

Pamela Savage-Knepshield
Jessie Chen *Editors*

Advances in Human Factors in Robots and Unmanned Systems

Proceedings of the AHFE 2016
International Conference on Human
Factors in Robots and Unmanned
Systems, July 27–31, 2016, Walt
Disney World[®], Florida, USA

Advances in Intelligent Systems and Computing

Volume 499

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagrass, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at <http://www.springer.com/series/11156>

Pamela Savage-Knepshield · Jessie Chen
Editors

Advances in Human Factors in Robots and Unmanned Systems

Proceedings of the AHFE 2016 International
Conference on Human Factors in Robots
and Unmanned Systems, July 27–31, 2016,
Walt Disney World[®], Florida, USA

 Springer

Editors

Pamela Savage-Knepshield
U.S. Army Research Laboratory
Aberdeen Proving Ground
Aberdeen, MD
USA

Jessie Chen
U.S. Army Research Laboratory
Aberdeen Proving Ground
Aberdeen, MD
USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-41958-9 ISBN 978-3-319-41959-6 (eBook)
DOI 10.1007/978-3-319-41959-6

Library of Congress Control Number: 2016943962

© Springer International Publishing Switzerland 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Advances in Human Factors and Ergonomics 2016

AHFE 2016 Series Editors

*Tareq Z. Ahram, Florida, USA
Waldemar Karwowski, Florida, USA*

7th International Conference on Applied Human Factors and Ergonomics

Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems, July 27–31, 2016, Walt Disney World[®], Florida, USA

<i>Advances in Cross-Cultural Decision Making</i>	<i>Sae Schatz and Mark Hoffman</i>
<i>Advances in Applied Digital Human Modeling and Simulation</i>	<i>Vincent G. Duffy</i>
<i>Advances in Human Factors and Ergonomics in Healthcare</i>	<i>Vincent G. Duffy and Nancy Lightner</i>
<i>Advances in Affective and Pleasurable Design</i>	<i>WonJoon Chung and Cliff(Sungsoo) Shin</i>
<i>Advances in Human Aspects of Transportation</i>	<i>Neville A. Stanton, Steven Landry, Giuseppe Di Bucchianico and Andrea Vallicelli</i>
<i>Advances in Ergonomics In Design</i>	<i>Francisco Rebelo and Marcelo Soares</i>
<i>Advances in Ergonomics Modeling, Usability & Special Populations</i>	<i>Marcelo Soares, Christianne Falcão and Tareq Z. Ahram</i>
<i>Advances in Social & Occupational Ergonomics</i>	<i>Richard Goossens</i>
<i>Advances in Neuroergonomics and Cognitive Engineering</i>	<i>Kelly S. Hale and Kay M. Stanney</i>
<i>Advances in Physical Ergonomics and Human Factors</i>	<i>Ravindra Goonetilleke and Waldemar Karwowski</i>
<i>Advances in The Ergonomics in Manufacturing: Managing the Enterprise of the Future</i>	<i>Christopher Schlick and Stefan Trzcielinski</i>
<i>Advances in Safety Management and Human Factors</i>	<i>Pedro Arezes</i>
<i>Advances in Human Factors, Software, and Systems Engineering</i>	<i>Ben Amaba</i>
<i>Advances in Human Factors and Sustainable Infrastructure</i>	<i>Jerzy Charytonowicz</i>

(continued)

(continued)

<i>Advances in The Human Side of Service Engineering</i>	<i>Tareq Z. Ahram and Waldemar Karwowski</i>
<i>Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries</i>	<i>Sacit Cetiner, Paul Fechtelkötter and Michael Legatt</i>
<i>Advances in Human Factors in Sports and Outdoor Recreation</i>	<i>Paul Salmon and Anne-Claire Macquet</i>
<i>Advances in Human Factors and System Interactions</i>	<i>Isabel L. Nunes</i>
<i>Advances in Human Factors, Business Management, Training and Education</i>	<i>Jussi Kantola, Tibor Barath, Salman Nazir and Terence Andre</i>
<i>Advances in Human Factors in Robots and Unmanned Systems</i>	<i>Pamela Savage-Knepshield and Jessie Chen</i>
<i>Advances in Design for Inclusion</i>	<i>Giuseppe Di Bucchianico and Pete Kercher</i>
<i>Advances in Human Factors in Cybersecurity</i>	<i>Denise Nicholson, Janae Lockett-Reynolds and Katherine Muse</i>

Preface

Researchers are conducting cutting-edge investigations in the area of unmanned systems to inform and improve how humans interact with robotic platforms. Many of the efforts are focused on refining the underlying algorithms that define system operation and on revolutionizing the design of human–system interfaces. The multifaceted goals of this research is to improve ease of use, learnability, suitability, and human–system performance, which in turn, will reduce the number of personnel hours and dedicated resources necessary to train, operate, and maintain the systems. As our dependence on unmanned systems grows along with the desire to reduce the manpower needed to operate them across both military and commercial sectors, it becomes increasingly critical that system designs are safe, efficient, and effective. Optimizing human–robot interaction and reducing cognitive workload at the user interface requires research emphasis to understand what information the operator requires, when they require it, and in what form it should be presented so they can intervene and take control of unmanned platforms when it is required. With a reduction in manpower, each individual’s role in system operation becomes even more important to the overall success of the mission or task at hand. Researchers are developing theories as well as prototype user interfaces to understand how best to support human–system interaction in complex operational environments. Because humans tend to be the most flexible and integral part of unmanned systems, the human factors and unmanned systems’ focus considers the role of the human early in the design and development process in order to facilitate the design of effective human–system interaction and teaming.

This book will prove useful to a variety of professionals, researchers, and students in the broad field of robotics and unmanned systems who are interested in the design of multi-sensory user interfaces (auditory, visual, and haptic), user-centered design, and task-function allocation when using artificial intelligence/automation to offset cognitive workload for the human operator. We hope this book is informative, but even more so that it is thought-provoking. We hope it provides inspiration,

leading the reader to formulate new, innovative research questions, applications, and potential solutions for creating effective human–system interaction and teaming with robots and unmanned systems.

We would like to thank the editorial board members for their contributions.

M. Barnes, USA
P. Bonato, USA
G. Calhoun, USA
R. Clothier, Australia
D. Ferris, USA
J. Fraczek, Poland
S. Hill, USA
M. Hou, Canada
C. Johnson, UK
K. Neville, USA
J. Pons, Spain
R. Taiar, France
J. Thomas, USA
A. Tvaryanas, USA
H. Van der Kooij, The Netherlands
H. Widroither, Germany
H. Zhou, UK

Aberdeen, USA
July 2016

Pamela Savage-Knepshield
Jessie Chen

Contents

Part I A Vision for Future Soldier-Robot Teams

Iterative Interface Design for Robot Integration with Tactical Teams	3
K.M. Ibrahim Asif, Cindy L. Bethel and Daniel W. Carruth	
Context Sensitive Tactile Displays for Bidirectional HRI Communications	17
Bruce Mortimer and Linda Elliott	
An Initial Investigation of Exogenous Orienting Visual Display Cues for Dismounted Human-Robot Communication	27
Julian Abich IV, Daniel J. Barber and Linda R. Elliott	
Five Requisites for Human-Agent Decision Sharing in Military Environments	39
Michael Barnes, Jessie Chen, Kristin E. Schaefer, Troy Kelley, Cheryl Giammanco and Susan Hill	
Initial Performance Assessment of a Control Interface for Unmanned Ground Vehicle Operation Using a Simulation Platform	49
Leif T. Jensen, Teena M. Garrison, Daniel W. Carruth, Cindy L. Bethel, Phillip J. Durst and Christopher T. Goodin	

Part II Confronting Human Factors Challenges

Examining Human Factors Challenges of Sustainable Small Unmanned Aircraft System (sUAS) Operations	61
Clint R. Balog, Brent A. Terwilliger, Dennis A. Vincenzi and David C. Ison	

**Mission Capabilities—Based Testing for Maintenance Trainers:
Ensuring Trainers Support Human Performance 75**
Alma M. Sorensen, Mark E. Morris and Pedro Geliga

Detecting Deictic Gestures for Control of Mobile Robots 87
Tobias Nowack, Stefan Lutherdt, Stefan Jehring, Yue Xiong,
Sabine Wenzel and Peter Kurtz

Effects of Time Pressure and Task Difficulty on Visual Search 97
Xiaoli Fan, Qianxiang Zhou, Fang Xie and Zhongqi Liu

Part III Human-Agent Teaming

**Operator-Autonomy Teaming Interfaces to Support
Multi-Unmanned Vehicle Missions. 113**
Gloria L. Calhoun, Heath A. Ruff, Kyle J. Behymer
and Elizabeth M. Mersch

**Shaping Trust Through Transparent Design: Theoretical
and Experimental Guidelines. 127**
Joseph B. Lyons, Garrett G. Sadler, Kolina Koltai, Henri Battiste,
Nhut T. Ho, Lauren C. Hoffmann, David Smith, Walter Johnson
and Robert Shively

**A Framework for Human-Agent Social Systems: The Role
of Non-technical Factors in Operation Success 137**
Monika Lohani, Charlene Stokes, Natalia Dashan, Marissa McCoy,
Christopher A. Bailey and Susan E. Rivers

**Insights into Human-Agent Teaming: Intelligent Agent
Transparency and Uncertainty 149**
Kimberly Stowers, Nicholas Kasdaglis, Michael Rupp, Jessie Chen,
Daniel Barber and Michael Barnes

**Displaying Information to Support Transparency for Autonomous
Platforms 161**
Anthony R. Selkowitz, Cintya A. Larios, Shan G. Lakhmani
and Jessie Y.C. Chen

**The Relevance of Theory to Human-Robot Teaming Research
and Development 175**
Grace Teo, Ryan Wohleber, Jinchao Lin and Lauren Reinerman-Jones

Part IV From Theory to Application: UAV and Human-Robot Collaboration

Classification and Prediction of Human Behaviors by a Mobile Robot 189
D. Paul Benjamin, Hong Yue and Damian Lyons

The Future of Human Robot Teams in the Army: Factors Affecting a Model of Human-System Dialogue Towards Greater Team Collaboration 197
A. William Evans, Matthew Marge, Ethan Stump, Garrett Warnell, Joseph Conroy, Douglas Summers-Stay and David Baran

Human-Autonomy Teaming Using Flexible Human Performance Models: An Initial Pilot Study 211
Christopher J. Shannon, David C. Horney, Kimberly F. Jackson and Jonathan P. How

Self-scaling Human-Agent Cooperation Concept for Joint Fighter-UCAV Operations 225
Florian Reich, Felix Heilemann, Dennis Mund and Axel Schulte

Experimental Analysis of Behavioral Workload Indicators to Facilitate Adaptive Automation for Fighter-UCAV Interoperability 239
Dennis Mund, Felix Heilemann, Florian Reich, Elisabeth Denk, Diana Donath and Axel Schulte

Part V Supporting Sensor and UAV Users

Model-Driven Sensor Operation Assistance for a Transport Helicopter Crew in Manned-Unmanned Teaming Missions: Selecting the Automation Level by Machine Decision-Making 253
Christian Ruf and Peter Stütz

Using Natural Language to Enable Mission Managers to Control Multiple Heterogeneous UAVs 267
Anna C. Trujillo, Javier Puig-Navarro, S. Bilal Mehdi and A. Kyle McQuarry

Adaptive Interaction Criteria for Future Remotely Piloted Aircraft 281
Jens Alfredson

Confidence-Based State Estimation: A Novel Tool for Test and Evaluation of Human-Systems 291
Amar R. Marathe, Jonathan R. McDaniel, Stephen M. Gordon and Kaleb McDowell

Human Robots Interactions: Mechanical Safety Data for Physical Contacts 305
 Alberto Fonseca and Claudia Pires

Part VI An Exploration of Real-World Implications for Human-Robot Interaction

Droning on About Drones—Acceptance of and Perceived Barriers to Drones in Civil Usage Contexts 317
 Chantal Lidynia, Ralf Philipsen and Martina Ziefle

Factors Affecting Performance of Human-Automation Teams 331
 Anthony L. Baker and Joseph R. Keebler

A Neurophysiological Examination of Multi-robot Control During NASA’s Extreme Environment Mission Operations Project 341
 John G. Blich

A Comparison of Trust Measures in Human–Robot Interaction Scenarios 353
 Theresa T. Kessler, Cintya Larios, Tiffani Walker, Valarie Yerdon and P.A. Hancock

Human-Robot Interaction: Proximity and Speed—Slowly Back Away from the Robot! 365
 Keith R. MacArthur, Kimberly Stowers and P.A. Hancock

Human Factors Issues for the Design of a Cobotic System 375
 Théo Moulières-Seban, David Bitonneau, Jean-Marc Salotti, Jean-François Thibault and Bernard Claverie

A Natural Interaction Interface for UAVs Using Intuitive Gesture Recognition 387
 Meghan Chandarana, Anna Trujillo, Kenji Shimada and B. Danette Allen

Part VII Optimizing Human-Systems Performance Through System Design

An Analysis of Displays for Probabilistic Robotic Mission Verification Results 401
 Matthew O’Brien and Ronald Arkin

A Neurophysiological Assessment of Multi-robot Control During NASA’s Pavilion Lake Research Project 415
 John G. Blich

**A Method for Neighborhood Gesture Learning Based on Resistance
Distance 427**
Paul M. Yanik, Anthony L. Threatt, Jessica Merino, Joe Manganelli,
Johnell O. Brooks, Keith E. Green and Ian D. Walker

Part I
A Vision for Future Soldier-Robot Teams

Iterative Interface Design for Robot Integration with Tactical Teams

K.M. Ibrahim Asif, Cindy L. Bethel and Daniel W. Carruth

Abstract This research investigated mobile user interface requirements for robots used in tactical operations and evaluated user responses through an iterative participatory design process. A longitudinal observational study (five sessions across six months) was conducted for the iterative development of robot capabilities and a mobile user interface. Select members of the tactical team wore the mobile interface and performed operations with the robot. After each training an after-action review and feedback was received about the training, the interface, robot capabilities, and desired modifications. Based on the feedback provided, iterative updates were made to the robotic system and the user interface. The field training studies presented difficulties in the interpretation of the responses due to complex interactions and external influences. Iterative designs, observations, and lessons learned are presented related to the integration of robots with tactical teams.

