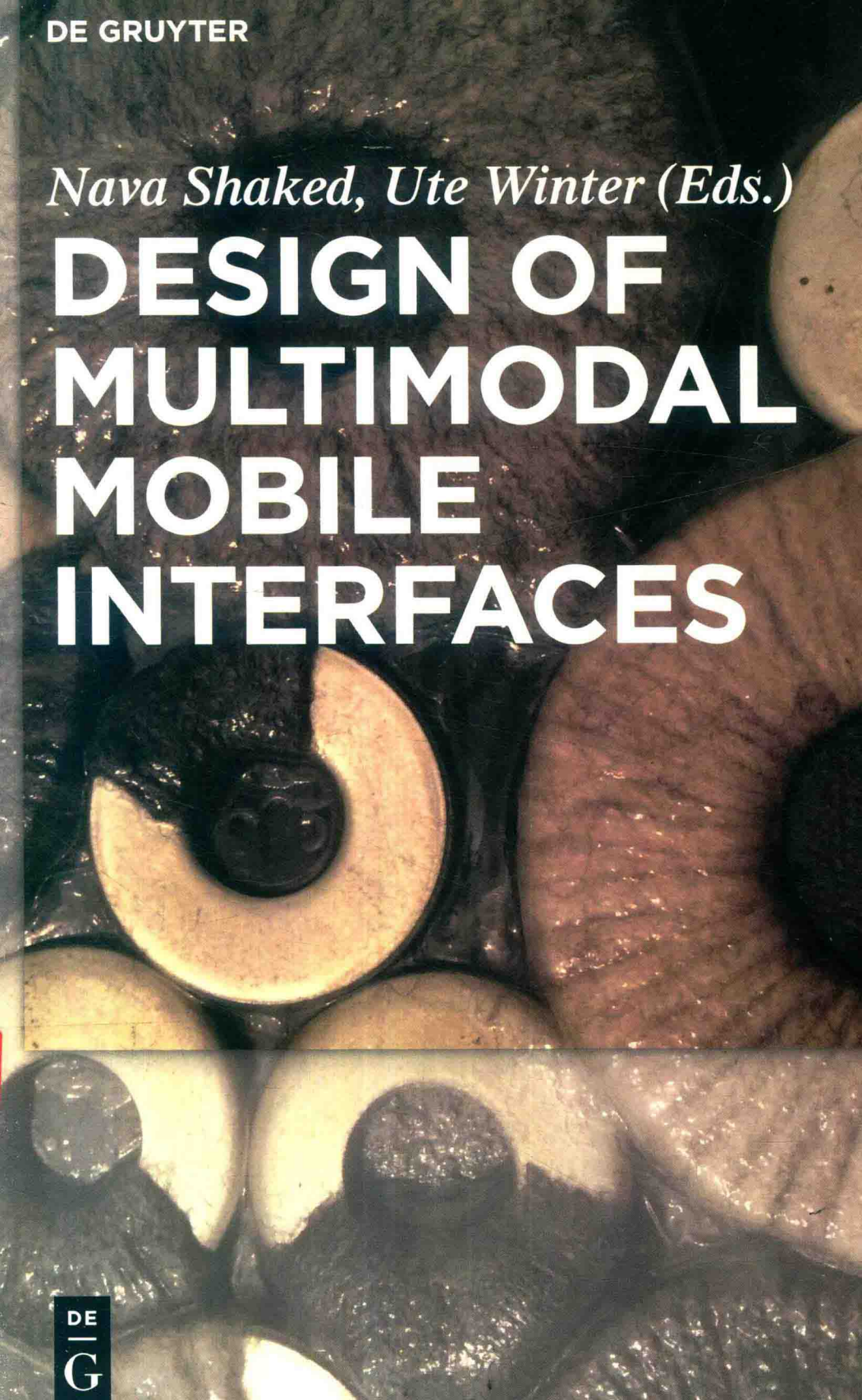
DE GRUYTER



Nava Shaked, Ute Winter (Eds.)

