm.org/10.1145/3173574.3174010, 2018.

[214] Kosch, T., Karolus, J., Ha, H., and Schmidt, A.: Your Skin Resists: Exploring Electrodermal Activity as Workload Indicator during Manual Assembly, in: Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, EICS '19, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3319499.3328230, 2019.

[215] Kosch, T., Wennrich, K., Topp, D., Muntzinger, M., and Schmidt, A.: The Digital Cooking Coach: Using Visual and Auditory In-situ Instructions to Assist Cognitively Impaired During Cooking, in: Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '19, pp. 156–163, ACM, New York, NY, USA, doi: 10.1145/3316782.3321524, URL http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/3316782.3321524, 2019.

[216] Kosch, T., Schmidt, A., Thanheiser, S., and Chuang, L. L.: One Does Not Simply RSVP: Mental Workload to Select Speed Reading Parameters Using Electroencephalography, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, pp. 1–13, Association for Computing Machinery, Honolulu, HI, USA, doi: 10.1145/3313831.3376766, 2020.

[217] Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., and Woźniak, P. W.: A Survey on Measuring Cognitive Workload in Human-Computer Interaction, ACM Comput. Surv., doi: 10.1145/3582272, URL https://doi.org/10.1145/3582272, 2023.

[218] Kosch, T., Welsch, R., Chuang, L., and Schmidt, A.: The Placebo Effect of Artificial Intelligence in Human-Computer Interaction, ACM Trans. Comput.-Hum. Interact., 29, doi: 10.1145/3529225, URL https://doi.org/10.1145/3529225, 2023.

[219] Kosti, M. V., Georgiadis, K., Adamos, D. A., Laskaris, N., Spinellis, D., and Angelis, L.: Towards an Affordable Brain Computer Interface for the Assessment of Programmers' Mental Workload, International Journal of Human-Computer Studies, 115, 52–66, doi: https://doi.org/10.1016/j.ijhcs.2018.03.002, 2018.

[220] Krakauer, J. W. and Shadmehr, R.: Consolidation of motor memory, Trends in Neurosciences, 29, 58–64, doi: 10.1016/j.tins.2005.10.003, 2006.

[221] Kujala, T.: Efficiency of Visual Time-Sharing Behavior: The Effects of Menu Structure on POI Search Tasks While Driving, in: Proceedings of the 1st International Confer-

ence on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '09, pp. 63–70, ACM, Essen, Germany, doi: 10.1145/1620509.1620522, 2009.

[222] Kumar, N. and Kumar, J.: Measurement of Cognitive Load in HCI Systems Using EEG Power Spectrum: An Experimental Study, Procedia Computer Science, 84, 70–78, doi: https://doi.org/10.1016/j.procs.2016.04.068, URL https://www.sciencedirect.com/science/article/pii/S1877050916300825, proceeding of the Seventh International Conference on Intelligent Human Computer Interaction (IHCI 2015), 2016.

[223] Kun, A. L., Shyrokov, A., and Heeman, P. A.: Spoken Tasks for Human-Human Experiments: Towards in-Car Speech User Interfaces for Multi-Threaded Dialogue, in: Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '10, pp. 57–63, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/1969773.1969784, URL https://doi.org/10.1145/1969773.1969784, 2010.

[224] Kun, A. L., Palinko, O., and Razumenić, I.: Exploring the Effects of Size and Luminance of Visual Targets on the Pupillary Light Reflex, in: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '12, pp. 183–186, ACM, Portsmouth, New Hampshire, doi: 10.1145/2390256.2390287, 2012.

[225] La Vaque, T.: The history of EEG hans berger: psychophysiologist. A historical vignette, Journal of Neurotherapy, 3, 1–9, 1999.

[226] Langley, J. N.: The Autonomic Nervous System. Part 1 Part 1, W. Heffer, Cambridge [England, 1921.

[227] Large, D. R., Burnett, G., Anyasodo, B., and Skrypchuk, L.: Assessing Cognitive Demand During Natural Language Interactions with a Digital Driving Assistant, in: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Automotive'UI 16, pp. 67–74, ACM, Ann Arbor, MI, USA, doi: 10.1145/3003715.3005408, 2016.

[228] Laugwitz, B., Held, T., and Schrepp, M.: Construction and evaluation of a user experience questionnaire, in: Symposium of the Austrian HCI and Usability Engineering Group, pp. 63–76, Springer, 2008.

[229] LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., and Poupyrev, I. P.: 3D user interfaces: theory and practice, Addison-Wesley Professional, 2017.

[230] Lazar, J., Feng, J. H., and Hochheiser, H.: Research methods in human-computer interaction, Morgan Kaufmann, 2017.

[231] Lee, B., Chung, K., and Kim, S.-H.: Interruption Cost Evaluation by Cognitive Workload and Task Performance in Interruption Coordination Modes for Human-Computer Interaction Tasks, Applied Sciences, 8, 1780, doi: 10.3390/app8101780, URL http://www.mdpi.com/2076-3417/8/10/1780, 2018.

[232] Lee, J. C. and Tan, D. S.: Using a low-cost electroencephalograph for task classification in HCI research, in: Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06, p. 81, ACM Press, Montreux, Switzerland, doi: 10.1145/1166253.1166268, URL http://dl.acm.org/citation.cfm?doid=1166253.1166268, 2006.

[233] Leifert, S.: The Influence of Grids on Spatial and Content Memory, in: CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, pp. 941–946, ACM, New York, NY, USA, doi: 10.1145/1979742.1979522, URL http://doi.acm.org/10.1145/1979742.1979522, 2011.

[234] LeMay, M. and Hird, E.: Operator Work Load: When is Enough Enough?, Commun. ACM, 29, 638–642, doi: 10.1145/6138.6147, URL https://doi.org/10.1145/6138.6147, 1986.

[235] Leslie, G. and Mullen, T.: MoodMixer : EEG-Based Collaborative Sonification, in: Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 296–299, Oslo, Norway, 2011.

[236] Liang, C., Karolus, J., Kosch, T., and Schmidt, A.: On the Suitability of Real-Time Assessment of Programming Proficiency Using Gaze Properties, in: Proceedings of the 7th ACM International Symposium on Pervasive Displays, PerDis '18, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3205873.3210702, URL https://doi.org/10.1145/3205873.3210702, 2018.

[237] Lin, T., Imamiya, A., and Mao, X.: Using multiple data sources to get closer insights into user cost and task performance, Interacting with Computers, 20, 364–374, doi: 10.1016/j.intcom.2007.12.002, URL https://doi.org/10.1016/j.intcom.2007.12.002, 2008.

[238] Lin, Y., Zhang, W., and Watson, L.: Using Eye Movement Parameters for Evaluating Human–machine Interface Frameworks under Normal Control Operation and Fault Detection Situations, International Journal of Human-Computer Studies, 59, 837–873, doi: https://doi.org/10.1016/S1071-5819(03)00122-8, 2003.

[239] Lindlbauer, D., Feit, A. M., and Hilliges, O.: Context-Aware Online Adaptation of Mixed Reality Interfaces, in: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, UIST '19, pp. 147–160, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3332165.3347945, 2019.

[240] Liu, C., Chapuis, O., Beaudouin-Lafon, M., Lecolinet, E., and Mackay, W. E.: Effects of Display Size and Navigation Type on a Classification Task, in: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pp. 4147–4156, ACM, New York, NY, USA, doi: 10.1145/2556288.2557020, URL http://doi.acm.org/10.1145/2556288.2557020, 2014.

[241] Liu, C., Chapuis, O., Beaudouin-Lafon, M., and Lecolinet, E.: Shared Interaction on a Wall-Sized Display in a Data Manipulation Task, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 2075–2086, ACM, New York, NY, USA, doi: 10.1145/2858036.2858039, URL http://doi.acm.org/10.1145/2858036.2858039, 2016.