Keywords Human-robot interaction · Mobile user interface · Human factors · SWAT · Robot integration

This research was sponsored by the U.S. Army Research Laboratory Under Grant W911NF-13-1-0481. The views and conclusions contained in this document are those of the author's and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

K.M. Ibrahim Asif (✉) · C.L. Bethel · D.W. Carruth
Department of Computer Science and Engineering, Mississippi State University,
Mississippi State, MS 39762, USA
e-mail: kia13@msstate.edu

C.L. Bethel
e-mail: cbethel@cse.msstate.edu

D.W. Carruth
e-mail: dwc2@cavs.msstate.edu

C.L. Bethel
Center for Advanced Vehicular Systems, Mississippi State University,
Mississippi State, MS 39762, USA

1 Introduction

The focus of this paper was to iteratively design and evaluate a mobile interface for control and communication with ground robots for use in law enforcement tactical team operations. This was investigated through the research efforts of Mississippi State University's Social, Therapeutic, and Robotics Systems (STaRS) Laboratory and the Starkville (Mississippi) Special Weapons and Tactics (SWAT) team. The research team was given access to monitor the SWAT team's activities and incorporate the use of the robot within the team's monthly training exercises. Select members of the tactical team wore the mobile user interface to perform different operations using the robot. After each training session, the team members completed an after-action review and provided feedback about the training, robot, and the mobile user interface. Based on the information provided by the team, modifications were made to the interface and/or the robot prior to the next training session and then reevaluated.

The members of law enforcement tactical teams have one of the most dangerous civilian occupations. Their tasks are challenging because of the need to operate in dynamic, unpredictable, and often unknown environments. The primary purpose of SWAT teams is to ensure safety and to save lives in a systematic approach in different high-risk situations, such as engaging with active shooters, subduing barricaded suspects, rescuing hostages, and similar incidences [1]. They want to take all possible actions to minimize danger during their mission responses.

The background and motivation for this paper is presented in Sect. 2. Related work is presented in Sect. 3. The methodology for the research is included in Sect. 4 of the paper. Section 5 includes details regarding the design and development of the iterative user interfaces using a participatory design process. In Sect. 6 challenges and lessons learned are presented, followed by conclusions and future work in Sect. 7.

2 Background and Motivation

Robots such as unmanned ground vehicles (UGV) are ideal for integration with tactical teams such as SWAT. The United States, Department of Defense (DoD), defines a UGV as "a powered physical system with (optionally) no human operator aboard the principal platform, which can act remotely to accomplish assigned tasks" [2]. UGVs can be programmed extensively to operate semi-autonomously in different terrains [3]. The U.S. Army's 30-year unmanned ground systems (UGS) campaign plan is expected to decrease the workloads on warfighters both physically and cognitively as well as increase their combat capabilities [2]. The same statement applies for SWAT teams, who have to operate in the civilian environment. For the rest of the paper, the term *robot* will be used specifically instead of UGV for ease of reading.

Historically, robots are imagined as ideal for jobs that have one or more of the three Ds: dirty, dull, or dangerous [4]. Takayama et al. state that this simplistic notion has changed over time in the last ten years [5]. Their analysis indicates that people now consider robots for jobs where memorization, strong perceptual skills, and service-orientation are key factors. Human-robot teaming is the collaboration between humans and intelligent robots to perform specific tasks. Woods et al. point out that in such teaming both the human and the robot contribute unique qualities [6]. While a robot may be equipped with advanced sensing capabilities superior to a human, today's robots require human intervention to understand and act on the information.

SWAT officers envision the possibility that robots may be able to make their jobs safer. However, they do not have a good understanding of the strengths, weaknesses, and operational requirements needed to incorporate robots into their activities. Robots could be a significant addition to SWAT and increase their capabilities. One of the most significant benefits of integrating robots into tactical environments is the minimization of risk to the lives of the SWAT team members. Robots could be sent into the environment ahead of rest of the team giving them the ability to have "eyes" on the scene first before having to directly encounter potential threats. The following are some other benefits for integrating robots into SWAT operations:

- i. For scouting to gain critical intelligence about the environment and potential threats.
- ii. For guarding a specific room or hallway, while the team conducts slow and methodical searches in other directions.
- iii. For diverting the attention of threats when encountered and providing team members with an advantage in these types of situations.

When a robot is used in SWAT operations, it is often considered as a camera on wheels [7–9]. Another common use of robots in law enforcement is for explosive ordnance disposal [10, 11]. How a robot could potentially be integrated as a member of the team instead of just another piece of equipment was initially explored in [12].

3 Related Work

In 1999, the Defense Advanced Research Project Agency (DARPA) was interested in research for tactical robots since it was evident that military conflicts would occur in urban areas with a large amount of building infrastructure and civilians [13]. Since this time, research has been conducted for the development of this type of robots and methods of communication with these robots. While the focus was on military operations, it also applied to other tactical teams such as SWAT teams, who have to operate in a similar manner in the civilian environment. Researchers

have discussed with SWAT team members, issues related to the integration of robots into tactical environments such as SWAT operations [10, 14].

Over time, robots have been successfully integrated and used in different instances with law enforcement teams for explosive ordnance disposal [10, 11] or as remote cameras to provide intelligence [7, 9, 15]. While some of these robots are small in size, which helps operators maneuver the robots more easily, sometimes law enforcement teams prefer larger robots, because they feel these robots could provide improved protection and enhanced capabilities that may save lives. Hence, they are willing to compensate for the challenges that may occur with these robots (e.g., large size, battery life, sound generation, etc.).

There has been limited research in the area of integrating these robots as teammates instead of considering the robot as a tool for performing one specific task with tactical teams such as SWAT [10, 11, 14, 15]. When deploying a robot with a SWAT team, it is usually thought of as an adjunct to the team or a tool to operate remotely. The issue with this concept is the team member who is operating the robot is practically unable to perform his own tasks as he is now acting solely as an operator for the robot. A typical operator control unit (OCU) is large in size and often the person using it cannot perform any additional tasks that would be part of their typical responsibilities as a SWAT team member. Moreover, OCUs have a high associated cost from both the hardware and software perspectives. This is a challenge for any SWAT teams as they try to use minimal manpower to respond to most tactical situations. Also there are challenges associated with the use of robots with SWAT teams, such as the robots being considered a distraction to SWAT operations and potentially slowing down the team in specific types of responses, especially with dynamic entry [14]. Research has been performed to identify the potential roles a robot could fulfill within a SWAT team, serving as a team member, and the acceptance of a robot integrated as part of a SWAT team [12].

Over the years, researchers have come up with different user interface (UI) concepts to interact with robots [16–19]. Some of these interfaces would not be feasible for use in a tactical environment. Researchers also have explored the use of telepresence, multimodal displays, and the effects of temporal latency in the effort to create a better OCU for military environments [20]. Traditionally, these OCUs require larger hardware and expensive software, which makes the operating cost of the robots very high. As the overall budget for the U.S. DoD will be reduced by \$489 billion over next ten years [2], DoD is investigating affordable, interoperable, integrated, and technologically advanced robots. The increasing popularity of low cost mobile devices and wearables in recent years has provided a great opportunity to use such devices as OCUs to communicate with different types of robots especially in tactical team responses. Moreover, the familiarity of using these devices among the members of the tactical response teams is a significant benefit. While mobile phones has been used previously as an OCU, it has only been considered as an alternative controller to joysticks [21]. Very limited research has been done to identify mobile interfaces as a completely viable OCU approach for implementation. The goal of the research presented in this paper is to identify the viability of

using mobile interface to control and communicate with the robots in SWAT team and how an iterative design process could help with achieving this.

4 Methodology

The method used in this study involved the development of a mobile interface to control and communicate with a ground robot for use with SWAT teams. In SWAT trainings, a robot was deployed in different runs to perform different scenarios and training/response operations. The training sessions were held in different locations such as a large multi-story building, an abandoned house, open air police training facility, and a controlled testbed facility designed as a small two-bedroom apartment similar to locations commonly investigated in SWAT responses. Initially, a basic concept interface was designed and an ethnographic interview was conducted with one of the SWAT team leaders. Based on the input, a working prototype was built to use during future SWAT team trainings. Training was held approximately once a month at these locations. After each training an after-action review was performed and feedback was solicited about the training, robot, and interface, with a focus on requested modifications. Based on feedback from the officers, iterative modifications were made to the robot platforms and the interface prior to the next training, when additional feedback was requested.

4.1 Hardware

A Jaguar V4 robotic platform from Dr. Robot is used in this research [22]. Modifications were made and additional components were added to address the requirements of the SWAT team, such as cameras, LED lighting features, and a speaker system for distraction sounds. An Arduino board was installed inside the robot to control the lights and sounds. A Lenovo laptop was used to run the robot. A PicoStation from Ubiquiti Networks was used to increase the range for controlling and communicating with the robot from a distance [23]. Android smartphones (both Nexus 5 and Nexus 6) were used to serve as the operator control unit (OCU) for this robot and to activate the lights and sounds diversion device. Logitech game controller and Wii remote with Nunchuck controller were additionally used for teleoperation of the robot.

4.2 Software

The software architecture was designed as a distributed control system to support a high degree of modularity for adding and/or removing hardware and software

control components as required for the control of the robot, sensors, and onboard distraction devices. It was implemented using the Robot Operating System (ROS). An Android application was developed to control and communicate with the robot. Over time, different iterations of this interface was introduced to address the requirements of the SWAT team. The following functionalities were implemented within the interface:

- i. Display live streaming video feeds broadcasted from the robot.
- ii. Toggle between the two different camera feeds on the robot: drive camera and top camera.
- iii. Control the onboard strobe lights and diversion sounds.
- iv. Drive the robot through the interface.

5 Iterative Interfaces

The traditional mechanism for driving the Jaguar V4 robot was using a laptop as an OCU, which was not feasible for use with the SWAT teams because of different factors such as size and weight. The use of a laptop strapped onto someone's chest and having him or her constantly looking at the screen completely negates any effectiveness of that member as part of the team. Other than operating the robot, the member would not be able to participate in the response operations at all. This is not ideal, because most SWAT teams are small in numbers and having a team member unusable because of the robot defeats the purpose of using a robot as a teammate to serve as a force multiplier and to increase the team's capabilities.

This prompted the researchers to develop a better solution that would allow the SWAT team to control and communicate with the robot effectively without compromising one of the team members. In the think aloud discussion with members of the team, they wanted a mechanism that would fit perfectly with their current equipment and would not be a distraction. Familiarity with the device and common interfaces and reduced learning curve were important to them as well. A smart phone strapped on their non-dominant arm with an armband (Fig. 4a) seemed a viable solution to this problem. This approach required significant development efforts on both the hardware and software aspects because there was no existing codebase for this robotic system that could be used as a starting point for the development of a mobile interface.

The researchers had to design and build the interface from scratch. As part of the process to develop a functional prototype, an ethnographical interview with one of the SWAT team leaders to identify the requirements for this type of an interface was performed. Since the SWAT team members had not used this type of an interface before and their technical knowledge was limited in terms of the design process, the two main requirements was a) no glare from the screen if possible to minimize distraction and b) to perform operations as easily as possible. It was determined that

the team members would wear the interface on their non-dominant arm and that the interface would stay in landscape mode only.

Driving in different directions, activating different lights/strobes and sounds, switch between cameras, etc. are defined as different operations that can be carried out through the interface. Similar operations were grouped together and considered as functionalities within the interface. The four functionalities were: driving the robot, activating lights/strobes, activating sounds, and switching between camera feeds. An iterative process of designing the mobile interface over time through a participatory design process was followed with feedback received from the SWAT team members after each training session. Three major iterations were performed and implemented during this six-month series of training exercises.

5.1 First Interface Iteration

For the first iteration the researchers decided to use a tabbed interface approach. The interface functionalities were in different tabs (Camera, Drive, Lights, and Sounds as shown in Fig. 1). The default view would provide the operator with a camera feed that could be toggled between different cameras with a single tap (Fig. 1a). For activating lights (Fig. 1c) or playing a sound (Fig. 1d), the operator could select the desired function and specific operation in the interface and it would activate immediately. To deactivate the operation, the operator would tap again the specific operation indicator. This was helpful as the officers could activate multiple lights and deactivate them using one action as long as the operator remained on the lights function panel.

The researchers then tested this interface in two training sessions and while the officers' responses were positive overall, their most requested change was to be able

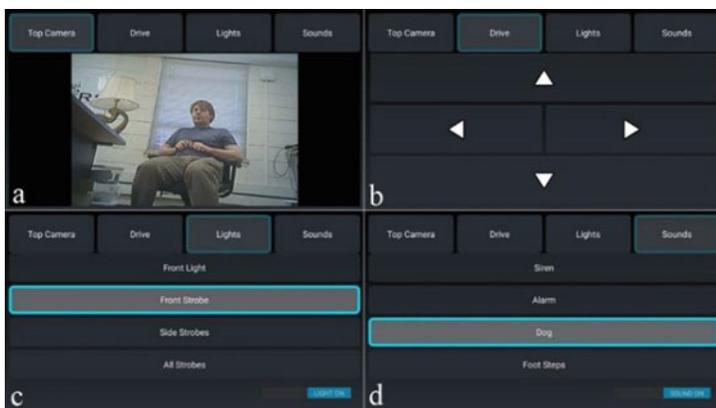


Fig. 1 First interface iteration: **a** broadcast panel from top camera, **b** drive controls panel, **c** lights and strobes options panel, and **d** diversion sound options panel

to perform operations while simultaneously being able to watch the video feed from the onboard cameras. This feedback directed the researchers' efforts to develop a second iteration of the interface.