DESIGN OF MULTIMODAL MOBILE INTERFACES

Design of Multimodal Mobile Interfaces

Edited by
Nava Shaked, Ute Winter

DE GRUYTER

Editors
Nava Shaked
HIT – Holon Institute of Technology
52 Golomb Street
POB 305 Holon 5810201
Israel
shakedn@hit.ac.il

Ute Winter
General Motors – Advanced Technical Center Israel
7 HaMada St.
7th Floor POB 12091
Herzliya Pituach 46733
Israel
ute.winter@gm.com

ISBN 978-1-5015-1084-7
e-ISBN (PDF) 978-1-5015-0273-6
e-ISBN (EPUB) 978-1-5015-0275-0
Set-ISBN 978-1-5015-0274-3

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Cover image: "Circles" Original design by Brit Shaked
Typesetting: PTP Protago-TEX-Production, Berlin
Printing and binding: CPI books GmbH, Leck
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

Preface

This book is a collection of research articles and reviews in the area of Mobile Multimodality. It is looking at the field through various perspectives and areas of expertise in order to describe the main issues and challenges of putting together an optimized multimodal interface for mobile environments such as automotive, wearables, mobile applications, and avatars and virtual agents.

The book also raises usability questions that are of major concern to interface designers and architects in general, and even more so in a multimodal environment. A recurring usability theme is the importance of proper information or knowledge management. Today we are busy collecting and analyzing enormous amounts of data. Nevertheless, how much data is sufficient in order to create the "right" user experience? How do we present this information to the user? How can we introduce context into the multimodal paradigm, presenting information at the right time and in the right place?

What role do NLP engines play in the processing stage? How compatible are the current NLP engines with Mobile Multimodal systems? And how are NLP engines being standardized to support mobile multimodality?

And finally, does multimodality pose social, moral or behavioral questions that we need to answer while putting together all other components? In our highly technological, mobile and always-connected world where communication is essential and fast, and information is at everyone's fingertips from such an early age, how much freedom of choice and privacy should we support?

Acknowledgments

This book is the result of two years of continuous exploration of a field that is constantly changing and evolving. We wish to thank all the people who supported us in the process. First, we wish to thank the authors themselves, who agreed to share their research and vision – from academic work to real-time industry projects. We also want to thank HIT – Holon Institute of Technology and General Motors Advanced Technical Center – Israel for supporting us and allowing this book to develop. A big thanks to Dr. Laura Rosenbaun, our research assistant, who contributed greatly to the research and editing as well as to maintaining our morale. And finally to our book cover designer Brit Shaked, for sharing with us her creation.

We hope this book will inspire our students and colleagues in the fields of Mobility, Multimodality, and Interfaces, to keep developing and researching these exciting and continuously advancing domains.

What's in the book

Mobile Multimodality has touched many industries and is constantly developing and changing. The chapters in this book cover a wide variety of Mobile Multimodality issues. We shed light on central questions concerning the usability, concepts and complexity of Mobile Multimodality, taking into consideration not only some of the most interesting technological aspects but also the designers' and the users' concerns.

Most of the chapters deal with the challenge of deploying multimodal interface methods and technologies in a wide variety of contexts – mobile phone, automotive, contact center, wearables, etc. Typically, these systems collect contextual information for the purpose of personalization and optimization. Although the technology is becoming more reliable, the next challenge is dynamic adaptation to the requirements of the user, the tasks or the domain, as well as to environment and/or infrastructure constraints. One way to meet this challenge is to change the technology, while another approach is to gather large amounts of information about the user and the environment in order to create and present a different user interface in each context.

In the first chapter, *Introduction to the evolution of mobile multimodality*, Shaked and Winter lay the groundwork through explaining the basic terms and concepts of Mobile Multimodality and Human-Machine Interaction in general, and Multimodal Interaction in particular. The chapter explores the evolution of user behavior and needs as well as the development of new and exciting interaction technologies. The authors conclude by raising the "chicken and egg" question: What are the determining factors that influence new interface developments? Are they the users' new needs and requirements? Or are the latter only a response to the emergence of new innovative technologies?