[242] Liu, Z., Nersessian, N., and Stasko, J.: Distributed cognition as a theoretical framework for information visualization, IEEE transactions on visualization and computer graphics, 14, 1173–1180, 2008.

[243] Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B.: A review of classification algorithms for EEG-based brain–computer interfaces, doi: 10.1088/1741-2560/4/2/r01, URL https://doi.org/10.1088/1741-2560/4/2/r01, 2007.

[244] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F.: A review of classification algorithms for EEG-based brain–computer inter-

faces: a 10 year update, doi: 10.1088/1741-2552/aab2f2, URL https://doi.org/10.1088/1741-2552/aab2f2, 2018.

[245] Lu, S., Zhang, M. Y., Ersal, T., and Yang, X. J.: Workload Management in Teleoperation of Unmanned Ground Vehicles: Effects of a Delay Compensation Aid on Human Operators' Workload and Teleoperation Performance, International Journal of Human-Computer Interaction, 35, 1820–1830, doi: 10.1080/10447318.2019.1574059, URL https://doi.org/10.1080/10447318.2019.1574059, 2019.

[246] Lucero, A., Holopainen, J., and Jokela, T.: Pass-them-around: Collaborative Use of Mobile Phones for Photo Sharing, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pp. 1787–1796, ACM, New York, NY, USA, 2011.

[247] Luck, S. J.: An introduction to the event-related potential technique, MIT press, 2014.

[248] Lukanov, K., Maior, H. A., and Wilson, M. L.: Using fNIRS in Usability Testing: Understanding the Effect of Web Form Layout on Mental Workload, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 4011–4016, ACM, Santa Clara, California, USA, doi: 10.1145/2858036.2858236, URL http://doi.acm.org/10.1145/2858036.2858236, 2016.

[249] Lystbæk, M. N., Rosenberg, P., Pfeuffer, K., Grønbæk, J. E., and Gellersen, H.: Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality, Proc. ACM Hum.-Comput. Interact., 6, doi: 10.1145/3530886, URL https://doi.org/10.1145/3530886, 2022.

[250] Lyu, Y., Luo, X., Zhou, J., Yu, C., Miao, C., Wang, T., Shi, Y., and Kameyama, K.-i.: Measuring Photoplethysmogram-Based Stress-Induced Vascular Response Index to Assess Cognitive Load and Stress, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 857–866, ACM, Seoul, Republic of Korea, doi: 10.1145/2702123.2702399, URL http://doi.acm.org/10.1145/2702123.2702399, 2015.

[251] Mackay, W. E. and Fayard, A.-L.: HCI, natural science and design: a framework for triangulation across disciplines, in: Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques, pp. 223–234, 1997.

[252] MacKenzie, I. S.: Fitts' Law as a Research and Design Tool in Human-Computer Interaction, Hum.-Comput. Interact., 7, 91–139, doi: 10.1207/s15327051hci0701_3, URL https://doi.org/10.1207/s15327051hci0701_3, 1992.

[253] Maior, H. A., Wilson, M. L., and Sharples, S.: Workload Alerts - Using Physiological Measures of Mental Workload to Provide Feedback During Tasks, ACM Trans. Comput.-Hum. Interact., 25, 9:1–9:30, doi: 10.1145/3173380, 2018.

[254] Malik, M. and Camm, A. J.: Heart rate variability, Clinical Cardiology, 13, 570–576, doi: 10.1002/clc.4960130811, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/clc.4960130811, 1990.

[255] Malmivuo, P., Malmivuo, J., and Plonsey, R.: Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields, Oxford University Press, google-Books-ID: H9CFM0TqWwsC, 1995.

[256] Marquardt, N., Ballendat, T., Boring, S., Greenberg, S., and Hinckley, K.: Gradual Engagement: Facilitating Information Exchange Between Digital Devices As a Function of Proximity, in: Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, ITS '12, pp. 31–40, ACM, New York, NY, USA, doi: 10.1145/2396636.2396642, URL http://doi.acm.org/10.1145/2396636.2396642, 2012.

[257] Marquardt, N., Hinckley, K., and Greenberg, S.: Cross-device Interaction via Micro-mobility and F-formations, in: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12, pp. 13–22, ACM, New York, NY, USA, doi: 10.1145/2380116.2380121, URL http://doi.acm.org/10.1145/2380116.2380121, 2012.

[258] Marshall, S. P.: The index of cognitive activity: Measuring cognitive workload, in: Proceedings of the IEEE 7th conference on Human Factors and Power Plants, pp. 7–7, IEEE, 2002.

[259] Marwecki, S., Wilson, A. D., Ofek, E., Gonzalez Franco, M., and Holz, C.: Mise-Unseen: Using Eye Tracking to Hide Virtual Reality Scene Changes in Plain Sight, in: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, UIST '19, pp. 777–789, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3332165.3347919, 2019.

[260] Mason, S. G. and Birch, G. E.: A general framework for brain-computer interface design, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11, 70–85, doi: 10.1109/TNSRE.2003.810426, 2003.

[261] Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., Huggings, J., and Mervielde, I.: Validation of a comprehensive stress state questionnaire: Towards a state "Big Three?", 7, 335–350, 1999.

[262] Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C., and Warm, J. S.: Profiling task stress with the dundee stress state questionnaire, Psychology of stress: New research, 1, 49–90, 2013.

[263] McDuff, D. J., Hernandez, J., Gontarek, S., and Picard, R. W.: COGCAM: Contact-free Measurement of Cognitive Stress During Computer Tasks with a Digital Camera, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 4000–4004, ACM, Santa Clara, California, USA, doi: 10.1145/2858036.2858247, URL http://doi.acm.org/10.1145/2858036.2858247, 2016.

[264] McGrath, W., Bowman, B., McCallum, D., Hincapié-Ramos, J. D., Elmqvist, N., and Irani, P.: Branch-explore-merge: Facilitating Real-time Revision Control in Collaborative Visual Exploration, in: Proc. of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, ITS '12, pp. 235–244, ACM, New York, NY, USA, 2012.

[265] Melamed, S., Shirom, A., Toker, S., Berliner, S., and Shapira, I.: Burnout and risk of cardiovascular disease: Evidence, possible causal paths, and promising research directions., Psychological bulletin, 132, 327, doi: https://psycnet.apa.org/doi/10.1037/0033-2909.132.3.327, 2006.

[266] Merletti, R. and Farina, D.: Surface Electromyography: Physiology, Engineering and Applications, IEEE Press Series on Biomedical Engineering, John Wiley & Sons, 2016.

[267] Miyake, S.: Multivariate workload evaluation combining physiological and subjective measures, International Journal of Psychophysiology, 40, 233–238, doi: https:

//doi.org/10.1016/S0167-8760(00)00191-4, URL https://www.sciencedirect.com/science/article/pii/S0167876000001914, psychophysiology in, 2001.

[268] Mock, P., Gerjets, P., Tibus, M., Trautwein, U., Möller, K., and Rosenstiel, W.: Using Touchscreen Interaction Data to Predict Cognitive Workload, in: Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI 2016, pp. 349–356, ACM, Tokyo, Japan, doi: 10.1145/2993148.2993202, 2016.

[269] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and PRISMA Group: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS medicine, 6, e1000 097, doi: 10.1371/journal.pmed.1000097, 2009.

[270] Morris, M. R., Lombardo, J., and Wigdor, D.: WeSearch: Supporting Collaborative Search and Sensemaking on a Tabletop Display, in: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW '10, pp. 401–410, ACM, New York, NY, USA, 2010.

[271] Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., schraefel, m. c., and Wobbrock, J. O.: Reducing Legacy Bias in Gesture Elicitation Studies, interactions, 21, 40–45, doi: 10.1145/2591689, URL http://doi.acm.org/10.1145/2591689, 2014.

[272] Mosaly, P. R., Guo, H., and Mazur, L.: Toward Better Understanding of Task Difficulty during Physicians' Interaction with Electronic Health Record System (EHRs), International Journal of Human-Computer Interaction, 35, 1883–1891, doi: 10.1080/10447318.2019.1575081, URL https://doi.org/10.1080/10447318.2019.1575081, 2019.