5.2 Second Interface Iteration

As requested by the SWAT team, the researchers redesigned the interface to be able to broadcast the video from the cameras while having the ability to perform other operations. This interface resulted in a three-pane system: (1) the left pane contained all of the available functionalities, (2) the center pane contained the video feed(s), and the right pane contained the operations, such as lighting, sound, and drive commands (refer to Fig. 2). It was natural to put the video pane in the center, for the other two panes the researchers did a heuristic evaluation for usability inspection [24]. Since the team members would often wear the interface on their left arms, it was easier to access needed buttons and features on the right side of the interface compared to the left as that would a further reach and closer to the body limiting access. In general, it was determined that Operations would be performed more frequently compared to selection of basic functionalities. Hence, the researchers designed the pane setup using this approach.

The researchers also wanted to test incorporating another function and set of operations—access to additional sensors on the robot and as a placeholder for a future expected enhancement. To accommodate this addition, the lights and sounds functionalities were combined and assigned in the interface to a button named Utilities. This added an extra layer of required actions forcing the team members to select the utilities panel, select desired operations, and activate specific buttons for the desired lights (Fig. 2c) and sounds (Fig. 2d) to activate them. An improvement

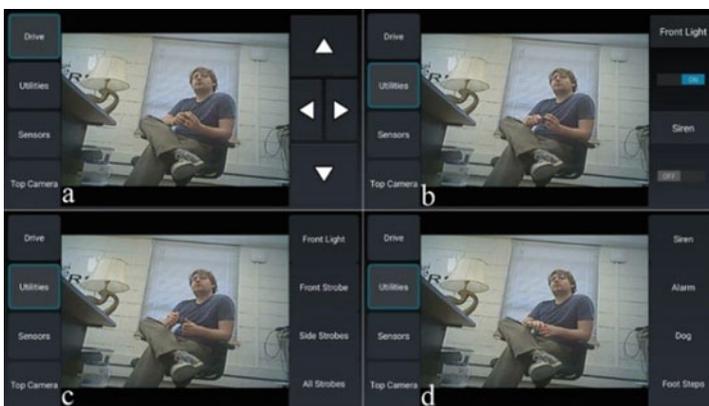


Fig. 2 Seconds interface iteration—camera broadcast always visible: **a** drive controls, **b** light and sound control as utilities, **c** lights and strobes options, and **d** diversion sounds options

from the previous interface iteration was the ability toggle between the cameras on the robot without having to change the current operation pane.

The researcher tested this iteration of the interface during the next two training sessions. The overall response from the SWAT team was more positive compared to the response to the first interface iteration. The team was happy to be able to view the video feed and simultaneously access specific operations. The researchers also queried the SWAT team about their thoughts on having the functionalities on the right side of the interface and the operations on the left with a minor adjustment to the screen. Most of the team members stated they preferred the current setup in the second interface iteration that located the functionality buttons on the left and the operation buttons on the right. The biggest disadvantage they revealed in this iteration was having to perform additional actions to activate the lights and sounds when needed and while under high stress. Their feedback was to accomplish a desired task with fewer required actions and in a more informative way if possible. This led the researchers to design and develop a third interface iteration, which is the current version used in training.

5.3 Third and Current Interface Iteration

The researchers tried to address the concerns expressed by the members of the SWAT team and developed the current (third) iteration for the interface. Since, there was no actual additional sensory data to show in the interface at this time, it was decided to return lights and sounds functions to individual buttons on the functionality pane of the interface and to remove the Utilities and Sensor functionality buttons from this version of the interface (see Fig. 3). This change allowed the lights (Fig. 3b) and sounds (Fig. 3c) functions (buttons) to be activated through direct operations. All of the operations were set as toggle buttons that allowed the

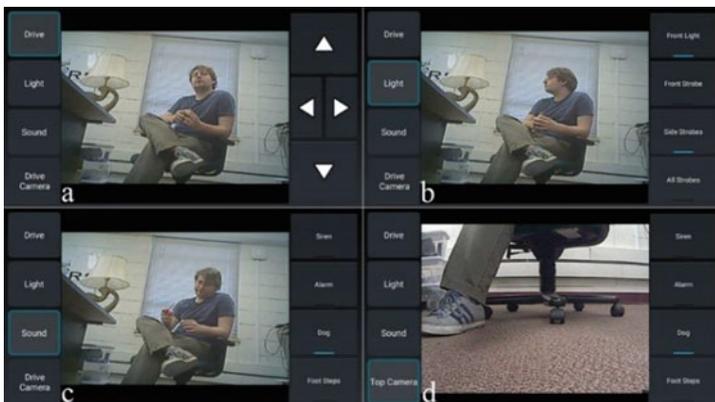


Fig. 3 Current iteration—camera broadcast always visible: **a** drive controls, **b** lights and strobes options, **c** diversion sounds options, and **d** toggled camera showing drive camera

SWAT team to be able to activate and deactivate any specific light or sound by tapping once when the appropriate function panel was selected. This reduced the number of actions required to activate an operation by 1. Toggling between cameras (see Fig. 3d) was modified to indicate which camera would be displayed on activation of the button.

The researchers evaluated the current (third) interface iteration over the next three training sessions and it seems the SWAT team really likes how the operations are more intuitive now. As possible minor modifications, the researchers explored the idea of being able to view both of the camera feeds at simultaneously by having them side by side or up and down in the center pane of the interface. The responses from the team were not supportive of this approach. They preferred the single camera view as it was easier to obtain scene understanding with the single video feed instead of the small frames displaying the two camera feeds. The size of display was an important factor in this decision.

During each iterations, the researchers explored the use of different controllers as well as virtual controller within the interface for teleoperation of the robot. This was to identify the acceptability and usability for teleoperation. SWAT officers seem to prefer physical controller in compare to virtual one. Moreover, Wii remote with Nunchuk was preferred compared to Logitech game controller as it can be operated with just grabbing and with the thumb. The researchers still exploring different mechanisms for teleoperation of the robot.

The researchers continue to evaluate and integrate small updates to the current iteration of the interface. The SWAT team has been using this iteration of the mobile interface to control and communicate with the robot in their training sessions (see Fig. 4).

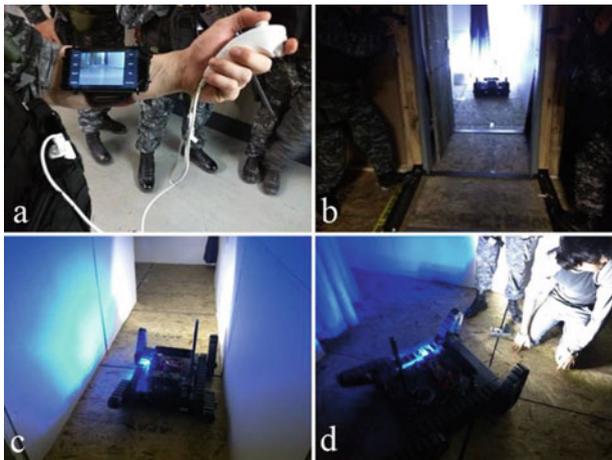


Fig. 4 Use of current interface in SWAT training **a** team member with the interface **b** robot entering the apartment, **c** robot checking the hallway with activated strobe and sound, and **d** robot assists in capturing the suspected threat

6 Challenges Faced and Lessons Learned

The initial discussion of system capabilities and requirements was limited by a lack of experience with the proposed system. In some cases, this is inherent in the early stages of design: a potential user cannot interact with a system that does not yet exist. In the researchers' case, the law enforcement officers initially had limited experience with controlling the robot and little to no knowledge of the robot's capabilities. This led to difficulty in identifying concrete requirements for the user interface.

The initial mobile device (Nexus 5) selected had a 4.95" screen with a 1920×1080 resolution. This provided limited space for the display of camera feeds and buttons for the activation of available operations. The subsequent mobile device (Nexus 6) has a 5.96" screen with a 2560×1440 resolution. While the increased size improved the available space for the camera display and controls, it has also negatively affected comfort. There is a clear tradeoff between comfort and the effectiveness of the display area, which is directly impacted by the size of the device. To accommodate all of the desired operations, we explored hierarchical menus in the second iteration to discover that the tradeoff of an increased number of actions per operation versus the total number of available operations. In the current case, the SWAT team members did not consider this to be an effective solution.

In addition to difficulties with providing effective access to light and sound operations, using digital (on/off) buttons to drive the robot was not effective. During the evaluation of the interfaces, the SWAT team members were also provided an external controller based on a Wii Nunchuk to provide analog control of the robots movements. This provided the team members with a single-handed controller that was well received and allowed for better control of the robot. It was able to be attached to their current gear and did not impede the performance of tasks by the officers.

Training was performed in a number of different scenarios that revealed different types of challenges for the SWAT team members. The variation in scenarios affected officer responses to the robot and the provided interfaces. While it would be tempting to control the scenario, the current study sought to observe integration in a range of training scenarios defined by the officers to understand the complexities of designing a user interface that would be effective in many scenarios commonly faced by law enforcement. A custom interface or robot for each scenario is not a practical solution for robot operations in law enforcement. The iterative design process provided the SWAT team members with repeated exposure to the robot and interfaces leading to increased experiences with the robot and increased understanding of robot and interface capabilities. This, in turn, significantly improved the quality of feedback and the ability for the SWAT team members to provide effective recommendations for modifications needed in the interface.

7 Conclusion and Future Work

The development and evaluation of a mobile interface for use as an OCU with tactical teams has not received significant attention from the research community. The importance of such interface is significant because it could help with the integration of robots with SWAT teams without the need to sacrifice a team member to serve as a dedicated robot operator. The process was challenging because the officers must function in different types of responses that require different capabilities in the robot and functionalities in the interface. The ability to provide an interface that can be easily used by officers operating in high stress environments, requires significant participatory design to develop both high usability as well as a positive user experience. Through a process of multiple training and three interface iterations, we have developed an interface that meets the current needs of the officers.

Because lives are dependent on the robot and the interface working to meet the needs of the officer, the involvement of the officers in different training exercises was critical to this project. It was discovered that what may work in one scenario (e.g., slow and methodical building searches) may not work for another scenario or response type such as fast, dynamic entry. It is important to consider what functionality is needed when an officer is in a safe location versus in immediate threat of danger. It is also important to consider what happens when the officers must work in low-to-no light conditions versus a daytime response. An immediate concern related to the latest version of the interface is backlighting the officers due to the ambient light from the display in a low-to-no light response, the officers could be placed in danger because they would be the focus of attention. This research will be ongoing and will continue to enhance all aspects of the robot functionality in addition to the necessary changes required in the interface to support tactical operations that include the integration of robots into tactical teams.

The focus for future work in this research is to collect and analyze data from two different user studies that address the usability and the acceptability of the mobile interface developed while continuing the iterative process of development for the interface. Two large studies have already been developed and performed in spring 2016. Future interface enhancements may include the use of tactic buttons or interfaces that do not require ambient lighting and may be safer and provide a better user experience.

Acknowledgments The authors would like to thank the Starkville City Police for allowing us to work with their SWAT teams. We would also like to thank Zachary Henkel, Christopher Hudson, Lucas Kramer, and Paul Barrett for their support and assistance throughout this project.

References

1. Association, N.T.O.: SWAT standards for law enforcement agencies. NTOA (2008)
2. Winnefeld, J.A., Kendall, F.: Unmanned Systems Integrated Roadmap: Fy2013-2038. DIANE Publishing Company (2014)
3. Gonzales, D., Harting, S.: Designing Unmanned Systems with Greater Autonomy: Using a Federated, Partially Open Systems Architecture Approach. RAND Corporation (2014)
4. Murphy, R.: Introduction to AI Robotics. MIT Press, Cambridge, MA (2000)
5. Takayama, L., Ju, W., Nass, C.: Beyond dirty, dangerous and dull: What everyday people think robots should do. In: 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 25–32. (2008)
6. Woods, D.D., Tittle, J., Feil, M., Roesler, A.: Envisioning human-robot coordination in future operations. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **34**, 210–218 (2004)
7. Schreiner, K.: Operation: microrobot. *IEEE Intell. Syst. Their Appl.* **14**, 5–7 (1999)
8. Kumagai, J.: Techno cops [police robotic and electronic technology]. *IEEE Spectr.* **39**, 34–39 (2002)
9. Yue, L., Qiang, H., Yuancan, H., Liancun, Z., Junyao, G., Ye, T.: A throwable miniature robotic system. In: IEEE International Conference on Automation and Logistics (ICAL), pp. 114–118. (2011)
10. Nguyen, H.G., Bott, J.P.: Robotics for law enforcement: applications beyond explosive ordnance disposal. In: International Society for Optics and Photonics Enabling Technologies for Law Enforcement, pp. 433–454 (Year)
11. Lundberg, C., Christensen, H.I.: Assessment of man-portable robots for law enforcement agencies. In: Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems, pp. 76–83. ACM, Washington, D.C. (2007)
12. Bethel, C.L., Carruth, D., Garrison, T.: Discoveries from integrating robots into SWAT team training exercises. In: IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pp. 1–8 (2012)
13. Krotkov, E., Blich, J.: The defense advanced research projects agency (DARPA) tactical mobile robotics program. *Int. J. Robot. Res.* **18**, 769–776 (1999)
14. Jones, H.L., Rock, S.M., Burns, D., Morris, S.: Autonomous robots in swat applications: research, design, and operations challenges. In: Symposium of Association for Unmanned Vehicle Systems International (2002)
15. Blich, L.T.C.: Semi-autonomous tactical robots for urban operations. In: Intelligent Control (ISIC). Held Jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), Intelligent Systems and Semiotics (ISAS), Proceedings, pp. 783–794 (1998)
16. Atrash, A., Kaplow, R., Villemure, J., West, R., Yamani, H., Pineau, J.: Development and validation of a robust speech interface for improved human-robot interaction. *Int. J. Social Robot.* **1**, 345–356 (2009)
17. Kumar, S., Sekmen, A.: Single robot—Multiple human interaction via intelligent user interfaces. *Knowl.-Based Syst.* **21**, 458–465 (2008)
18. Lapedes, P., Sharlin, E., Sousa, M.C.: Three dimensional tangible user interface for controlling a robotic team. In: 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 343–350 (2008)
19. Waldherr, S., Romero, R., Thrun, S.: A gesture based interface for human-robot interaction. *Auton. Robots* **9**, 151–173
20. Jentsch, F., Barnes, M., Harris, P.D., Salas, E., Stanton, P.N.A.: Human-Robot Interactions in Future Military Operations. Ashgate Publishing Limited (2012)
21. Walker, A.M., Miller, D.P., Ling, C.: Spatial orientation aware smartphones for tele-operated robot control in military environments: a usability experiment. *Proc. Human Factors Ergon. Soc. Ann. Meet.* **57**, 2027–2031 (2013)