Interaction is often about how the machine processes and perceives natural language automatically. In the chapter *Integrating natural language resources in mobile applications*, Dhal explores the contribution of Natural Language Processing (NLP) resources to the development of mobile applications with Multimodal Interfaces. The chapter describes the NLP components that are useful for integration into mobile applications. This integration, however, is extremely challenging and standardized specifications such as EMMA and MMI play a major role in re-

ducing the effort required to create a seamless interface between NLP systems and other interfacing software.

Omnichannel architecture is an emerging technology that uses NLP components. In their chapter, *Omnichannel natural language – Bringing omnichannel and prediction to the multimodal user interface*, Brown, Walia and Urban discuss the fusion of contextual data for the purpose of prediction and service optimization. The authors describe the process of combining next generation UI technology with NLP, Omnichannel prediction, Big Data, and a real-time learning platform. Multimodal Interaction is defined and determined not only by the collection of data from various channels (SMS, voice calls, chats, web, etc.) but also by consolidation of the modalities used in those channels such as text, speech, gesture and more. The objective is to create a holistic contextual experience based on the user's current and past interaction history.

The utilization of various interaction data is also closely connected to the field of wearables. Both wearable computing and the Omnichannel infrastructure require building and maintaining an ecosystem to ensure effective interaction. The next chapter, *Wearable computing – The context challenge and a media ecology approach*, by Lawo, Logan and Pasher, explores the relationship between the user and her wearable environment. They argue that understanding the environment created by wearable computing as well as its psychological and social impact is crucial to implementing this technology. They introduce the concept of Wearable Ecology as a framework for understanding these impacts. The chapter reviews past and recent research on context frameworks for wearable computing and provides considerations for future directions in the development of wearable technology.

An additional use case of interaction and context data is laid out in *Spoken dialog systems adaptation for domains and for users* by Sun and Rudnicky. In this chapter Spoken Dialog Systems are the framework for a discussion of adaptation strategies that a machine can apply to accommodate user needs and domain specifics. How do individual users prefer to complete tasks? What is their mental model of the system and the domains that the system is designed for? How do user expectations and preferences, as well as the understanding of system capabilities and domain functionality, change over time? The authors significantly contribute to the questions raised through exploring adaptive strategies for language understanding and learning of user intention in applications that attempt to ensure a personalized user experience.

One of the most appealing interaction methodologies for mobile users are Avatars and Virtual Agents (A&VA). Virtual personal assistants, health caregivers, and game avatars have become part of our everyday activities. Their interaction modalities are discussed in the next chapter *The use of multimodality in avatars*

and virtual agents by Shaked and Artelt. Like wearables, a special relationship between user and machine is created through A&VAs. Furthermore, the chapter argues that there is a direct relation between the development and progress of A&VA applications and the development and availability of multimodal interaction. The research question is how to determine the correct interaction design for A&VAs. The chapter examines a set of classification methods to characterize the types of user-machine relationship and interaction. This approach is then used to determine (a) which multimodal environment is best suited for the different interaction and relationship types, and (b) the optimal set of features for each case. As in the Omnichannel and the wearable environments, the authors recommend establishing a well-defined ecosystem in order to achieve the desired interaction.

Personal assistants may be presented to the user as voice persona, such as the popular Siri, Cortana, and Google Now speech-enabled personal assistants. A very compelling use case is the speech-enabling of an in-car infotainment system. The mobile industry has discovered this field of late, and has been developing tools to project user phone content and capabilities into the car infotainment system. For example, Siri can be used in a car via Apple CarPlay. To ensure efficient and pleasant communications of this type, user studies must be conducted in the specific driving context. Van Over, Molina-Markham, Lie, and Carbaugh have undertaken a near naturalistic study, which focusses on the cultural nature of communication between a human and machine in this setting. *Managing interaction with an in-car infotainment system* presents an investigation of turn exchange, a major aspect of any communication sequence. It further focusses on interactional misalignments as a result of misplaced verbal and multimodal cues. Which of those misalignments are culturally constituted? How do users adapt to such imperfect systems in order to accomplish their goals? Their findings lead to considerations for future multimodal in-vehicle system designs.