[273] Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T., and Cauwenberghs, G.: Real-time neuroimaging and cognitive monitoring using wearable dry EEG, doi: 10.1109/TBME.2015.2481482, 2015.

[274] Müller, J., Rädle, R., Jetter, H.-C., and Reiterer, H.: An Experimental Comparison of Vertical and Horizontal Dynamic Peephole Navigation, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 1523–1526, ACM, New York, NY, USA, doi: 10.1145/2702123.2702227, URL http://doi.acm.org/10.1145/2702123.2702227, 2015.

[275] Müller, J., Rädle, R., and Reiterer, H.: Virtual Objects As Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 1245–1249, ACM, New York, NY, USA, doi: 10.1145/2858036.2858043, URL http://doi.acm.org/10.1145/2858036.2858043, 2016.

[276] Müller, J., Rädle, R., and Reiterer, H.: Remote Collaboration With Mixed Reality Displays: How Shared Virtual Landmarks Facilitate Spatial Referencing, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 6481–6486, ACM, New York, NY, USA, doi: 10.1145/3025453.3025717, URL https://doi.org/10.1145/3025453.3025717, 2017.

[277] Müller, J., Zagermann, J., Wieland, J., Pfeil, U., and Reiterer, H.: A qualitative comparison between augmented and virtual reality collaboration with handheld devices, in: Proceedings of Mensch und Computer 2019, pp. 399–410, 2019.

[278] Nacenta, M. A., Gutwin, C., Aliakseyeu, D., and Subramanian, S.: There and back again: Cross-display object movement in multi-display environments, Human–Computer Interaction, 24, 170–229, 2009.

[279] Naseer, N. and Hong, K.-S.: fNIRS-based brain-computer interfaces: a review, Frontiers in Human Neuroscience, 9, 3, doi: 10.3389/fnhum.2015.00003, URL https://www.frontiersin.org/article/10.3389/fnhum.2015.00003, 2015.

[280] Nater, U. M., Skoluda, N., and Strahler, J.: Biomarkers of stress in behavioural medicine:, Current Opinion in Psychiatry, 26, 440–445, doi: 10.1097/YCO.0b013e328363b4ed, URL http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001504-201309000-00004, 2013.

[281] Neerincx, M.: Cognitive task load analysis: allocating tasks and designing support, Handbook of Cognitive Task Design. Chapter 13. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 283–305, 2003.

[282] Neumayr, T., Jetter, H.-C., Augstein, M., Friedl, J., and Luger, T.: Domino: A Descriptive Framework for Hybrid Collaboration and Coupling Styles in Partially Distributed Teams, Proc. ACM Hum.-Comput. Interact., 2, doi: 10.1145/3274397, URL https://doi.org/10.1145/3274397, 2018.

[283] Newman, J., Bornik, A., Pustka, D., Echtler, F., Huber, M., Schmalstieg, D., and Klinker, G.: Tracking for distributed mixed reality environments, in: Workshop on Trands and Issues in Tracking for Virtual Environments at the IEEE Virtual Reality Conference (VR'07), 2007.

[284] Nicolas-Alonso, L. F. and Gomez-Gil, J.: Brain Computer Interfaces, a Review, Sensors, 12, 1211–1279, doi: 10.3390/s120201211, URL https://www.mdpi.com/1424-8220/12/2/1211, 2012.

[285] Nielsen, J.: Enhancing the Explanatory Power of Usability Heuristics, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '94, pp. 152–158, ACM, New York, NY, USA, doi: 10.1145/191666.191729, URL http://doi.acm.org/10.1145/191666.191729, 1994.

[286] Nielsen, J. and Levy, J.: Measuring Usability: Preference vs. Performance, Commun. ACM, 37, 66–75, doi: 10.1145/175276.175282, URL http://doi.acm.org/10.1145/175276.175282, 1994.

[287] Nikulin, V. V., Nolte, G., and Curio, G.: A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition, NeuroImage, 55, 1528–1535, doi: 10.1016/j.neuroimage.2011.01.057, URL http://www.sciencedirect.com/science/article/pii/S1053811911000930, 2011.

[288] Nourbakhsh, N., Wang, Y., Chen, F., and Calvo, R. A.: Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks, in: Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI '12, pp. 420–423, ACM, Melbourne, Australia, doi: 10.1145/2414536.2414602, 2012.

[289] Nourbakhsh, N., Wang, Y., and Chen, F.: GSR and Blink Features for Cognitive Load Classification, in: Human-Computer Interaction - INTERACT 2013, edited by Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., and Winckler, M., vol. 8117, pp. 159–166, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-642-40483-2_11, URL http://link.springer.com/10.1007/978-3-642-40483-2_11, 2013.

[290] Novak, D., Mihelj, M., and Munih, M.: Using Psychophysiological Measurements in Physically Demanding Virtual Environments, in: Human-Computer Interaction - INTERACT 2009, edited by Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R. O., and Winckler, M., vol. 5726, pp. 490–493, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-642-03655-2_55, URL http://link.springer.com/10.1007/978-3-642-03655-2_55, 2009.

[291] Novak, D., Beyeler, B., Omlin, X., and Riener, R.: Workload Estimation in Physical Human-Robot Interaction Using Physiological Measurements, Interacting with Computers, 27, 616–629, doi: 10.1093/iwc/iwu021, URL https://doi.org/10.1093/iwc/iwu021, 2014.

[292] O'Hara, K., Sellen, A., and Bentley, R.: Supporting Memory for Spatial Location While Reading from Small Displays, in: CHI '99 Extended Abstracts on Human Factors in Computing Systems, CHI EA '99, pp. 220–221, ACM, New York, NY, USA, 1999.

[293] Oleksik, G., Jetter, H.-C., Gerken, J., Milic-Frayling, N., and Jones, R.: Towards an Information Architecture for Flexible Reuse of Digital Media, in: Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia, MUM '13, pp. 12:1–12:10, ACM, New York, NY, USA, 2013.

[294] Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P. L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P. A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S. B., Kohli, P., Lutchyn, Y., Keskin, C., and Izadi, S.: Holoportation: Virtual 3D Teleportation in Real-Time, in: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16, pp. 741–754, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/2984511.2984517, URL https://doi.org/10.1145/2984511.2984517, 2016.

[295] Oulasvirta, A.: When Users "Do" the Ubicomp, interactions, 15, 6–9, doi: 10.1145/1340961.1340963, URL http://doi.acm.org/10.1145/1340961.1340963, 2008.

[296] Oulasvirta, A. and Hornbæk, K.: Counterfactual thinking: What theories do in design, International Journal of Human–Computer Interaction, 38, 78–92, 2022.

[297] Ouwehand, K., Kroef, A. v. d., Wong, J., and Paas, F.: Measuring Cognitive Load: Are There More Valid Alternatives to Likert Rating Scales?, Frontiers in Education, 6, doi: 10.3389/feduc.2021.702616, URL https://www.frontiersin.org/articles/10.3389/feduc.2021.702616, 2021.

[298] Oviatt, S.: Ten Myths of Multimodal Interaction, Commun. ACM, 42, 74–81, doi: 10.1145/319382.319398, URL https://doi.org/10.1145/319382.319398, 1999.

[299] Oviatt, S. and Cohen, P.: Perceptual User Interfaces: Multimodal Interfaces That Process What Comes Naturally, Commun. ACM, 43, 45–53, doi: 10.1145/330534.330538, URL https://doi.org/10.1145/330534.330538, 2000.

[300] Oviatt, S., Lunsford, R., and Coulston, R.: Individual Differences in Multimodal Integration Patterns: What Are They and Why Do They Exist?, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05, pp. 241–249, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/1054972.1055006, URL https://doi.org/10.1145/1054972.1055006, 2005.