22. http://jaguar.drrobot.com/specification_V4.asp
23. <https://www.ubnt.com/airmax/picostationm/>
24. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 249–256. ACM, Seattle, Washington, USA (1990)

Context Sensitive Tactile Displays for Bidirectional HRI Communications

Bruce Mortimer and Linda Elliott

Abstract Tactile displays have shown high potential to support critical communications, by relaying information, queries, and direction cues from the robot to the Soldier in a hands-free manner. Similarly, Soldiers can communicate to the robot through speech, gesture, or visual display. They could also respond to robot messages by pressing a button to acknowledge or request a repeat message. A series of studies have demonstrated the ability of Soldiers to interpret tactile direction and information cues during waypoint navigation in rough terrain, during day and night operations. The design of tactile display systems must result in reliable message detection. We have proposed a framework to ensure salience of the tactile cues. In this presentation we summarize research efforts that explore factors affecting the efficacy of tactile cues for bidirectional soldier-robot communications. We will propose methods for changing tactile salience based on the symbology and context.

Keywords Human factors · Tactile · Adaptive · Salience · Communication

1 Introduction

Tactical situation awareness (SA) is critical for ensuring the Warfighters situational dominance, especially during combat. SA can improve security, survivability and optimize lethality. Combat environments can subject Warfighters to extreme conditions, testing the limits of both their physical and cognitive abilities. So, while there have been considerable advancements in military and commercial efforts to integrate and display context sensitive SA information, displaying this information

B. Mortimer

Engineering Acoustics, Inc, 406 Live Oaks Blvd, Casselberry, FL 32707, USA
e-mail: bmort@eaiinfo.com

L. Elliott (✉)

Army Research Laboratory Human Research and Engineering Directorate,
Fort Benning, Georgia, USA
e-mail: linda.r.elliott.civ@mail.mil

© Springer International Publishing Switzerland 2017

P. Savage-Knepshield and J. Chen (eds.), *Advances in Human Factors in Robots and Unmanned Systems*, Advances in Intelligent Systems and Computing 499,
DOI 10.1007/978-3-319-41959-6_2

and interacting with the system remains a challenge on the battlefield. One of the general problems in all sensory perception is information overload. Tactile cueing has proven to be a particularly effective means of conveying direction and spatial orientation information and, when implemented effectively, can increase performance (e.g., speed, accuracy) and lower cognitive workload [1–5].

Tactile communication is as old as human communication itself. Touch is also a natural information channel that complements our sight, hearing, and the sense of balance. Tactile displays have been investigated as a sensory or cognitive augmentation mechanism for delivering feedback to users. Tactile cueing can also be used to convey information [6]. However, there is a difference between feeling a stimulus and identifying its meaning. While body-referenced tactile orientation is a natural or intuitive representation of spatial information, tactile communication signals and their associations must first be learnt and then detected and interpreted by the user. The maximum information rate for the auditory channel, extrapolated from the perception of normal rate speech, is estimated at 75 bits/s. The tactile channel is typically lower, with rates of 25–50 bits/s possible for highly experienced Braille users [7]. Despite this limitation, the transfer of tactile information can potentially occur on many different levels or “channels”. Combat ground soldiers do not need the same type that higher echelons need—too much information may be a distraction and induce cognitive overload [8]. Therefore complex tactile language constructs that require concentration are not likely to be useful.

Several studies have shown that there are a range of touch variables that can be used to convey useful information to Soldiers [9]. We have recently demonstrated a Navigation and Communication system, *NavCom*, that includes GPS driven tactile navigation cues. In field trials, *NavCom* reduced mission times and increased navigation accuracy [10]—Soldiers reported being more aware of their surroundings and having better control of their weapon. In one specific trial, Soldiers using *NavCom* during a 300 m navigation task at night checked their visual (map) display on average 1.2 times, versus 17.7 times for those without *NavCom*; one demonstration of how *NavCom* can improve effectiveness and lower cognitive workload, effort, and frustration. We also showed that tactile navigation cueing and tactile coding for messages could be presented simultaneously. *NavCom* was also fielded at the 2016 AEWE exercise in Ft Benning.

Although task performance enhancements have been demonstrated, care must be taken when considering transitioning from one tactile display mode to another. For example, the usual construct for tactile cueing may be to alert the wearer of an impending threat through the use of a body-referenced vibrotactile stimulus in a particular sector. The same touch variable may be used in navigation where the tactile cueing may be in terms of a tactile “instruction” to head towards a particular sector. Clearly, in this example the two modes are ambiguous and could in fact be contradictory! The effective transition from tactile display modes or context remains a challenge.

Typically, tactile stimuli are defined by dimensions such as the frequency, intensity, force, location, and duration of the signal. However, these definitions and their associated thresholds, in isolation, are of little value if one does not consider

characteristics of the user or situational context. Potential contributions of tactile cues may be obscured through inattention to moderating factors, fuzzy construct definitions, or imprecise measures. A common limitation of investigations of tactile cues is over-reliance on performance or workload outcomes, without consideration or measurement of mediating factors such as the ease by which tactile cues can be perceived and interpreted. We have introduced a framework for describing and quantifying the performance of tactile systems [11].

As unmanned assets become more autonomous, Soldier-robot communications will evolve from tele-operation and focused oversight to more tactical, multimodal, critical communications. Therefore the requirements for tactile and Soldier multi-modal displays will need careful analysis and optimization to meet these expanded needs. In this paper, we discuss our research towards designing context sensitive and adaptive displays using our salience model.

2 Tactile Salience

One of the general problems in all sensory perception is information overload, but we humans are adept at using selective attention to quickly prioritize large amounts of information and give attention to what is most important. Salience is the property of a stimulus that allows it to stand out and be noticed. Salience is widely used in describing visual system performance of humans and in computational models that drive computer vision [12, 13] but has not been extensively or systematically applied to the tactile modality.

Tactile salience can be simply defined as the probability that the tactile cue will be detected. In controlled laboratory settings, salience can often be modeled as a function of tactor engineering and the vibratory stimuli characteristics—i.e. physical characteristics of the signal itself, when context, or “noise”, is very low. However, as context becomes more complex, additional factors become significant [14].

2.1 *Tactile Salience Construct*

It is clear that tactile salience is affected or moderated by many factors in addition to engineering characteristics of the factors. While issues of tactor engineering are clearly important to the concept of salience, the salience of any tactile cue or tactor will be affected, perhaps to a great extent, by many factors, including body location, characteristics of the user, the task demands, and the environment in general.

Perhaps even more important, these factors interact with each other, such that the interactions may be more highly predictive than particular characteristics per se. Predictions of operator performance in naturalistic settings require the consideration of these characteristics as they interact in a particular setting. To better emphasize

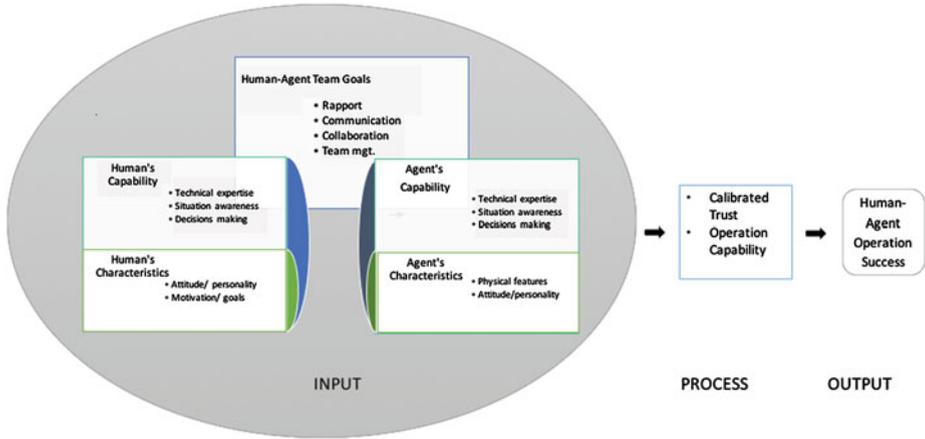


Fig. 1 Core factors and interactions affecting tactile salience

the interplay of these factors, Fig. 1 shows the construct of tactile salience as mediated by three core factors—characteristics pertaining to the user, the technology, and the environment—and their interactions. The effect of any one characteristic cannot be precisely predicted without consideration and/or control of other core factors.

The three main sources of influences on tactile salience, as shown in Fig. 1 are:

- *Technology characteristics/capabilities.* Many studies have addressed a multitude of features related to the design of tactile stimulators and the construction of the stimulus signal. There is no doubt that these factors affect salience, and can predict user perception, localization, and interpretation, in controlled settings. For example, abrupt onset (or changes) in stimuli and high frequency (200–300 Hz) tone burst vibrations are known to be naturally salient. Effects of characteristics such as amplitude, frequency, and ISI are well summarized in a number of publications [15–17].
- *Individual/User differences.* Salience can also be affected by characteristics of the user. These include sensory and perceptual characteristics common to all operators, such as sensory processing limitations that limit tactile discrimination and body location [18]. These can also include individual differences in cognitive abilities, personality, training, experience, age, or posture. Differences can always occur with regard to user motivation or focus of attention.
- *Environmental factors and task demands.* It has been shown that factors such as physical demands and workload can significantly affect tactile cue perception. In addition, a variety of contextual aspects, such as operational tempo, physical demands, nature and type of distracters, level of threat, and consequences of failure should be considered. Features such as environmental noise can certainly impact the perception, recognition and thus effectiveness of tactile signals.

While each category can act as a main effect on tactile salience, interactions among the categories are also important.

Interactions between the user and the environment/task context produce factors such as perceptions of stress, workload, or fatigue that are likely to affect attention, the need for alerts, and/or the ability of the user to attend to alerts. As an example, individuals with higher levels of neuroticism, emotional reactivity, and/or lower stress tolerance are likely to experience work situations more intensely [19]. While simple direction cues may prove valuable when the user is stressed or fatigued, more complex cues may be less likely to be attended to. Thus, map-based visual information and complex audio information (e.g., turn north after the second street on your left) can become much less effective than tactile direction cues (e.g., go this way). One can see that issues regarding multisensory integration would fit here.

Interactions between environmental/task context and technology include the degree of match between the operational context and technology features or capabilities. Basic examples include situations that require factors that are very quiet (ex, covert communications), or that augment attention management (complex decision making), or that are very easily perceived (e.g., during strenuous movements).

Interactions between the user and technology basically address the traditional domain of human factors engineering. The mismatch between user and technology characteristics can result in poor performance, when technology does not address operator norms that affect their ability to perceive and easily interpret tactile signals.

Thus, it is reasonable to posit that tactile salience depends on main effects and interactions among characteristics of the user, the technology, and the environment.

Tactile salience depends on the user, the technology, and the environment. To be salient for diverse users, tactile display technology should provide a wide range of recognizable touch characteristics, and do so in a small lightweight, efficient system that is not limited by the mounting or usage. Research has shown that features such the abrupt onset (or changes) in stimuli and high frequency (200–300 Hz) tone burst vibrations are known to be naturally salient, while lower frequency tone burst vibrations are typically less salient. A complete consideration of tactile salience is essential to accomplish successful critical signaling, through three fundamental steps: (1) perception: the signal must get their attention, (2) the signal must provide recognizable and distinct stimuli (patterns, locations, areas, changes in patterns, timing, tempo, melody) and (3) provide some confidence or support that substantiates the selection (either repeat or give multimodal so that their identification is confirmed).

2.2 Adaptive Salience

The requirements for tactile salience are therefore complex and situation-dependent. We have identified the need for technology that meets requirements for adaptable

saliency [20]. In visual saliency, the spotlight of attention is based on both exogenous properties of the object (bottom-up saliency) as well as cognitive processes (top-down) [21]. Top-down attention is usually associated with the completion of a specific task, while bottom-up saliency is a fixed characteristic of the stimuli and does not vary with changes in task or situation [22]. Similarly, we can expect that some aspects of tactile saliency will be naturally salient, and other aspects will vary as a function of context (e.g., noise, competing demands for attention).

The concept of saliency can contribute to perception and effectiveness, particularly when adapted to dynamic situations. Tactile saliency can represent priority, such that more salient signals communicate higher importance or urgency [23]. A signal could start at a low level and move to higher saliency if not attended to, or if priority increases.