Two chapters discuss information organization on displays from different perspectives. *Classification and organization of information: The case of the head up display* by Heymann and Degani raises the question of how to organize and present information to drivers in a car environment using emerging technologies and features ranging from safety control to driver comfort to entertainment and communication capabilities. The chapter uses head up displays as an example of how overwhelming quantities of information can be categorized and prioritized. Their approach leads to a proposal of what elements of information should be presented to the driver on a head up display vs. a cluster or center stack display. Several existing head up displays are reviewed and a set of principles is presented to guide designers in their effort to organize information. The presented approach is relevant to display design in other environments and for other devices.

The second information organization chapter – *Towards objective methods in display design* by Shmueli-Friedland, Zelman, Degani and Asherov – discusses a complementary method for the display of information, in this case demonstrated through the example of the visual automotive instrument cluster. It applies a formal approach to the analysis and design of driver displays, using a series of methods involving domain experts' estimation as well as applying concepts of graph theory. The authors thus generate a cluster hierarchy that uncovers relationships among the informational elements. They illustrate the usefulness of their approach by analyzing a generic cluster display, and by revealing problems and missed opportunities in the design. In the end, they propose that this methodology can be generalized to include additional modalities, such as sound and haptics.

List of contributing authors

Sanjay Agarwal
[24]7 Inc., Bangalore, India
sanjay.a@247-inc.com
Chapter 3

Detlev Artelt
AIXVOX GmbH
Germany
da@aixvox.net
Chapter 6

Ron Asherov
General Motors R&D
Advanced Technical Center Israel
rasherov@gmail.com
Chapter 9

Kathy Brown
[24]7 Inc., Campbell CA
kathy.brown@247-inc.com
Chapter 3

Donal Carbaugh
Department of Communication
University of Massachusettsm Amherst, MA
carbaugh@comm.umass.edu
Chapter 7

Abir Chakraborty
[24]7 Inc., Bangalore, India
abir.c@247-inc.com
Chapter 3

Deborah A. Dahl
Conversational Technologies, USA
dahl@conversational-technologies.com
Chapter 2

Asaf Degani
General Motors R&D
Advanced Technical Center Israel
asaf.degani@gm.com
Chapters 8 and 9

Michael Heymann
Department of Computer Science,
Technion, Israel Institute of Technology
heymann@cs.technion.ac.il
Chapter 8

Prashant Joshi
[24]7 Inc., Bangalore, India
prashant.joshi@247-inc.com
Chapter 3

Michael Lawo
Centre for Computing and Communication
Technologies (TZI) at the Universität Bremen
mlawo@tzi.de
Chapter 4

Sunny Lie
Pomona Department of Communication
California State Polytechnic University
slie@cpp.edu
Chapter 7

Robert K. Logan
Department of Physics and St. Michael's
College at the University of Toronto
logan@physics.utoronto.ca
Chapter 4

Elizabeth Molina-Markham
Independent Scholar
emolmark@gmail.com
Chapter 7

Brion van Over
Department of Communication
Manchester Community College
Manchester, CT
bvanover@manchestercc.edu
Chapter 7

Edna Pasher
EP Management Consultants
edna@pasher.co.il
Chapter 4

Alexander I. Rudnicky
Language Technologies Institute
Carnegie Mellon University
air@cs.cmu.edu
Chapter 5

Nava Shaked
HIT – Holon Institute of Technology
Israel
shakedn@hit.ac.il
Chapters 1 and 6

Yael Shmueli-Friedland
General Motors R&D
Advanced Technical Center Israel
yael.shmueli@gm.com
Chapter 9

Ming Sun
Language Technologies Institute
Carnegie Mellon University
mings@cs.cmu.edu
Chapter 5

Anmol Walia
[24]7 Inc., Campbell CA
anmol.walia@247-inc.com
Chapter 3

Ute Winter
General Motors R&D
Advanced Technical Center Israel
ute.winter@gm.com
Chapter 1

Ido Zelman
General Motors R&D
Advanced Technical Center Israel
ido.zelman@gm.com
Chapter 9