[301] Paay, J., Raptis, D., Kjeldskov, J., Skov, M. B., Ruder, E. V., and Lauridsen, B. M.: Investigating Cross-Device Interaction Between a Handheld Device and a Large Display, in:

Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 6608–6619, ACM, New York, NY, USA, doi: 10.1145/3025453.3025724, URL http://doi.acm.org/10.1145/3025453.3025724, 2017.

[302] Palinko, O. and Kun, A. L.: Exploring the Effects of Visual Cognitive Load and Illumination on Pupil Diameter in Driving Simulators, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA '12, pp. 413–416, ACM, New York, NY, USA, doi: 10.1145/2168556.2168650, 2012.

[303] Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P.: Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator, in: Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications, ETRA '10, pp. 141–144, ACM, New York, NY, USA, doi: 10.1145/1743666.1743701, 2010.

[304] Palleis, H. and Hussmann, H.: Indirect 2D Touch Panning: How Does It Affect Spatial Memory and Navigation Performance?, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 1947–1951, ACM, New York, NY, USA, doi: 10.1145/2858036.2858334, URL http://doi.acm.org/10.1145/2858036.2858334, 2016.

[305] Partala, T. and Surakka, V.: Pupil Size Variation as an Indication of Affective Processing, International Journal of Human-Computer Studies, 59, 185–198, doi: https://doi.org/10.1016/S1071-5819(03)00017-X, applications of Affective Computing in Human-Computer Interaction, 2003.

[306] Pauzie, A.: A method to assess the driver mental workload: The driving activity load index (DALI), IET Intelligent Transport Systems, 2, 315–322, doi: 10.1049/iet-its:20080023, 2008.

[307] Peck, E. M., Carlin, E., and Jacob, R.: Designing Brain-Computer Interfaces for Attention-Aware Systems, Computer, 48, 34–42, doi: 10.1109/MC.2015.315, 2015.

[308] Peck, E. M. M., Yuksel, B. F., Ottley, A., Jacob, R. J., and Chang, R.: Using fNIRS Brain Sensing to Evaluate Information Visualization Interfaces, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pp. 473–482, ACM, New York, NY, USA, doi: 10.1145/2470654.2470723, 2013.

[309] Peck, T. C., Good, J. J., and Bourne, K. A.: Inducing and Mitigating Stereotype Threat through Gendered Virtual Body-Swap Illusions, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, pp. 1–13, Association for Computing Machinery, Honolulu, HI, USA, doi: 10.1145/3313831.3376419, 2020.

[310] Pejović, V., Gjoreski, M., Anderson, C., David, K., and Luštrek, M.: Toward Cognitive Load Inference for Attention Management in Ubiquitous Systems, IEEE Pervasive Computing, 19, 35–45, 2020.

[311] Pfleging, B., Fekety, D. K., Schmidt, A., and Kun, A. L.: A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 5776–5788, ACM, Santa Clara, California, USA, doi: 10.1145/2858036.2858117, URL http://doi.acm.org/10.1145/2858036.2858117, 2016.

[312] Pike, M. F., Maior, H. A., Porcheron, M., Sharples, S. C., and Wilson, M. L.: Measuring the Effect of Think Aloud Protocols on Workload Using fNIRS, in: Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pp.

3807–3816, ACM, Toronto, Ontario, Canada, doi: 10.1145/2556288.2556974, URL http://doi.acm.org/10.1145/2556288.2556974, 2014.

[313] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team: nlme: Linear and Nonlinear Mixed Effects Models, URL http://CRAN.R-project.org/package=nlme, r package version 3.1-122, 2015.

[314] Pirolli, P. and Card, S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis, in: Proceedings of international conference on intelligence analysis, vol. 5, pp. 2–4, McLean, VA, USA, 2005.

[315] Plaisant, C., Grinstein, G., Scholtz, J., Whiting, M., and et al.: Evaluating Visual Analytics at the 2007 VAST Symposium Contest, IEEE Comput. Graph. Appl., 28, 12–21, 2008.

[316] Plank, T., Jetter, H.-C., Rädle, R., Klokmose, C. N., Luger, T., and Reiterer, H.: Is Two Enough?! Studying Benefits, Barriers, and Biases of Multi-Tablet Use for Collaborative Visualization, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 4548–4560, ACM, New York, NY, USA, doi: 10.1145/3025453.3025537, URL http://doi.acm.org/10.1145/3025453.3025537, 2017.

[317] Platt, J. R.: Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others., science, 146, 347–353, 1964.

[318] Pollmann, K., Stefani, O., Bengsch, A., Peissner, M., and Vukelić, M.: How to Work in the Car of the Future? A Neuroergonomical Study Assessing Concentration, Performance and Workload Based on Subjective, Behavioral and Neurophysiological Insights, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, pp. 1–14, Association for Computing Machinery, Glasgow, Scotland Uk, doi: 10.1145/3290605.3300284, 2019.

[319] Probst, K.: Peripheral Interaction in Desktop Computing: Why It's Worth Stepping Beyond Traditional Mouse and Keyboard, pp. 183–205, Springer International Publishing, Cham, doi: 10.1007/978-3-319-29523-7_9, URL http://dx.doi.org/10.1007/978-3-319-29523-7_9, 2016.

[320] Putze, F. and Schultz, T.: Investigating Intrusiveness of Workload Adaptation, in: Proceedings of the 16th International Conference on Multimodal Interaction, ICMI '14, pp. 275–281, ACM, Istanbul, Turkey, doi: 10.1145/2663204.2663279, 2014.

[321] Rädle, R., Jetter, H.-C., Butscher, S., and Reiterer, H.: The Effect of Egocentric Body Movements on Users' Navigation Performance and Spatial Memory in Zoomable User Interfaces, in: Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, ITS '13, pp. 23–32, ACM, New York, NY, USA, doi: 10.1145/2512349.2512811, URL http://doi.acm.org/10.1145/2512349.2512811, 2013.

[322] Rädle, R., Jetter, H.-C., and Reiterer, H.: TwisterSearch: A Distributed User Interface for Collaborative Web Search, in: Distributed User Interfaces: Usability and Collaboration, edited by Lozano, M. D., Gallud, J. A., Tesoriero, R., and Penichet, V. M., Human-Computer Interaction Series, pp. 53–67, Springer London, 2013.

[323] Rädle, R., Jetter, H.-C., Marquardt, N., Reiterer, H., and Rogers, Y.: HuddleLamp: Spatially-Aware Mobile Displays for Ad-hoc Around-the-Table Collaboration, in: Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, ITS '14, pp. 45–54, ACM, New York, NY, USA, 2014.

[324] Rädle, R., Jetter, H.-C., Müller, J., and Reiterer, H.: Bigger is Not Always Better: Display Size, Performance, and Task Load During Peephole Map Navigation, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, pp. 4127–4136, ACM, New York, NY, USA, doi: 10.1145/2556288.2557071, URL http://doi.acm.org/10.1145/2556288.2557071, 2014.

[325] Rädle, R., Jetter, H.-C., Schreiner, M., Lu, Z., Reiterer, H., and Rogers, Y.: Spatially-Aware or Spatially-Agnostic? Elicitation and Evaluation of User-Defined Cross-Device Interactions, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 3913–3922, ACM, New York, NY, USA, doi: 10.1145/2702123.2702287, URL https://doi.org/10.1145/2702123.2702287, 2015.

[326] Rädle, R., Nouwens, M., Antonsen, K., Eagan, J. R., and Klokmose, C. N.: Codestrates: Literate computing with webstrates, in: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 715–725, 2017.

[327] Rädle, R., Jetter, H.-C., Fischer, J., Gabriel, I., Klokmose, C. N., Reiterer, H., and Holz, C.: PolarTrack: Optical Outside-In Device Tracking That Exploits Display Polarization, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 497:1–497:9, ACM, New York, NY, USA, doi: 10.1145/3173574.3174071, URL http://doi.acm.org/10.1145/3173574.3174071, 2018.