While single factor cueing can vary in saliency depending on factor characteristics and context, the issue of saliency becomes more complex when using multi-factor arrays which can vary along a number of features. Controlled comparisons showed some patterns to be more salient than others [11]. In addition to factor characteristics and context, multi-factor arrays vary along additional dimensions that can also affect saliency. For example, these arrays may communicate a pattern through sequential activation—such that an illusion of movement can be communicated (e.g., activating factors around the torso would be felt as a circling movement, in the same way that sequential activation of visual cues can create an illusion of movement—e.g., neon signs). Other patterns may be based on simultaneous activation of factors, which may be a single burst, or repeated. It's been shown that tempo of sequential activations create “melody” type of sensations that are easily recognized and distinguished based on their rhythmic features [24]. Simply changing the frequency of activation from slower to faster can change perceptions of urgency [23]. Patterns may be communicated across multiple body locations, such that an additional cue in a particular location can increase saliency and indicate urgency.

As understanding accumulates, tactile patterns can be developed to different levels of saliency, and to automatically change levels of saliency as the situation changes, or until an action is taken. From a user perspective, operators have indicated they would want the ability to control, to some degree, the saliency of incoming communications, and be able to “turn down the volume” when the operator is in quiet conditions, while turning it up as needs arise.

2.3 Design Guidance for Tactile Displays

Saliency will vary depending on the specific stimulus characteristics associated with the task at hand. For example, factors found to be effective when users are standing

still have been found to be much less effective when the users are engaged in strenuous activity [25], whereas factors more specifically engineered for salience were more readily perceived regardless of activity. As another example, a tap on the shoulder from a tactile navigational aid can mean different things to a helicopter pilot than to an infantryman clearing an urban area.

Ideally, tactile displays should be salient and readily interpretable. For example, localized mechanical vibration to the front of the torso will probably be salient because it is not usually experienced. Saliency can be increased further by using a vibration frequency that maximally stimulates the Pacinian corpuscles. Even though deep in abdominal tissue, the exquisite sensitivity of the Pacinian receptors to high frequency (~ 250 Hz) stimuli allows appropriate vibration stimuli to be detected readily [26]. The interpretability of tactile displays can be ensured by intuitive somatotopic mapping of direction and spatial orientation information [27, 28]. Usable stimulation patterns can be generated by placing multiple factors on many different body sites and providing a wide range of tactile stimuli, such as vibration, temperature, and light touch.

The engineering challenge is to develop efficient actuators that can display such a wide range of touch qualities in a small, lightweight package that is not adversely affected by variations in mounting or skin attachment. To preserve salience, input characteristics should be adjusted as changes occur in external factors. This adjustment is similar to environmental or cognitive forms of masking, or habituation, except that less is known about tactile masking.

Tactile displays have typically been implemented as an “artificial” language that has no natural relationship between the spatial and temporal elements of the original information and the display output. To decrease the required training time and increase the probability of correct interpretation, the language needs to be related to the user’s task and be intuitive. An example of an intuitive set of tactile commands was the TACTICS system developed by UCF and EAI [9]. Tactile messages were created using four standard Army arm and hand signals, as described in the Army Field Manual FM 21–60 chapter 2. The five signals chosen for the experiment were, “Attention”, “Halt”, “Rally”, “Move Out”, and “NBC”. Overall accuracy rates depended on training but were high in spite of minimal subject training and even when tested in a stressed environment (subjects were led with full kit on an obstacle course). A key design factor that contributed to the success of these particular tactile messages was that the tactile patterns had a temporal and spatial similarity to the original visual representation of the particular hand signal. For example, the hand signal for “Rally” is a circular motion by the hand over the head. Similarly, the tactile signal for “Rally” is a circular motion around the torso. The hand signal for “NBC” is for arms and hands to touch the left and right side of the head. Similarly, the tactile signal for “NBC” are sharp simultaneous signals to the left and right of the torso. Given this approach to design, naïve participants were able to correctly guess the meaning of each signal 51 % of the time, and a group given five minutes of training were accurate 75 % across all five signals. Thus, it is clear that the design of tactile hand-signal messages should attempt to use the visual reference where relevant.

3 Experiments

In addition to human-human communications, tactile displays can be used to enhance human-robot critical communications. A recent experiment [11] explored this concept in two ways. Several multi-factor cues were developed that varied in factor and factor pattern features, to represent alerts and status updates that could be received by a robotic squad member. In one approach, experimenters used paired comparisons with forced-choice and independent scaled ratings of various multi-factor patterns. Results showed significant differences due to factor design characteristics, such that the EAI C-3 actuator was consistently perceived as more salient. However, participants also had no problem perceiving the EAI EMR factors, which have the advantage of a low frequency stimulus, low acoustic signature, and higher efficiency regarding power usage. Significant differences in salience were also associated with different factions, such that some factions were more salient. In addition interaction effects suggest that differences in salience due to factor type can vary among different factions. While forced-choice approaches to measurement of salience can reliably identify differences, the number of paired comparisons becomes quite large when comparing more than a few factions; thus the use of independent ratings based on 5 pt response scales offer a more efficient approach to measurement.

While the investigation above was based on participants who were stationary, a corresponding effort presented factions while participants were on the move, during night operations. For this experiment, participants received two types of signals. Navigation direction cues were presented using lower frequency EMR factors, which continually indicated the direction to the next navigation waypoint, and also indicated when participants needed to avoid an exclusion zone. During this navigation task (e.g., 3 waypoints, totaling 900 m), participants also received four different factions indicating threat or status updates. Participants indicated each time they perceived an incoming factor and identify the nature of the communication. Responses were 93 % accurate, which was very high given that they were given each signal only once, with no warning signal and no repeats.

4 Discussion

Tactile displays have been shown to enhance performance and reduce workload across a range of performance domains. In this report we discuss some fundamental issues with regard to the conceptualization, measurement, and usefulness of tactile salience. Recent experiments with Soldiers proved that tactile displays, engineered to present salient signals, can significantly enhance navigation and critical communication. This underlines the need for systematic investigations of tactile salience, its measurement, and moderators. We propose a framework to guide such investigations. It is clear that the design of tactile display systems must achieve

effective levels of salience. As a further step, tactile displays can be engineered to be adaptive, such that levels of salience can adapt to levels of operator activity, environmental factors, and task demands. Salience can thus increase to demand an operator response, or decrease, either automatically or upon request.

5 Conclusion

The design of tactile displays should be guided by systematic consideration of operator characteristics, environmental and task demands, and technology options, in order to achieve necessary levels of tactile salience for that situation—one size does not fit all. The notion of adaptive salience is predicted to further enhance the contribution of tactile cueing in dynamic circumstances.

Acknowledgments Acknowledgements go to Drs. Michael Barnes and Susan Hill of the US Army Research Laboratory's Human Research and Engineering Directorate for support of efforts described in this report, as part of ongoing multiyear investigations of human-robot interaction.

References

1. Chiasson, J., McGrath, B., Rupert, A.: Enhanced situation awareness in sea, air, and land environment. In: Proceedings of NATO RTO Human Factors & Medicine Panel Symposium on "Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures," No. TRO-MP-086, pp. 1–10. La Coruña, Spain, Available at: http://www.faa.gov/library/online_libraries/aerospace_medicine/sd/media/MP-086-32.pdf. (2002)
2. Elliott L., Redden, E.: Reducing workload: a multisensory approach. In: P. Savage-Knepshield (ed.) *Designing Soldier Systems: Current Issues in Human Factors*. Ashgate. (2013)
3. Prewett, M., Elliott, L., Walvoord, A., Covert, M.: A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* **42**(1), 123–132 (2012)
4. Self, B., van Erp, J., Eriksson, L., Elliott, L.R.: Human factors issues of tactile displays for military environments. In: van Erp, J., Self, B.P. (eds.) *Tactile Displays for Orientation, Navigation, and Communication in Air, Sea, and Land Environments*. NATO. RTO-TR-HFM-122 (2008)
5. Van Erp, J.B.F., Self, B.P.: Tactile displays for orientation, navigation, and communication in air, sea, and land environments. Final RTO Technical Report RTO-TR-HFM-122. North Atlantic Treaty Organization Research and Technology Organization. (2008)
6. Geldard, F.A.: Adventures in tactile literacy. *Am. Psychol.* **12**, 115–124 (1957)
7. Spirkovska, L.: Summary of Tactile User Interfaces Techniques and Systems, NASA Ames Research Center, Report NASA/TM-2005-213451 (2005)
8. Redden, E.: Virtual Environment Study of Mission-Based Critical Informational Requirements: ARL-TR-2636. US Army Research Laboratory, Aberdeen Proving Ground, MD (2002)
9. Gilson, R., Redden, E., Elliot, L. (eds.): *Remote Tactile Displays for Future Soldiers*, ARL-SR-0152. Army Research Laboratory, Aberdeen Proving Ground, MD (2007)

10. Pomranky-Hartnett, G., Elliott, L., Mortimer, B., Mort, G., Pettitt, R., Zets, G.: Soldier-based assessment of a dual-row tactor display during simultaneous navigational and robot-monitoring tasks. ARL-TR-7397. US Army Research Laboratory, Aberdeen Proving Ground, MD (2015)
11. Elliott, L., Mortimer, B., Cholewiak, R., Mort, G., Zets, G., Pomranky-Hartnett, G., Pettitt, R., Wooldridge, R.; Saliency of tactile cues: an examination of tactor actuator and tactile cue characteristics. ARL-TR-7392. US Army Research Laboratory, Aberdeen Proving Ground, MD (2015)
12. Borji, A., Sihiti, D., Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Trans. Image Process.* **22**(1), 55–69 (2013)
13. Frintop, S., Rome, E., Christensen, H.: Computational models of visual selective attention and their cognitive foundations: a survey. *ACM Trans. Appl. Percept.* **7**(1), 1–46 (2010)
14. Mortimer, B., Zets, G., Mort, G., Shovan, C.: Implementing effective tactile symbology for orientation and navigation. In: *Proceedings of the 14th International Conference on Human Computer Interaction, HCI, Orlando, FL* (2011)
15. Cholewiak, R., Wollowitz, M.: The design of vibrotactile transducers. In: Summers, R. (ed.) *Tactile Aids for the Hearing Impaired*. Whurr Publishers, London (1992)
16. Jones, L., Sarter, N.: Tactile displays: guidance for their design and application. *Hum. Factors* **50**, 90–111 (2008)
17. Mortimer, B., Zets, G., Cholewiak, R.: Vibrotactile transduction and transducers. *J. Acoust. Soc. Am.* **121**, 2970 (2007)
18. Cholewiak, R., Collins, A.: Sensory and physiological bases of touch. In: Heller, M., Schiff, W. (eds.) *The Psychology of Touch*, pp. 23–60. Lawrence Erlbaum Associates, Hillsdale, NJ (1991)
19. Hogan R.: Personality and personality measurement. In: Dunnette, M.D., Hough, L.M. (eds.) *Handbook of Industrial Psychology*. Consulting Psychologists Press, Palo Alto, CA (1991)
20. Hancock, P.A., Lawson, B.D., Cholewiak, R., Elliott, L., van Erp, J.B.F., Mortimer, B.J.P., Rupert, A.H., Redden, E., Schmeisser, E.: Tactile Cueing to Augment Multisensory Human-Machine Interaction, *Engineering in Design*, pp. 4–9 (2015)
21. Itti, L., Koch, C.: Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2**(3), 194–203 (2001)
22. Navalpakkam, V., Itti, L.: Modeling the influence of task on attention. *Vision. Res.* **45**(2), 205–231 (2005)
23. White, T., Krausman, A.: Effects of inter-stimulus interval and intensity on the perceived urgency of tactile patterns. *Appl. Ergon.* **48**, 121–129 (2015)
24. Brown, L., Brewster, S., Purchase, H.: Tactile crescendos and sforzandos: applying musical techniques to tactile icon design. In: *CHI'06 Extended Abstracts on Human factors in Computing Systems*, pp. 610–615 (2006)
25. Redden, E.S., Carstens, C.B., Turner, D.D., Elliott, L.R.: Localization of tactile signals as a function of tactor operating characteristics. ARL-TR-3971. Army Research Laboratory (US), Aberdeen Proving Ground (MD) (2006)
26. Gescheider, G., Bolanowski, S., Hall, K., Hoffman, K., Verrillo, R.: The effects of aging on information processing channels in the sense of touch: 1. Absolute sensitivity. *Somatosens. Motor Res.* **11**, 4 (1994)
27. van Erp, J.B.F., Tactile navigation display. In: Brewster, S., Murray-Smith, R. (eds.) *Haptic HCI 2000, LNCS 2058*, pp. 165–173 (2001)
28. van Erp, J.: Tactile displays for navigation and orientation: perception and behavior. Mostert and van Onderen, Leiden, The Netherlands (2007)

An Initial Investigation of Exogenous Orienting Visual Display Cues for Dismounted Human-Robot Communication

Julian Abich IV, Daniel J. Barber and Linda R. Elliott

Abstract The drive to progress dismounted Soldier-robot teaming is toward more autonomous systems with effective bi-directional Soldier-robot dialogue, which in turn requires a strong understanding of interface design factors that impact Soldier-robot communication. This experiment tested effects of various exogenous orienting visual display cues on simulation-based reconnaissance and communication performance, perceived workload, and usability preference. A 2×2 design provided four exogenous orienting visual display designs, two for navigation route selection and two for building identification. Participants' tasks included signal detection and response to visual prompts within a tactical multimodal interface (MMI). Within the novice non-military sample, results reveal that all display designs elicited low perceived workload, were highly accepted in terms of usability preference, and did not have an effect on task performance regarding responses to robot assistance requests. Results suggest inclusion of other factors, such as individual differences (experience, ability, motivation) to enhance a predictive model of task performance.