Contents

Nava Shaked and Ute Winter

1 Introduction to the evolution of Mobile Multimodality

Abstract: Human-Machine Interaction technologies and design have been at the center of both academic and industry research for several decades. Recent developments in the fields of mobility, sensors, Big Data, and their underlying algorithms have created an opportunity for a unique fusion of multiple factors to bring interaction to the next level. This chapter presents a model that describes the mutual influence and constant adaptation of technology and user needs in the area of Mobile Multimodality. We start by defining the key terms as they are used in the literature, with a focus on Mobility, User Interface, and Multimodality. We then describe major factors of influence such as social interaction, user behavior, contextual factors, and technological developments, among others. Having laid the groundwork for our model, we look closely at the two parties involved and sharing any interface: the users and the underlying technologies. We discuss both the evolution of user requirements and needs relative to mobile device interfaces as well as the historical development, readiness and availability of the underlying technologies. We then propose a cycle of mutual influence and adaptation between user needs and technology development, which leads to the constant evolution of Mobile Multimodal Interfaces.

1.1 User Interfaces: Does vision meet reality?

Movies and television shows traditionally transport us into a world of imagination and inspiration. They explore ideas and let us dream what could be if we just make it happen. One of the most appealing multimodal mobile interfaces of the 1980s was KITT, Michael Knight's car and communication partner in the series Knight Rider. Episode 12 in Season 2 contains the following spoken interaction sequence, when KITT had just failed to warn Michael in time of an imminent dangerous encounter with villains, and had to watch the incident from afar:

KITT: Michael! Michael! *(Pause while KITT waits to see if Michael got harmed)*
KITT: I'm terribly sorry, Michael, but he had such a head start.
Michael: It's ok, fell. What you'd get on the car?
KITT: It's out of scale of range. But I did run the plates. Unfortunately they were stolen from another car yesterday. Michael, are you alright?
Michael: Yeah, I am alright.

Apparently KITT has impressive conversational skills with a genuine human touch, makes observations via sensors with astonishing precision, and is connected to external networks and data. Its black displays though, representing the utterance only by some flashing red bars that simulate KITTs speech signal, seems highly unimaginative today after 30 years of continuous technology development. Still, KITT was undoubtedly multimodal and in many ways predicted the future.

After years of interfacing with KITT, what would Michael's reaction be to one of the current speech enabled multimodal interfaces? Despite the unquestionable progress in the technology and design of such interfaces, would he perceive this interaction to be natural and intuitive? Would he find the interface easy to use? Or perhaps users would prefer that multimodal interfaces be similar to KITT?

1.2 Discussion of terms: Mobility and User Interface

Like KITT, who accompanies Michael wherever he chooses to take his vehicle, multimodal user interfaces are often mobile. Mobility is therefore an important concept to examine in any discussion of multimodal interfaces.

1.2.1 Mobility

The Merriam-Webster dictionary defines mobility as being "able to move from one place to another", without or "with the use of vehicles", or as being "able to be moved". However, mobility should not be viewed only as a spatial phenomenon, associated with physical movement. As Ishii (2006, p. 347) points out, "mobility should be understood in a broader sense to include at least three interrelated dimensions of human interaction; namely, spatial, temporal, and contextual mobility" (see also Kakihara & Sorensen 2002). Users can interact with the interface and perform their intended task anywhere and anytime. Mobile interfaces thus have a huge impact on the social interaction between users and on the organization of daily tasks, thus contributing to social interaction and efficiency (Srivastava 2005; Schroeder 2010; Campbell & Kwak 2011). The third and more comprehensive factor, namely contextual mobility, may be of even higher social significance. Any human interaction, whether with humans or with devices, is context dependent (Hymes 1972; Gumperz 1982; Hagen et al. 2005; Tamminen et al. 2004; Stivers & Sidnell 2005). Space and time are only part of what constitutes context and contributes to the meaning of interaction; there are many other dimensions, such as cultural background, degree of familiarity among the interaction partners and the device, mood, use cases, etc. Mobile devices allow users to choose the context for