[328] Rafiqi, S., Nair, S., and Fernandez, E.: Cognitive and Context-Aware Applications, in: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '14, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/2674396.2674445, 2014.

[329] Rafiqi, S., Wangwiwattana, C., Kim, J., Fernandez, E., Nair, S., and Larson, E. C.: PupilWare: Towards Pervasive Cognitive Load Measurement Using Commodity Devices, in: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '15, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/2769493.2769506, 2015.

[330] Rajan, R., Selker, T., and Lane, I.: Task Load Estimation and Mediation Using Psycho-Physiological Measures, in: Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16, pp. 48–59, ACM, New York, NY, USA, doi: 10.1145/2856767.2856769, 2016.

[331] Rashid, U., Nacenta, M. A., and Quigley, A.: Factors Influencing Visual Attention Switch in Multi-Display User Interfaces: A Survey, in: Proceedings of the 2012 International Symposium on Pervasive Displays - PerDis '12, pp. 1–6, ACM, Porto, Portugal, doi: 10.1145/2307798.2307799, 2012.

[332] Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., and Tan, C.: An On-Road Assessment of the Impact of Cognitive Workload on Physiological Arousal in Young Adult Drivers, in: Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '09, pp. 115–118, ACM, Essen, Germany, doi: 10.1145/1620509.1620531, 2009.

[333] Rekimoto, J.: Pick-and-drop: A Direct Manipulation Technique for Multiple Computer Environments, in: Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology, UIST '97, pp. 31–39, ACM, New York, NY, USA, doi: 10.1145/263407.263505, URL http://doi.acm.org/10.1145/263407.263505, 1997.

[334] Robert J. Hockey, G.: Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework, Biological Psychology, 45, 73–93, doi: 10.1016/S0301-0511(96)05223-4, URL http://www.sciencedirect.com/science/article/pii/S0301051196052234, 1997.

[335] Rogers, Y. and Lindley, S.: Collaborating around vertical and horizontal large interactive displays: which way is best?, Interacting with Computers, 16, 1133–1152, 2004.

[336] Rohani Ghahari, R., George-Palilonis, J., Gahangir, H., Kaser, L. N., and Bolchini, D.: Semi-aural Interfaces: Investigating Voice-controlled Aural Flows, Interacting with Computers, 28, 826–842, doi: 10.1093/iwc/iww004, URL https://doi.org/10.1093/iwc/iww004, 2016.

[337] Roscoe, A. H. and Ellis, G. A.: A Subjective Rating Scale for Assessing Pilot Workload in Flight: A decade of Practical Use, p. 18.

[338] Roy, R. N., Charbonnier, S., Campagne, A., and Bonnet, S.: Efficient mental workload estimation using task-independent EEG features, doi: 10.1088/1741-2560/13/2/026019, URL https://doi.org/10.1088/1741-2560/13/2/026019, 2016.

[339] Rozado, D. and D́unser, A.: Combining EEG with Pupillometry to Improve Cognitive Workload Detection, Computer, 48, 18–25, doi: 10.1109/MC.2015.314, 2015.

[340] Rudmann, D. S., McConkie, G. W., and Zheng, X. S.: Eyetracking in Cognitive State Detection for HCI, in: Proceedings of the 5th International Conference on Multimodal Interfaces, ICMI '03, pp. 159–163, ACM, New York, NY, USA, 2003.

[341] Ruiz, N., Taib, R., Shi, Y. D., Choi, E., and Chen, F.: Using Pen Input Features As Indices of Cognitive Load, in: Proceedings of the 9th International Conference on Multimodal Interfaces, ICMI '07, pp. 315–318, ACM, Nagoya, Aichi, Japan, doi: 10.1145/1322192.1322246, 2007.

[342] Ruiz, N., Feng, Q. Q., Taib, R., Handke, T., and Chen, F.: Cognitive Skills Learning: Pen Input Patterns in Computer-Based Athlete Training, in: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI '10, pp. 41:1–41:4, ACM, Beijing, China, doi: 10.1145/1891903.1891955, 2010.

[343] Ruiz, N., Taib, R., and Chen, F.: Freeform Pen-Input As Evidence of Cognitive Load and Expertise, in: Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI '11, pp. 185–188, ACM, Alicante, Spain, doi: 10.1145/2070481.2070511, 2011.

[344] Ryall, K., Forlines, C., Shen, C., and Morris, M. R.: Exploring the Effects of Group Size and Table Size on Interactions with Tabletop Shared-display Groupware, in: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, CSCW '04, pp. 284–293, ACM, New York, NY, USA, 2004.

[345] Saffo, D., Batch, A., Dunne, C., and Elmqvist, N.: Through Their Eyes and In Their Shoes: Providing Group Awareness During Collaboration Across Virtual Reality and Desktop Platforms, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3544548.3581093, URL https://doi.org/10.1145/3544548.3581093, 2023.

[346] Sauseng, P., Klimesch, W., Schabus, M., and Doppelmayr, M.: Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory, International Journal of Psychophysiology, 57, 97–103, doi: https://doi.org/10.1016/j.ijpsycho.2005.03.018, URL https://www.sciencedirect.com/science/article/pii/S0167876005000851, eEG Coherence, 2005.

[347] Sayers, B.: Analysis of Heart Rate Variability, Ergonomics, 16, 17–32, doi: 10.1080/00140137308924479, 1973.

[348] Scarr, J., Cockburn, A., and Gutwin, C.: Supporting and exploiting spatial memory in user interfaces, Interaction, 6, 1–84, 2012.

[349] Schapkin, S. A. and Weißbecker-Klaus, X.: Test Battery for Assessment of Cognitive Function in Older Employees: Performance, Brain Processes, and Cardiovascular "Costs", in: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '15, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/2769493.2769553, 2015.

[350] Schartmüller, C., Weigl, K., Wintersberger, P., Riener, A., and Steinhauser, M.: Text Comprehension: Heads-Up vs. Auditory Displays: Implications for a Productive Work Environment in SAE Level 3 Automated Vehicles, in: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '19, pp. 342–354, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3342197.3344547, URL https://doi.org/10.1145/3342197.3344547, 2019.

[351] Schaule, F., Johanssen, J. O., Bruegge, B., and Loftness, V.: Employing Consumer Wearables to Detect Office Workers' Cognitive Load for Interruption Management, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2, doi: 10.1145/3191764, URL https://doi.org/10.1145/3191764, 2018.

[352] Schmidt, A.: Implicit human computer interaction through context, Personal technologies, 4, 191–199, 2000.

[353] Schmidt, A.: Ubiquitous computing-computing in context, Lancaster University (United Kingdom), 2003.

[354] Schmidt, A.: Ubiquitous computing: Are we there yet?, Computer, 43, 95–97, 2010.

[355] Schmidt, A., Beigl, M., and Gellersen, H.-W.: There is more to context than location, Computers & Graphics, 23, 893–901, doi: https://doi.org/10.1016/S0097-8493(99)00120-X, URL https://www.sciencedirect.com/science/article/pii/S009784939900120X, 1999.

[356] Schneegass, S., Pfleging, B., Broy, N., Heinrich, F., and Schmidt, A.: A Data Set of Real World Driving to Assess Driver Workload, in: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '13, pp. 150–157, ACM, Eindhoven, Netherlands, doi: 10.1145/2516540.2516561, 2013.

[357] Schnotz, W. and Kürschner, C.: A reconsideration of cognitive load theory, Educational psychology review, 19, 469–508, doi: 10.1007/s10648-007-9053-4, 2007.

[358] Schreiner, M., Rädle, R., Jetter, H.-C., and Reiterer, H.: Connichiwa: A Framework for Cross-Device Web Applications, in: Proceedings of the 33rd Annual ACM Conference

Extended Abstracts on Human Factors in Computing Systems, CHI EA '15, pp. 2163–2168, ACM, New York, NY, USA, 2015.