Keywords Human-robot interaction • Human-robot teams • Multimodal communication • Exogenous orientation • Visual displays

J. Abich IV (✉) · D.J. Barber
Institute for Simulation and Training, University of Central Florida, Orlando,
FL 32826, USA
e-mail: jabich@ist.ucf.edu

D.J. Barber
e-mail: dbarber@ist.ucf.edu

L.R. Elliott
Army Research Laboratory, Human Research and Engineering Directorate Field Element,
Ft. Benning, GA 31905, USA
e-mail: linda.r.elliott.civ@mail.mil

1 Introduction

A picture is worth a thousand words...but at what cost? Visual displays offer the ability to convey a vast amount of information in a temporally succinct manner within a relatively economical real-estate [1]. The manner in which content is presented on a visual display can greatly influence the amount of attention allocation required to scan and process the information [2], having an impact on task performance [3], perceived workload [3], and usability preference [4]. Orienting of visual attention is governed by an interaction between internal (endogenous) goals and external (exogenous) demands [5]. Endogenous mechanisms involve “more purposeful and effortful orienting” (e.g. searching for a tumor in an x-ray image), while exogenous mechanisms are “triggered reflexively by a salient sensory event” (e.g. attending to a pulsating light) [6]. The focus here is on designing visual cues that trigger exogenous orientation within a visual display to facilitate the conveyance of information from a robot to a human teammate.

As the implementation of unmanned systems, such as autonomous robots, continues to transcend operational domains, specifically military, it is imperative to investigate the factors that contribute to effective human-robot communication (HRC) [7]. The military has a growing interest in equipping dismounted Soldiers (i.e. Soldiers conducting operations on foot) with personal mobile technology [8], which can serve as a potential interface for HRC [9]. The major caveat is that visual displays will allocate visual attention away from other tasks, therefore visual displays must be designed to present pertinent information in a way that supports timely information gathering without adding to the cognitive burden of other tasks, ultimately enabling efficient human-robot team communication.

1.1 *Experimental Aim*

The purpose for this experiment was to explore unidirectional communication transaction from a robot to a dismounted teammate through a visual display during a surveillance task. The specific goal was to investigate the effects of exogenous orientating visual display cues within robot-generated assistance request reports on participant’s task performance, perceived workload, and usability preference. Specifically, this study explored how the manner in which exogenous orienting visual cues are designed affect the (1) task performance regarding response to visual reports, (2) level of perceived workload while interacting with a visual display and (3) ratings for usability preference. The exploratory approach taken utilized highlighting as an exogenous cue in two types of visual reports which required human intervention to support robot decision-making: (1) choose a navigational route based on environmental features and (2) identify the correct target (i.e. building) to monitor for potential threats. Additionally, using a non-military novice sample provides a comparison group for a follow-on study with a military-based sample to evaluate population differences regarding responses to the above stated factors.

2 Methods

2.1 Participants

Fifty-six university students, 33 males and 23 females (mean age = 20.5, $SD = 3.9$) participated in the experiment. Participation in the experiment was completely voluntary and compensation was awarded in the form of class credit. Participants all had normal or corrected to normal vision. None of the participants had prior experience with the simulated environment or multimodal interface. The responses to the demographic questionnaire indicate participants were familiar with video games (e.g. first person shooter and multiplayer games) but had little knowledge about robotics technology. Almost all participants had no prior military experience except for three males (two were National Guardsmen and one was in the U.S. Air Force Reserve Officer Training Corps).

2.2 Experiment Task

Apparatus. The simulation was presented using a standard desktop computer or laptop with equal specifications (3.2 GHz, Intel Core i7 processor) connected to a 22" (16:10 aspect ratio) monitor. Responses to the tasks were collected using the left mouse button and keyboard.

Simulation. The experimental scenario was performed within the Mixed Initiative eXperimental test bed (MIX) [10]. The scenario simulated a reconnaissance and surveillance task where the Soldier and robot were traveling along separate non-overlapping routes within the same urban environment.

Threat Detection (TD) Task. The MIX test bed was customized to represent the first-person perspective of a Soldier traveling through a generic Middle Eastern urban environment (Fig. 1, left). The Soldier's route was pre-planned and did not require the participants to control the Soldier's movement. The participants' primary role was to identify potential threats that populated the environment (i.e. a signal detection task) by capturing photos to help populate a robot teammate's database with examples so it can more effectively carry out mission tasks autonomously. There were four categories of characters (i.e. events) within the environment: friendly Soldiers, friendly civilians, enemy Soldiers, and insurgents (Fig. 1, right), and each category included at least five different types of characters. Enemy Soldiers and insurgents were classified as threats (i.e. signals) and an equal number of each were presented. Equal number of each category of non-threats were also present. All threats were identified by left-clicking with a mouse on the character within the environment. No feedback was provided regarding accuracy of detection, but participants did hear the sound of a camera shutter to indicate they were capturing photos. The event rate of the TD task was a presentation of 30 per minute with a signal probability of 13.33 % based on previous research [11].



Fig. 1 (*Left*) This is a screenshot of the MIX test bed representing the first person perspective of the Soldier traveling through the Middle Eastern urban environment. (*Right*) Example of characters used within the threat detection task. From *left to right*: friendly soldier, friendly civilian, enemy soldier, and insurgent (armed civilian)

Performance on the TD task was assessed as the overall percentage of correct responses.

Robot Assistance (RA) Task. While participants were performing the TD task, visual prompts randomly appeared and required participants to respond to RA queries generated by a robot teammate. The visual prompts were a virtual representation of the visual display component of a Multimodal Interface (MMI) for human-robot communication (Fig. 2). The contextual application told to participants was the robot was traveling a separate route and may need assistance if it cannot deduce the best option based on its limited intelligence capabilities. Participants responded to reports by left-clicking on one of the buttons located at the top of the screen indicating their choice to the robot's query (Fig. 2). No feedback was provided to the participants regarding accuracy of choices made. Performance was assessed as providing the correct response to the RA queries and as the average reaction time to respond.

RA Reports. The types of assistance the robot requested was regarding (1) navigational routes and (2) building identification (Figs. 3 and 4). The navigation assistance requests asked participants to decide the best travel route for the robot to avoid obstacles. The building identification assistance requests asked participants which building a robot should monitor. The information needed to make a decision for the robot was gathered from the MMI. Every MMI prompt was comprised of two images: (1) the right image represented the point-of-view (POV) of the robot traveling through the environment and (2) the left image represented an aerial view of the operational area (Fig. 2). Both images represented the same scene but from different angles and therefore, the RA queries could be answered by gathering the pertinent information from either image. The research interest here lies in evaluating the response performance, workload impact, and preference of visual display format by comparing the types of display formats for each query (i.e. Navigation A vs. B & Building A vs. B).

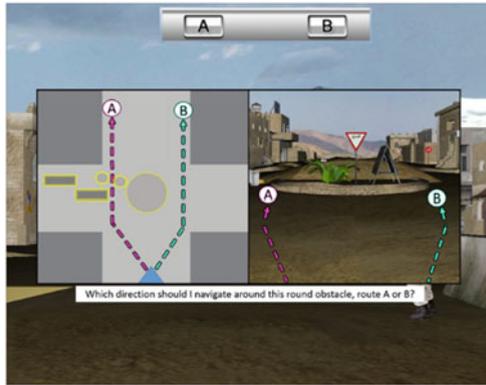


Fig. 2 This image represents the virtual multimodal interface as a prompt on the screen within the MIX environment

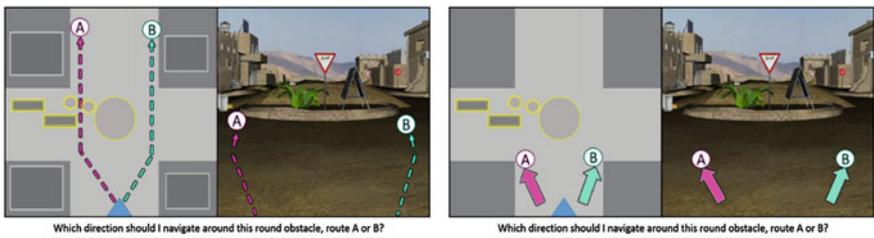


Fig. 3 These two images represent the navigation robot assistance query display formats presented in the MMI. The image on the *left* has extended directional arrows (*Navigation A*). The image on the *right* has short **bold** directional arrows (*Navigation B*)

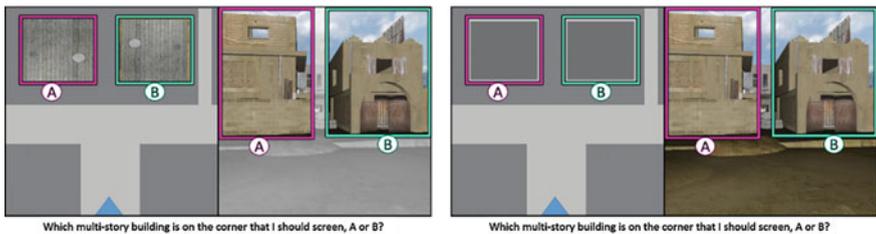


Fig. 4 These two images represent the building robot assistance query display formats presented in the MMI. The image on the *left* has highlighted boxes around the buildings (*Building A*). The image on the *right* has highlighted boxes around the buildings with modifications to the scene and aerial images (*Building B*)

2.3 Questionnaires

Demographics. This questionnaire was developed in-house and gathered background information regarding age, gender, visual acuity, academic education, military experience, computer use, video game exposure/experience, and robotics knowledge.

NASA-Task Load Index (TLX). The TLX [12] is comprised of six perceived workload subscales. Each subscale rating used a 100-point sliding scale with five-point increments. The unweighted-TLX was administered by computer at the end of each block throughout each scenario.

System Usability Scale (SUS). This 10-item questionnaire focused on perceived usability of the system (i.e. hardware, software, equipment) [13]. Ratings were indicated using 5-point Likert items. The composite score of the items provided a single number representing the overall usability of the system and ranged from 0 to 100. These questions were focused on the interaction with the device during the scenario and was administered by computer at the end of each block.

Free Response Questionnaire. This 6-item open-ended questionnaire covered positive and negative aspects of the participant's interaction with the simulated multimodal interface device, their preference for display design, and suggestions for improvement. These questions were focused on the interaction with the device during the scenario and was presented to participants after completing the whole scenario.

2.4 Procedure

When participants arrived they were instructed to first read over the informed consent. Upon consent, they then completed the demographics form and cube comparisons test. Task training followed completion of the pre-study questionnaires. An 18-min narrated PowerPoint presentation was used to support consistency of training for each participant. Training was accomplished in two phases. The first phase was 12-min and it instructed participants about the continuous task they were to perform (i.e. TD task) and how to respond to the TLX. They were then given the chance to ask any questions pertaining to the task, practice performing the task for about 1-m and responding to the TLX. The next phase continued in the same format. The second phase was 6-min and concentrated on explaining the purpose and elements of the four different RA query display formats. It also provided instructions on how to respond to the usability and free response questionnaires. They were then given the chance to ask any questions pertaining to the task and practice responding to two display formats of RA queries to clearly illustrate the way they would be presented and to make sure they understood how to respond to the queries.

3 Results

Paired-sampled *t*-tests were used to assess the impact of various RA query designs on task performance, perceived workload, and usability preference. Cohen’s *d* effect sizes using the conventional scale of 0.2, 0.5, & 0.8 (small, medium, & large, respectively) are also reported for specific comparisons of RA query designs. A bivariate correlation was run to find relationships among demographic characteristics, performance, perceived workload, and usability preference. After removing outliers the sample size for all analyses was $n = 49$.

3.1 TD Performance

Two *t*-tests were run to assess the effects of responding to navigation and building RA report design separately on TD task performance accuracy. Results show a significant difference was found when responding to navigation RA reports, but not for building reports ($p > 0.15$; Table 1).

3.2 Robot Assistance (RA) Performance

A series of *t*-tests were run to assess the effects of RA report designs on RA response accuracy and response time. Results show no statistically significant results were found for the difference in percent of correct responses and for reaction time to RA reports (Table 2).

3.3 TLX

A series of *t*-tests were run to assess the effects of RA report designs on perceived mental workload. Results show no statistically significant results were found for

Table 1 Results of *t*-tests for threat detection task performance for both navigation and building display designs

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> -test	<i>d</i>
Threat detection accuracy (% correct)	<i>Navigation A</i>		<i>Navigation B</i>			
	95.24	9.07	97.28	7.00	-2.274*	0.25
	<i>Building A</i>		<i>Building B</i>			
	95.69	8.43	97.51	6.52	-1.478	0.24

*Note: $p < 0.05$

Table 2 Results of *t*-tests for robot assistance response accuracy and reaction time for both navigation and building display designs

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> -test	<i>d</i>
Correct response accuracy (% correct)	<i>Navigation A</i>		<i>Navigation B</i>			
	62.25	20.17	64.03	20.35	-0.516	0.09
	<i>Building A</i>		<i>Building B</i>			
	52.04	23.15	58.67	22.13	-1.139	0.29
Average reaction time (s)	<i>Navigation A</i>		<i>Navigation B</i>			
	4.58	2.64	4.00	1.63	1.308	0.26
	<i>Building A</i>		<i>Building B</i>			
	3.99	2.21	4.61	2.89	-1.271	0.24

Note: $p > 0.15$

Table 3 Results of *t*-tests for global workload for both navigation and building display designs

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> -test	<i>d</i>
Global TLX ratings	<i>Navigation A</i>		<i>Navigation B</i>			
	17.24	17.43	16.56	15.62	0.599	0.04
	<i>Building A</i>		<i>Building B</i>			
	16.75	17.32	16.58	15.31	0.170	0.01

Note: $p > 0.17$

either display type ($p > 0.17$). Only global workload means and standard deviations are reported because they show the trend for all subscales (Table 3).