[359] Schröder, J.-H., Schacht, D., Peper, N., Hamurculu, A. M., and Jetter, H.-C.: Collaborating Across Realities: Analytical Lenses for Understanding Dyadic Collaboration in Transitional Interfaces, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3544548.3580879, URL https://doi.org/10.1145/3544548.3580879, 2023.

[360] Scott, S. D., Carpendale, M. S. T., and Inkpen, K. M.: Territoriality in Collaborative Tabletop Workspaces, in: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, CSCW '04, pp. 294–303, ACM, New York, NY, USA, 2004.

[361] Scott, S. D., Besacier, G., and McClelland, P. J.: Cross-device transfer in a collaborative multi-surface environment without user identification, in: 2014 International Conference on Collaboration Technologies and Systems (CTS), pp. 219–226, doi: 10.1109/CTS.2014.6867568, 2014.

[362] Seppelt, B., Seaman, S., Angell, L., Mehler, B., and Reimer, B.: Differentiating Cognitive Load Using a Modified Version of AttenD, in: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '17, pp. 114–122, ACM, Oldenburg, Germany, doi: 10.1145/3122986.3123019, 2017.

[363] Shapiro, L.: Embodied cognition, Routledge, 2019.

[364] Sjölie, D., Bodin, K., Elgh, E., Eriksson, J., Janlert, L.-E., and Nyberg, L.: Effects of Interactivity and 3D-motion on Mental Rotation Brain Activity in an Immersive Virtual Environment, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, pp. 869–878, ACM, Atlanta, Georgia, USA, doi: 10.1145/1753326.1753454, URL http://doi.acm.org/10.1145/1753326.1753454, 2010.

[365] Skarbez, R., Smith, M., and Whitton, M. C.: Revisiting Milgram and Kishino's Reality-Virtuality Continuum, Frontiers in Virtual Reality, 2, doi: 10.3389/frvir.2021.647997, URL https://www.frontiersin.org/article/10.3389/frvir.2021.647997, 2021.

[366] Skopik, A. and Gutwin, C.: Improving Revisitation in Fisheye Views with Visit Wear, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05, pp. 771–780, ACM, New York, NY, USA, doi: 10.1145/1054972.1055079, URL http://doi.acm.org/10.1145/1054972.1055079, 2005.

[367] Skov, M. B., Kjeldskov, J., Paay, J., Jensen, H. P., and Olsen, M. P.: Investigating Cross-Device Interaction Techniques: A Case of Card Playing on Handhelds and Tablets, in: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, OzCHI '15, pp. 446–454, ACM, New York, NY, USA, doi: 10.1145/2838739.2838763, URL http://doi.acm.org/10.1145/2838739.2838763, 2015.

[368] Solovey, E., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., and Jacob, R.: Brainput: Enhancing Interactive Systems with Streaming Fnirs Brain Input, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, pp. 2193–2202, Association for Computing Machinery, Austin, Texas, USA, doi: 10.1145/2207676.2208372, 2012.

[369] Solovey, E. T., Lalooses, F., Chauncey, K., Weaver, D., Parasi, M., Scheutz, M., Sassaroli, A., Fantini, S., Schermerhorn, P., Girouard, A., and Jacob, R. J.: Sensing Cognitive Multitasking for a Brain-Based Adaptive User Interface, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pp. 383–392, Association for Computing Machinery, Vancouver, BC, Canada, doi: 10.1145/1978942.1978997, 2011.

[370] Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., and Mehler, B.: Classifying Driver Workload Using Physiological and Driving Performance Data: Two Field Studies, in: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pp. 4057–4066, ACM, Toronto, Ontario, Canada, doi: 10.1145/2556288.2557068, URL http://doi.acm.org/10.1145/2556288.2557068, 2014.

[371] Sridhar, P. K., Chan, S. W. T., and Nanayakkara, S.: Going Beyond Performance Scores: Understanding Cognitive-affective States in Kindergarteners, in: Proceedings of the 17th ACM Conference on Interaction Design and Children, IDC '18, pp. 253–265, ACM, New York, NY, USA, doi: 10.1145/3202185.3202739, URL http://doi.acm.org/10.1145/3202185.3202739, event-place: Trondheim, Norway, 2018.

[372] Srivastava, N., Velloso, E., Lodge, J. M., Erfani, S., and Bailey, J.: Continuous Evaluation of Video Lectures from Real-Time Difficulty Self-Report, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, pp. 1–12, Association for Computing Machinery, Glasgow, Scotland Uk, doi: 10.1145/3290605.3300816, 2019.

[373] St. John, M., Kobus, D. A., Morrison, J. G., and Schmorrow, D.: Overview of the DARPA Augmented Cognition Technical Integration Experiment, International Journal of Human-Computer Interaction, 17, 131–149, doi: 10.1207/s15327590ijhc1702_2, URL http://www.tandfonline.com/doi/abs/10.1207/s15327590ijhc1702_2, 2004.

[374] Stanford, G. and Stanford, B. D.: Learning discussion skills through games, New York : Citation Press, 75 pages, 1969.

[375] Steichen, B., Carenini, G., and Conati, C.: User-Adaptive Information Visualization: Using Eye Gaze Data to Infer Visualization Tasks and User Cognitive Abilities, in: Proceedings of the 2013 International Conference on Intelligent User Interfaces, IUI '13, pp. 317–328, ACM, New York, NY, USA, doi: 10.1145/2449396.2449439, 2013.

[376] Stephan, K. E. and Roebroeck, A.: A short history of causal modeling of fMRI data, NeuroImage, 62, 856–863, doi: https://doi.org/10.1016/j.neuroimage.2012.01.034, URL https://www.sciencedirect.com/science/article/pii/S1053811912000511, 20 YEARS OF fMRI, 2012.

[377] Strimbu, K. and Tavel, J. A.: What are biomarkers?, Current opinion in HIV and AIDS, 5, 463–466, doi: 10.1097/COH.0b013e32833ed177, URL https://pubmed.ncbi.nlm.nih.gov/20978388, 2010.

[378] Sweller, J.: Cognitive Load During Problem Solving: Effects on Learning, Cognitive Science, 12, 257–285, doi: 10.1207/s15516709cog1202_4, URL https://onlinelibrary-wiley-com.emedien.ub.uni-muenchen.de/doi/abs/10.1207/s15516709cog1202_4, 1988.

[379] Sweller, J.: Evolutionary Bases of Human Cognitive Architecture: Implications for Computing Education, in: Proceedings of the Fourth International Workshop on

Computing Education Research, ICER '08, pp. 1–2, ACM, Sydney, Australia, doi: 10.1145/1404520.1404521, 2008.

[380] Sweller, J.: Cognitive Load Theory and Computer Science Education, in: Proceedings of the 47th ACM Technical Symposium on Computing Science Education, SIGCSE '16, pp. 1–1, ACM, Memphis, Tennessee, USA, doi: 10.1145/2839509.2844549, 2016.

[381] Sweller, J., Van Merrienboer, J. J., and Paas, F. G.: Cognitive architecture and instructional design, Educational psychology review, 10, 251–296, doi: 10.1023/A:1022193728205, 1998.

[382] Szafir, D. and Mutlu, B.: Pay Attention! Designing Adaptive Agents That Monitor and Improve User Engagement, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, pp. 11–20, Association for Computing Machinery, Austin, Texas, USA, doi: 10.1145/2207676.2207679, 2012.

[383] Takahashi, K., Nakayama, M., and Shimizu, Y.: The Response of Eye-Movement and Pupil Size to Audio Instruction While Viewing a Moving Target, in: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, ETRA '00, pp. 131–138, ACM, New York, NY, USA, doi: 10.1145/355017.355043, 2000.

[384] Tan, D. S., Pausch, R., Stefanucci, J. K., and Proffitt, D. R.: Kinesthetic Cues Aid Spatial Memory, in: CHI '02 Extended Abstracts on Human Factors in Computing Systems, CHI EA '02, pp. 806–807, ACM, New York, NY, USA, doi: 10.1145/506443.506607, URL http://doi.acm.org/10.1145/506443.506607, 2002.