3.4 SUS

Two *t*-tests were run to assess the effects of responding to navigation and building RA report design separately on usability preference. Results show no significant difference was found for either navigation or building display formats separately (Table 4).

Table 4 Results of *t*-tests for usability preference ratings for both navigation and building display designs

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> -test	<i>d</i>
SUS ratings	<i>Navigation A</i>		<i>Navigation B</i>			
	82.53	13.70	82.00	15.70	0.512	0.04
	<i>Building A</i>		<i>Building B</i>			
	82.59	14.27	82.45	14.45	0.185	0.01

Note: $p > 0.6$

3.5 Free Response Questionnaire

Three independent raters conducted an evaluation of the data for common themes across participant responses to each item. Raters began by organizing the textual responses into common ideas and then comparing the frequency with which certain themes occurred in the text for each item. Common themes that were identified as being mentioned by at least three respondents for each item were retained. The three assessments were then compared for overlapping patterns identified by all three raters. Table 5 shows the results from the assessment. Within the table are the common themes identified, the number of participants that made that comment, and an example response they provided.

3.6 Bivariate Correlations

Bivariate correlations were run to find relationships among demographics characteristics, task performance, perceived workload, and usability preference. To clarify the significant variables listed in Table 6 refer to how often participants use a computer, how much experience they have working with any type of video game, reaction time when responding to the RA reports, global TLX associated with responding to RA reports, and RA report design usability preference. The RA response accuracy and reaction time, global TLX and SUS ratings were averaged across both display format types separately since the *t*-test showed no significant differences.

Table 5 Table of common themes among participant responses to the Free Response Questionnaire regarding the robot assistance display formats

Item/question	Theme	Example response
Positive aspects of the device	(a) Ease of use (28) (b) Multiple views (10) (c) Representation of environment	(a) It is straightforward and easy to use (b) There was more than one view to make decisions with (c) The two views give a clear representation to make decisions
Negative aspects of the device	(a) Diversion (12) (b) Ambiguous (10)	(a) Device requires full attention of the user (b) The information was not detailed enough
Navigation information	(a) No preference for layout (9) (b) Aerial view preference (17)	(a) It did a really good job at providing information (b) Both views provided info but the aerial view was most helpful
Building identification	(a) Clarity of information (7) (b) Ease of use	(a) All of the information was available in a clear view (b) It was surprisingly easy to use

The numbers in parentheses next to each theme represent the number of participants that reported each theme. An example of each theme response is provided

Table 6 Bivariate correlation matrix for statistically significant relationships among demographics, task performance, global workload, and usability preference

	1	2	3	4	5	6
1. Video game experience						
2. First person shooter game experience	0.813**					
3. Multiplayer video game experience	0.794**	0.879**				
4. Average correct RA response for Navigation display	0.363*	0.414**	0.374**			
5. Average correct RA response for Building display	-0.287*	-0.343*	-0.273	-0.285*		
6. SUS average rating for Navigation display	0.363*	0.343*	0.278	0.147	0.066	
7. SUS average rating for Building display	0.332*	0.296*	0.191	0.177	0.115	0.956**

Note: * $p < 0.05$, ** $p < 0.01$

The correlation matrix shows relationships among video game experience, including first person shooters and multiplayer games, seemed to correlate positively with SUS ratings and correct response to RA Navigation queries, but negatively with correct response to RA Building queries.

4 Discussion

The goal for the study was to evaluate the effects of various exogenous orientation visual cues within robot-generated visual reports on a mobile-based platform to convey squad-level information to a teammate in a dismounted scenario. Findings show that based on responses from a novice non-military based sample, all display design types elicited low levels of perceived mental workload, were highly accepted in terms of usability preference, and had little effect on performance related to RA report responses. Correlations reveal individual differences, such as video game experience, could factor into task performance and usability ratings.

The first finding shows that TD task performance differed when participants responded to navigation RA reports but not for building ones. This difference should not be dwelled upon because task performance for all conditions was above 95 % (Table 1). It was expected that this task would not be hindered by the interaction with the MMI since the TD task was kept at a constant event rate and in the past shown to result in the same level of performance [11]. Also, the TD task was paused when the RA reports were automatically prompted, therefore the MMI did not occlude the user's visual field while simultaneously performing the TD task.

In order to quantify the performance costs associated with the time spent viewing a visual display, future research should focus on the effects of interacting with the MMI while the primary task (e.g. TD task) is continuously occurring.

Looking next at the performance associated with responding to the RA reports shows no differences were found for performance accuracy or response time. Although on average it took participants about 4.3 s to respond to any type of report, the average accuracy of their responses was just above 50 % (Table 2). This raises concern that participants were either unable to decipher the correct answer based on the visually displayed information or they may simply have been guessing when presented with the RA generated queries. Since there is a difference in their navigation and building response accuracy, with the former having a higher accuracy (average 63 %), it can be argued that participants were contributing a level of effort as indicated by their TLX ratings. The detailed results of the TLX subscales were not reported but effort and mental demand were the two largest factors associated with both display types (i.e. navigation and building reports). Performance could have also been effected by the lack of feedback regarding their response to the RA queries. Although the lack of feedback was ecologically valid, a follow up study might want to assess the effects of immediate feedback regarding response accuracy since it has been shown to impact task performance [13].

The subjective measures will be discussed together. Overall, all of the display designs elicited low perceived workload, in accordance with past research that found low cognitive demands associated with exogenous attention [14]. Contrastingly, the usability preference for all display designs was high. Indicating, any design could be implemented and the outcome would be the same regarding task performance, perceived workload, and usability preference. Hence, deciding which of the exogenous orienting visual cues for navigation route selection or building identification should be generated could alternatively be based on computer processing capabilities. Meaning, if the robot is processing large amounts of intelligence and sensory data but needs to send a report to a Soldier teammate then the most efficient, or easiest to process programmatically, way to display the cues should be executed.

Theoretical contributions indicate that perceived assessment of subjective state might not be fully indicative of task performance, and therefore interface design evaluations should not rely solely on subjective measures. Practical applications show that various interface display design types could be generated to provide the same type of information to a human teammate, therefore if other constraints such as battery life or processing power is limited, the interface display design that is more energy efficient could be utilized. Additionally, individual differences in experience and exposure to computer technology can moderately affect workload and usability responses during interaction through computerized visual displays indicating a potential selection criteria for human-robot team personnel.

Acknowledgments This research was sponsored by the U.S. Army Research Laboratory (ARL) and was accomplished under Cooperative Agreement Number W911NF-10-2-0016. The views and conclusions contained in this document are those of the authors and should not be

interpreted as representing the official policies, either expressed or implied, of ARL or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

1. Chittaro, L.: Visualizing information on mobile devices. *Comp.* **39**(3), 40–45 (2006)
2. Atkinson, R., Holmgren, J., Juola, J.: Processing time as influenced by the number of elements in a visual display. *Percept. Psychol.* **6**(6A), 321–326 (1969)
3. Chao, C., Lin, C.-H., Hsu, S.-H.: An assessment of the effects of navigation maps on driver's mental workloads. *Percept. Motor Skills* **118**(3), 709–731 (2014)
4. Glumm, M., Branscome, T., Patton, D., Mullins, L., Burton, P.: The effects of an auditory versus a visual presentation of information on soldier performance. *Hum. Fact. Ergo. Soc. 43rd Anl. Mtng.* **43**, 1041–1043 (1999)
5. Berger, A., Henik, A., Rafal, R.: Competition between endogenous and exogenous orienting of visual attention. *J. Exp. Psych. Gen.* **134**(2), 207–221 (2005)
6. Hopfinger, J., West, V.: Interactions between endogenous and exogenous attention of cortical visual processing. *NeuroImage* **31**(1), 774–789 (2006)
7. Robotics Collaborative Technology Alliance [RCTA]: FY 2014 Annual program plan. Contract No. W911NF-10-2-0016 (2014)
8. Young, R.: Federal mobile computing summit. Retrieved from [www.fedsummits.com: http://www.fedsummits.com/wp-content/uploads/2014/09/Young-slides-Mobile-Summit-2014-08-201.pdf](http://www.fedsummits.com/wp-content/uploads/2014/09/Young-slides-Mobile-Summit-2014-08-201.pdf) (2014)
9. Barber, D., Abich, J.I., Phillips, E., Talone, A., Jentsch, F., Hill, S.: Field assessment of multimodal communication for dismounted human-robot teams. In: Proceedings of the 59th Human Factors and Ergonomics Society Annual Meeting (2015)
10. Reinerman-Jones, L., Barber, D., Lackey, S., Nicholson, D.: Developing methods for utilizing physiological measures. In: Tadeusz, M., Waldemar, K., Rice, V. (eds.) *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations*. CRC Press, Boca Raton (2010)
11. Abich, J.I., Reinerman-Jones, L., Taylor, G.: Establishing workload manipulations utilizing a simulated environment. In: Shumaker R. (ed.) *Virtual, Augmented and Mixed Reality. Systems and Applications: 5th International Conference, VAMR 2013, Held as Part of HCI International 2013*, vol. 8022, pp. 211–220. Springer, Berlin (2013)
12. Hart, S., Staveland, L.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (eds.) *Human Mental Workload*. North Holland Press, Amsterdam (1988)
13. Brooke, J.: SUS—A quick and dirty usability scale. *Usab. Eval. Ind.* **189**(194), 4–7 (1996)
14. Blank, L., Cohen, J.: Feedback improves performance: validating a first principle. *Arch. Ped. Adol. Med.* **161**(1), 103–104 (2007)

Five Requisites for Human-Agent Decision Sharing in Military Environments

Michael Barnes, Jessie Chen, Kristin E. Schaefer, Troy Kelley,
Cheryl Giammanco and Susan Hill

Abstract Working with industry, universities and other government agencies, the U.S. Army Research Laboratory has been engaged in multi-year programs to understand the role of humans working with autonomous and robotic systems. The purpose of the paper is to present an overview of the research themes in order to abstract five research requirements for effective human-agent decision-making. Supporting research for each of the five requirements is discussed to elucidate the issues involved and to make recommendations for future research. The requirements include: (a) direct link between the operator and a supervisory agent, (b) interface transparency, (c) appropriate trust, (d) cognitive architectures to infer intent, and e) common language between humans and agents.

Keywords Autonomy · Intelligent agent · Human agent teaming · Decision making

M. Barnes (✉) · J. Chen · K.E. Schaefer · T. Kelley · C. Giammanco · S. Hill
U.S. Army Research Laboratory, Human Research and Engineering Directorate,
Aberdeen Proving Ground, Maryland, USA
e-mail: michael.j.barnes.civ@mail.mil

J. Chen
e-mail: yun-sheng.c.chen.civ@mail.mil

K.E. Schaefer
e-mail: kristin.e.schaefer2@mail.mil

T. Kelley
e-mail: troy.d.kelley6.civ@mail.mil

C. Giammanco
e-mail: cheyl.a.giammanco.civ@mail.mil

S. Hill
e-mail: susan.d.hill.civ@mail.mil

1 Introduction

Combat by its very nature is volatile and unpredictable. Greater use of autonomous systems promises to change the battlefield equation. Fewer soldiers exposed to enemy fire, larger force to soldier ratios, and improved capabilities are some of the objectives of the U.S. Department of Defense's autonomy research programs [1–4]. An important advantage of autonomy is the possibility of improving friendly force ratios by increasing the number of unmanned vehicles (UVs) controlled per operator. Unfortunately, the first generation of UVs has not reduced the ratio of soldiers to systems. In many instances, the results have been quite the opposite: the Shadow unmanned aerial vehicle (UAV) requires a crew of 2–3 operators per UV and also requires a larger force for maintenance and security [4]. Even with increased autonomy, the ratio of systems to humans is not always favorable because of human cognitive limitations and degradations due to the “fog of war” [2–5]. Lewis [6] reviewed various UV control approaches that use scheduling algorithms based on the characteristics of the UV (average neglect time without negative consequences) and the operator (average time the operator interacts with an UV). The metrics' purposes are to develop a method of optimally assigning operator attentional focus per UV and to determine a method to compute “fan-out” (limits to the number of UVs assigned per operator or crew). Even with more sophisticated assumptions about variance, individual differences, and multitasking the metrics predict performance under ideal conditions (UV independence and near uniform neglect time during a mission). In reality, uncertainty and UVs interactions with each other and with manned systems make these assumption impractical for many combat situations [4]. On the contrary, systems must be able to respond to each other and change tactics as the mission unfolds. Unfortunately, human span of control does not necessarily improve as n-systems are automated [5]. Operators must maintain overall situation awareness (SA), multitask, and respond to changes during periods of increased volatility while at the same time supervise more than one system. The purpose of this paper is to discuss the implications of a general class of mixed initiative solutions [5] that involve shared decision making between an intelligent agent (IA) and a human operator.

Our definition of an IA is a software agent that acts autonomously, responds to environmental changes, and is able to compute a course of action (COA) to achieve its objectives [7]. We are assuming a class of agents that interact with other agents including its human supervisor. However, IAs are not a solution in themselves. There are various research issues that must be addressed before human-agent teams can take full advantage of the synergy between the algorithmic precision of IAs and the flexibility of humans [5]. The problem set that U.S. Army Research Laboratory researchers are addressing include how to interact with multiple systems without overwhelming the operator, how to develop appropriate trust and transparency between the IA and its human supervisor, and how to build mutual understanding between humans and software agents that results in a human-to-human like natural dialogue.

The themes are interrelated by the assumption that an IA acts as an assistant that is able to monitor UVs thus freeing its human supervisor to focus on the combat situation. This requires transparency between the operator and the IA enabling the operator to trust the IA to perform its functions but also to know when to intervene. A transparency model and the implications of our empirical evaluations of the model are discussed as the basis for calibrating trust. However, because trust is an important variable aside from transparency issues, additional research efforts related to trust are discussed below. Transparency needs to be bi-directional to achieve a nuanced partnership between humans and agents. In this regard, we review two technical research areas that enhances the humans and agent's capability to have similar knowledge structures (cognitive architectures) and permits two-way communications that enable an IA to infer meaning based on context (controlled English).