[385] Tan, D. S., Gergle, D., Scupelli, P., and Pausch, R.: With Similar Visual Angles, Larger Displays Improve Spatial Performance, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03, pp. 217–224, ACM, New York, NY, USA, doi: 10.1145/642611.642650, URL http://doi.acm.org/10.1145/642611.642650, 2003.

[386] Tan, D. S., Gergle, D., Scupelli, P. G., and Pausch, R.: Physically Large Displays Improve Path Integration in 3D Virtual Navigation Tasks, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04, pp. 439–446, ACM, New York, NY, USA, doi: 10.1145/985692.985748, URL http://doi.acm.org/10.1145/985692.985748, 2004.

[387] Tang, A., Tory, M., Po, B., Neumann, P., and Carpendale, S.: Collaborative Coupling over Tabletop Displays, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06, pp. 1181–1190, ACM, New York, NY, USA, 2006.

[388] Tattersall, A. J. and Foord, P. S.: An experimental evaluation of instantaneous self-assessment as a measure of workload, Ergonomics, 39, 740–748, doi: 10.1080/00140139608964495, URL http://www.tandfonline.com/doi/abs/10.1080/00140139608964495, 1996.

[389] Thulasidas, M., Guan, C., and Wu, J.: Robust classification of EEG signal for brain-computer interface, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14, 24–29, doi: 10.1109/TNSRE.2005.862695, 2006.

[390] Toker, D., Lallé, S., and Conati, C.: Pupillometry and Head Distance to the Screen to Predict Skill Acquisition during Information Visualization Tasks, in: Proceedings of the 22nd International Conference on Intelligent User Interfaces, IUI '17, pp. 221–231, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3025171.3025187, 2017.

[391] Treacy Solovey, E., Afergan, D., Peck, E. M., Hincks, S. W., and Jacob, R. J. K.: Designing Implicit Interfaces for Physiological Computing: Guidelines and Lessons Learned Using fNIRS, ACM Trans. Comput.-Hum. Interact., 21, 35:1–35:27, doi: 10.1145/2687926, 2015.

[392] Trutschel, U., Heinze, C., Sirois, B., Golz, M., Sommer, D., and Edwards, D.: Heart Rate Measures Reflect the Interaction of Low Mental Workload and Fatigue During Driving Simulation, in: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '12, pp. 261–264, ACM, Portsmouth, New Hampshire, doi: 10.1145/2390256.2390299, 2012.

[393] Tsang, P. S. and Velazquez, V. L.: Diagnosticity and multidimensional subjective workload ratings, Ergonomics, 39, 358–381, doi: 10.1080/00140139608964470, 1996.

[394] Turk, M.: Multimodal Interaction: A Review, Pattern Recognition Letters, 36, 189–195, doi: 10.1016/j.patrec.2013.07.003, 2014.

[395] Tyndiuk, F., Thomas, G., Lespinet-Najib, V., and Schlick, C.: Cognitive Comparison of 3D Interaction in Front of Large vs. Small Displays, in: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST '05, pp. 117–123, ACM, New York, NY, USA, doi: 10.1145/1101616.1101641, URL http://doi.acm.org/10.1145/1101616.1101641, 2005.

[396] Vaish, R., Wyngarden, K., Chen, J., Cheung, B., and Bernstein, M. S.: Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time, in: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pp. 3645–3654, ACM, Toronto, Ontario, Canada, doi: 10.1145/2556288.2556996, URL http://doi.acm.org/10.1145/2556288.2556996, 2014.

[397] Van Benthem, K. D., Cebulski, S., Herdman, C. M., and Keillor, J.: An EEG Brain-Computer Interface Approach for Classifying Vigilance States in Humans: A Gamma Band Focus Supports Low Misclassification Rates, International Journal of Human-Computer Interaction, 34, 226–237, doi: 10.1080/10447318.2017.1342942, URL https://www.tandfonline.com/doi/full/10.1080/10447318.2017.1342942, 2018.

[398] van Dam, A.: Post-WIMP User Interfaces, Commun. ACM, 40, 63–67, doi: 10.1145/253671.253708, URL https://doi.org/10.1145/253671.253708, 1997.

[399] Vetek, A. and Lemmelä, S.: Could a Dialog Save Your Life?: Analyzing the Effects of Speech Interaction Strategies While Driving, in: Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI '11, pp. 145–152, ACM, Alicante, Spain, doi: 10.1145/2070481.2070506, 2011.

[400] Vinson, N. G.: Design Guidelines for Landmarks to Support Navigation in Virtual Environments, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '99, pp. 278–285, ACM, New York, NY, USA, doi: 10.1145/302979.303062, URL http://doi.acm.org/10.1145/302979.303062, 1999.

[401] Vizer, L. M. and Sears, A.: Efficacy of Personalized Models in Discriminating High Cognitive Demand Conditions Using Text-Based Interactions, International Journal of Human-Computer Studies, 104, 80–96, doi: https://doi.org/10.1016/j.ijhcs.2017.03.001, 2017.

[402] Vock, K., Hubenschmid, S., Zagermann, J., Butscher, S., and Reiterer, H.: IDIAR: Augmented Reality Dashboards to Supervise Mobile Intervention Studies, in: Mensch und Computer 2021, pp. 248–259, 2021.

[403] Vukovic, M., Sethu, V., Parker, J., Cavedon, L., Lech, M., and Thangarajah, J.: Estimating cognitive load from speech gathered in a complex real-life training exercise, International Journal of Human-Computer Studies, 124, 116 – 133, doi: https://doi.org/10.1016/j.ijhcs.2018.12.003, URL http://www.sciencedirect.com/science/article/pii/S1071581918306803, 2019.

[404] Wagner, U., Lystbæk, M. N., Manakhov, P., Grønbæk, J. E. S., Pfeuffer, K., and Gellersen, H.: A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/3544548.3581423, URL https://doi.org/10.1145/3544548.3581423, 2023.

[405] Wallace, J. R., Scott, S. D., and MacGregor, C. G.: Collaborative Sensemaking on a Digital Tabletop and Personal Tablets: Prioritization, Comparisons, and Tableaux, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pp. 3345–3354, ACM, New York, NY, USA, 2013.

[406] Wang, L., Gu, T., Liu, A. X., Yao, H., Tao, X., and Lu, J.: Assessing User Mental Workload for Smartphone Applications with Built-in Sensors, IEEE Pervasive Computing, 18, 59–70, 2019.

[407] Wang, W., Li, Z., Wang, Y., and Chen, F.: Indexing Cognitive Workload Based on Pupillary Response Under Luminance and Emotional Changes, in: Proceedings of the 2013 International Conference on Intelligent User Interfaces, IUI '13, pp. 247–256, ACM, New York, NY, USA, doi: 10.1145/2449396.2449428, 2013.

[408] Wangwiwattana, C., Ding, X., and Larson, E. C.: PupilNet, Measuring Task Evoked Pupillary Response using Commodity RGB Tablet Cameras: Comparison to Mobile, Infrared Gaze Trackers for Inferring Cognitive Load, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 1, doi: 10.1145/3161164, URL https://doi.org/10.1145/3161164, 2018.

[409] Weiser, M.: The computer for the 21st century, Scientific American, 265, 66–75, doi: 10.1145/329124.329126, URL http://doi.acm.org/10.1145/329124.329126, 1991.

[410] Weiser, M., Gold, R., and Brown, J. S.: The origins of ubiquitous computing research at PARC in the late 1980s, IBM Systems Journal, 38, 693–696, doi: 10.1147/sj.384.0693, 1999.

[411] Wickens, C. D.: Processing Resources in Attention, Dual Task Performance, and Workload Assessment., Tech. rep., Illinois University at Urbana Engineering-Psychology Research Lab, 1981.

[412] Wickens, C. D.: Processing resources in attention, Varieties of Attention, 1984.

[413] Wickens, C. D.: Multiple Resources and Mental Workload, Human Factors: The Journal of the Human Factors and Ergonomics Society, 50, 449–455, doi: 10.1518/001872008X288394, URL http://journals.sagepub.com/doi/10.1518/001872008X288394, 2008.

[414] Wieland, J., Zagermann, J., Müller, J., and Reiterer, H.: Separation, Composition, or Hybrid? - Comparing Collaborative 3D Object Manipulation Techniques for Handheld Augmented Reality, in: 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 403–412, doi: 10.1109/ISMAR52148.2021.00057, 2021.

[415] Wigdor, D. and Wixon, D.: Brave NUI World: Designing Natural User Interfaces for Touch and Gesture, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edn., 2011.

[416] Wilson, J. R. and Sharples, S.: Evaluation of human work, CRC press, 2015.

[417] Wobbrock, J. O. and Kientz, J. A.: Research contributions in human-computer interaction, interactions, 23, 38–44, 2016.

[418] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M.: Brain-computer interfaces for communication and control, Clinical Neurophysiology, 113, 767–791, doi: 10.1016/S1388-2457(02)00057-3, URL http://www.sciencedirect.com/science/article/pii/S1388245702000573, 2002.

[419] Wozniak, P., Goyal, N., Kucharski, P., Lischke, L., Mayer, S., and Fjeld, M.: RAMPARTS: Supporting Sensemaking with Spatially-Aware Mobile Interactions, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 2447–2460, ACM, New York, NY, USA, doi: 10.1145/2858036.2858491, URL http://doi.acm.org/10.1145/2858036.2858491, 2016.

[420] Xu, J., Wang, Y., Chen, F., and Choi, E.: Pupillary Response Based Cognitive Workload Measurement under Luminance Changes, in: Human-Computer Interaction - INTERACT 2011, edited by Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., and Winckler, M., vol. 6947, pp. 178–185, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-642-23771-3_14, URL http://link.springer.com/10.1007/978-3-642-23771-3_14, 2011.

[421] Yamakoshi, T., Yamakoshi, K.-i., Tanaka, S., Nogawa, M., Shibata, M., Sawada, Y., Rolfe, P., and Hirose, Y.: A preliminary study on driver's stress index using a new method based on differential skin temperature measurement, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 722–725, IEEE, 2007.

[422] Yang, X. and Kim, J. H.: Measuring Workload in a Multitasking Environment Using Fractal Dimension of Pupil Dilation, International Journal of Human-Computer Interaction, URL https://www.tandfonline.com/doi/pdf/10.1080/10447318.2018.1525022?needAccess=true, 2018.

[423] Yang, Y., Reimer, B., Mehler, B., Wong, A., and McDonald, M.: Exploring Differences in the Impact of Auditory and Visual Demands on Driver Behavior, in: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '12, pp. 173–177, ACM, Portsmouth, New Hampshire, doi: 10.1145/2390256.2390285, 2012.

[424] Yin, B., Ruiz, N., Chen, F., and Khawaja, M. A.: Automatic Cognitive Load Detection from Speech Features, in: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, OZCHI '07, pp. 249–255, ACM, Adelaide, Australia, doi: 10.1145/1324892.1324946, 2007.

[425] Yu, K., Epps, J., and Chen, F.: Cognitive Load Evaluation of Handwriting Using Stroke-Level Features, in: Proceedings of the 16th International Conference on Intelligent User Interfaces, IUI '11, pp. 423–426, ACM, New York, NY, USA, doi: 10.1145/1943403.1943481, 2011.

[426] Yuksel, B. F., Afergan, D., Peck, E., Griffin, G., Harrison, L., Chen, N., Chang, R., and Jacob, R.: BRAAHMS: A Novel Adaptive Musical Interface Based on Users' Cognitive State, in: Proceedings of the International Conference on New Interfaces for Musical Expression, edited by Berdahl, E. and Allison, J., pp. 136–139, Louisiana State University, Baton Rouge, Louisiana, USA, 2015.

[427] Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., and Jacob, R. J.: Learn Piano with BACh: An Adaptive Learning Interface That Adjusts Task Difficulty Based on Brain State, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 5372–5384, ACM, Santa Clara, California, USA, doi: 10.1145/2858036.2858388, URL http://doi.acm.org/10.1145/2858036.2858388, 2016.

[428] Zagermann, J., Pfeil, U., Rädle, R., Jetter, H.-C., Klokmose, C., and Reiterer, H.: When Tablets Meet Tabletops: The Effect of Tabletop Size on Around-the-Table Collaboration with Personal Tablets, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 5470–5481, ACM, New York, NY, USA, doi: 10.1145/2858036.2858224, URL http://doi.acm.org/10.1145/2858036.2858224, 2016.

[429] Zagermann, J., Pfeil, U., and Reiterer, H.: Measuring cognitive load using eye tracking technology in visual computing, in: Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization, pp. 78–85, 2016.

[430] Zagermann, J., Pfeil, U., Fink, D., von Bauer, P., and Reiterer, H.: Memory in Motion: The Influence of Gesture- and Touch-Based Input Modalities on Spatial Memory, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pp. 1899–1910, ACM, New York, NY, USA, doi: 10.1145/3025453.3026001, URL http://doi.acm.org/10.1145/3025453.3026001, 2017.

[431] Zagermann, J., Pfeil, U., and Reiterer, H.: Studying eye movements as a basis for measuring cognitive load, in: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–6, 2018.

[432] Zagermann, J., Pfeil, U., von Bauer, P., Fink, D., and Reiterer, H.: "It's in my other hand!" – Studying the Interplay of Interaction Techniques and Multi-Tablet Activities, pp. 1–13, Association for Computing Machinery, New York, NY, USA, URL https://doi.org/10.1145/3313831.3376540, 2020.

[433] Zagermann, J., Hubenschmid, S., Balestrucci, P., Feuchtner, T., Mayer, S., Ernst, M. O., Schmidt, A., and Reiterer, H.: Complementary Interfaces for Visual Computing, it - Information Technology, 64, 145–154, doi: doi:10.1515/itit-2022-0031, 2022.

[434] Zbrodoff, N. J. and Logan, G. D.: What everyone finds: The problem-size effect., in: Handbook of mathematical cognition., pp. 331–345, Psychology Press, New York, NY, US, 2005.

[435] Zhang, X., Lyu, Y., Hu, X., Hu, Z., Shi, Y., and Yin, H.: Evaluating Photoplethysmogram as a Real-Time Cognitive Load Assessment during Game Playing, International Journal of Human-Computer Interaction, 34, 695–706, doi: 10.1080/10447318.2018.1461763, URL https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1461763, 2018.

[436] Zheng, R. Z.: Cognitive load measurement and application, Routledge, 2017.

[437] Ziegler, J., Graube, M., Suhrbier, A., Wessel, N., Malberg, H., and Urbas, L.: The Influence of the Spatial Separation of Control Elements on the Workload for Mobile Information Systems, in: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '11, pp. 191–200, Association for Computing Machinery, New York, NY, USA, doi: 10.1145/2037373.2037403, 2011.

[438] Zijlstra, F. and Van Doorn, L.: The construction of a scale to measure perceived effort, University of Technology, 1985.

[439] Zimmerer, C., Krop, P., Fischbach, M., and Latoschik, M. E.: Reducing the Cognitive Load of Playing a Digital Tabletop Game with a Multimodal Interface, in: CHI Conference on Human Factors in Computing Systems, pp. 1–13, ACM, New Orleans LA USA, doi: 10.1145/3491102.3502062, 2022.

[440] Züger, M. and Fritz, T.: Interruptibility of Software Developers and Its Prediction Using Psycho-Physiological Sensors, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 2981–2990, Association for Computing Machinery, Seoul, Republic of Korea, doi: 10.1145/2702123.2702593, 2015.