2 IA As an Interface

Lewis [6] delineated the control space of human operators (O) to UVs as $O(1)$, $O(n)$ and $O(>n)$ or one-to-one, additive or complex (interactive elements). Chen and her colleagues proposed using an IA as an interface to dampen the effects of complex interactions so the resulting paradigm is closer to a one-to-one control environment [5, 8, 9]. They simulated a mixed initiative paradigm wherein a single agent (RoboLeader) acts as an intermediate supervisor to multiple partially autonomous UVs [5, 7–11]. The main impetus of the research is the human factors issues involved in having a single IA interface directly with the operator and also with multiple UVs. Issues explored included system factors (number of UVs, error rate and type, degrees of autonomy and mission parameters) and operator factors (multitasking and individual differences). IA errors affected performance but the effects of false alarms were minimized if the situation display allowed the operator to quickly rectify errors [7]. Not surprisingly, the advantages of the IA were more manifest when the tasking environment became more difficult and the operator was involved in multitasking [11]. Individual differences in spatial abilities, gaming experience, perceived attentional control, etc. proved to be important determinates of how operators interacted with an IA, suggesting that IAs should be tuned to individual attributes [5, 7–10]. The RoboLeader paradigm takes advantage of hierarchy among agents by letting UVs conduct their mission including interacting with both manned and unmanned elements while the supervisory IA interacts with a human to suggest changes as events unfolded [5]. For example, during the RoboLeader experiments simulating convoy operations, the operator supervised a manned vehicle, a UAV and an unmanned ground vehicle while conducting 360° threat monitoring around his or her own vehicle. The results indicated that as complexity and the possibility of RoboLeader errors increase, operator transparency into the IA's reasoning process becomes an important consideration.

3 Transparency Model

The key to developing an effective teaming relationship is a mutual understanding of each other's assumptions as well as an understanding the objectives of the mission [5]. Lee and his colleagues suggested basic tenants of automation transparency arguing the operator should understand the purpose, process and performance of the automated system in order to interact effectively with the system [12, 13]. Chen et al. developed a Situation awareness-based Agent Transparency (SAT) model that adapted concepts from Mica Endsley's SA model to fully articulate the shared knowledge necessary for humans to gain insight into the IA's purpose, planning and process [12–15]. As Fig. 1 illustrates, the SAT model posits that human SA of autonomous systems has three levels (L): L1—understanding of the IA's intentions-plan, L2—comprehension of the IA reasoning, and L3—projection of the IA's expected outcomes [14, 15]. ARL researchers were able to evaluate SAT as part of the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT) project funded by the U.S. Department of Defense's Autonomy Research Pilot Initiative program [16, 17].

The purpose of the IMPACT program is to develop a suite of intelligent systems to aid in the planning and execution of multiple heterogeneous UV missions. ARL's function is to develop and verify transparency concepts for the IMPACT program. We examined various visualization concepts and evaluated their efficacy in terms of the three SAT levels over two experiments [18, 19]. The experimental paradigm consisted of a comparison of two IA generated plans detailing the type of UVs and their routing plans during 24 base defense vignettes. Updates, during the simulated mission, supported either the IA's favored option A (in 5/8th of the vignettes) or its

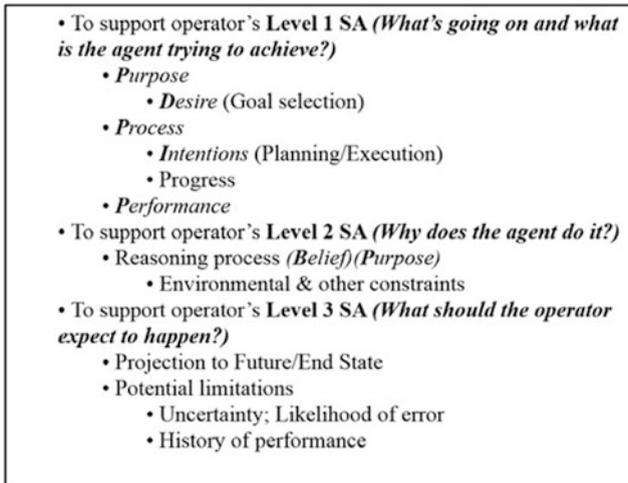


Fig. 1 The SAT model [15]

alternative option B (in 3/8th of the cases). The results were assessed in terms of correct usage (when A was the better option and correct rejections (when B was the better option).

Taken together, the research supports the SAT model's premise that adding levels of SAT of the IA's intent, reasoning, and projected outcomes help to calibrate operator reliance on courses of actions suggestions from an imperfect IA. The results also indicate that the visualization techniques do matter. The positive findings for the addition of L3 information in the second experiment were partially due to display improvements of outcome projections developed by Air Force researchers [7, 18, 19]. In summary, using an IA as an intermediate supervisor proved to be effective strategy for controlling multiple heterogeneous UVs during varied combat simulations given that the IA's intentions were transparent to the operator.

4 Trust

ARL's research on trust is connected to our research on transparency but trust is also separate research domain. Trust is the attitude of the operator towards autonomy but it also relates to an operator's behavior (calibration of reliance and compliance). There is not always a one-to-one correspondence between trust and transparency. Thus even when automated system are reliable and transparent, an operator may over-ride autonomy because of a basic mistrust in automation or conversely the operator may over-rely on autonomy because of a bias towards automation [12, 13]. An important focus of ARL researchers is to improve human-agent teaming especially for the supervision of ground robots. As robotic systems with increased intelligence and decision-making capability continue to be designed to function successfully as an integrated part of a Soldier-agent team, the issue of soldier trust becomes a key priority. Without appropriate trust, there is an increased probability that the technology will underutilized or even used inappropriately which can negatively impact the mission [12, 13, 20, 21]. The key priorities of this work began with developing a broader understanding of the trust domain, identification and evaluation of trust measurement techniques, and experimentation supporting specific Army applications.

To understand the trust and teaming problem space, large scale literature reviews and meta-analyses were conducted in the areas of human-robot interaction (HRI) [22], human-automation interaction (HAI) [23], and human-animal interaction [24]. Findings pointed to the importance of trust antecedents relating to the human, the partner (robot/agent, automation, or animal), and the environment. Overall, results of the meta-analyses provided insight into the known variables of importance (e.g., reliability of the system), as well as a number of areas of needed exploration (e.g., human states, transparency, errors/failures, and communication).

Research experimentation in both simulation and upcoming field studies will be used to further understand the process of trust development and calibration across a

number of Army applications. This research addresses key barriers related to achieving adequate levels of Soldier-robot integration and trust through developing appropriate methods for real-world reliable communication across different types of Soldier-robot teams. Overall, the development of accurate and effective user interfaces support the development of a common ground and a shared understanding of the mission space which can then be designed to engender appropriate trust in the system. For example, previous research has shown the importance of the relationship between trust and user interfaces for both base operations and field vehicles [25–27].

5 Cognitive Architectures for IAs

As part of the DOD emphasis on increased autonomy, ARL is investigating various architectures that improve autonomous control capabilities [28]. A traditional problem with both transparency and communications between humans and IAs is the lack of a common knowledge structure. The underlying logic of an optimization algorithm may be opaque to an operator. Developing knowledge structures using cognitive architectures ameliorates problems related to the operator not understanding the underpinning of the IA’s algorithms [2, 5]. Cognitive architectures have been used successfully for robotics manipulation and control [28, 29]. The advantage of using a cognitive architecture for robotics control is that a cognitive architecture represents problem solving in much the same way that humans represent problem solving. This similarity allows for interactions with IAs to be more seamless and robust. ARL has developed the Symbolic and Sub-symbolic Robotics Intelligence Control System (SS-RICS), which is based on previous cognitive architectures [28]. To emulate human decision making, SS-RICS uses goals for the representations of behaviors. This goal structure has been particularly valuable for expressing the actions of the robot under a variety of different situations.

Our experimentation with SS-RICS continues to determine the best representation of goal based behavior. One investigation for this representation is that the robotics architecture expresses goals as generalizable behavior. For example, if the goal is presented from the operator to the robot as “Go Around the Building” the goal is represented internally to the robot as “Go Around Object” with the building being the object of interest. This can be communicated to the operator from the robot as “I am going around the building” or “I am going around the object of interest”. Either one might be appropriate depending on the situation. Additionally, we have found two other critical types of information, in addition to the goal, for reporting to an operator, so that the operator can infer the intent of the robot. These are the *time* associated with the execution of a goal, and the *strategy* used for the execution of the goal. The length of time associated with the goal may be a clue that the goal is taking too long given the current strategy and that the robot may be having problems solving the goal of going around an object.

However, there is a problem with operators understanding strategies involved with problem solving, especially if the strategy is not a typical one or is not intuitive to the situation. For example, perhaps a robot should not report that it is using the D* algorithm for planning and search since the operator might not know the advantages and disadvantages of the algorithm. The operator would have to learn the advantages and disadvantages in order to interact with the robot. Additionally, the operator needs some understanding of the strategy in order to make accurate decisions involving the strategy selection. Perhaps edge detection as a strategy takes a long time to execute? Perhaps the end state in edge detection is difficult to define? What is D*? Knowledge of strategies and how they apply in different situations is critical for the efficient execution of goal based behavior between robots and operators. Choosing effective strategies is simplified if the IA understands the operator's intent and bi-directional communications permits the IA to query the operator in order to explain the advantages of specific strategies.

6 Language Processing

Although compatibility of knowledge structures is important for communication, compatibility by itself is not sufficient for two way communications. Language does not require either text or spoken dialogue; it does requires syntax, semantics and pragmatics to convey meaning during two way communications [5]. However, the pragmatics of creating a dialogue may make open-ended natural language processing impractical. For example, the implications of intelligence means something quite different in every day speech than in military applications. The purpose of the following research is to develop a framework for language processing that translates text in terms of the context of the phrases, not simply in terms of their literal interpretations. This research is the precursor to systems that generate dialogues between IAs and humans that are sensitive to military context and objectives.

A fully realized system will mean that IAs can query its human partner and ask for clarifications as the mission unfolds, emulating the peer-to-peer relationships of soldiers engaged in combat [5]. Controlled English (CE) is a natural language representation to support human-machine interaction developed by alliance members within ARL and UK Ministry of Defence International Technology Alliance in Network and Information Sciences. CE is a human writable and machine readable representation of the English language based on domain semantics and First Order Predicate Logic. The user's conceptual model is written in CE as concepts, facts and logical inference rules representing things and their relationships, properties, and attributes.

Our research uses DELPH-IN linguistic resources, the English Resource Grammar (ERG) and Minimal Recursion Semantics (MRS) in combination with the CE system [30]. The ERG system is used to deep parse natural language sentences for syntactic structure, which is output as MRS predicates and arguments representing the linguistic semantics. The linguistic semantics are transformed into

generic semantics (situations and roles), then mapped to domain semantics for output as CE facts within the CE system. Our research with the ERG/MRS and CE systems illustrates how CE can be used as a common language to support human-machine communication and reasoning. It is important to note that CE can be used to capture rationale as a set of reasoning steps leading to inferences [31, 32]. Logical inference rules can be applied to extract CE facts and assumptions and represent them in a rationale graph or “proof table” as the reasoning steps leading to conclusions. The CE system may be used to test multiple hypotheses, detecting incompatible conclusions based on inconsistent assumptions [33]. The International Technology Alliance Collaborative Planning Model (CPM) was developed using CE to assist coalition planners in the management of dynamic plans and to facilitate automated reasoning of mundane tasks [34]. CPM enabled coalition planners to exchange task specific planning information (subplans) contained within an overall plan representation for use within various automated planning tools.

CE has also been used to develop domain models with logical inference rules for solving logic problems similar to those used in training Intelligence Analysts [31]. It is worth noting that the representation of hypotheses and assumptions using CE aids in the identification of cognitive biases such as “mirroring” in which the analyst assumes others share his world view. In our current research, we are using CE to develop a civil-military domain model for fact extraction and reasoning about civil considerations [35]. This CE domain model may be used in future research to support natural language processing and reasoning within the Security and Stability Operations Training Technology, an Intelligent Tutoring System being developed for the Civil Affairs community. However, the general utility of the research goes beyond specific applications. CE and similar paradigms form the basis for developing future intelligent teams part-human and part-artificially intelligent.

7 Conclusions

The purpose of this paper is to discuss requisites for human-agent shared decision making. We reviewed research that indicated that an IA which supervised n-less capable UVs could reduce the operator’s workload and improve overall performances. The research also indicated that a successful partnership requires that the IA’s intent, reasoning, and projected outcomes be transparent to its human supervisor. However, to be successful transparency must be bi-directional.; ARL researchers collaborating with researchers from the academic and international communities are investigating technologies whose purpose is to develop a more natural teaming relationship among IAs and humans. Research efforts discussed include better understanding of human-agent trust, IA architectures designed to emulate human cognitive process, and the use of Controlled English as a basis for a natural dialogue between humans and their IA assistants.

References

1. Department of Defense: Briefing on Autonomy Initiatives (2012)
2. Defense Science Board: Role of Autonomy in DoD Systems. Office of the Undersecretary of Defense, Washington, D.C. (2012)
3. Endsley, M.: *Autonomous Horizons: System Autonomy in the Air Force—A Path to the Future* (Volume I: Human Autonomy Teaming). US Department of the Air Force, Washington, D.C. (2015)
4. Barnes, M.J., Chen, J.Y.C., Jentsch, F., Oron-Gilad, T., Redden, E.S., Elliott, L., Evans, A.: Designing for humans in autonomous systems: Military applications. Technical report ARL-TR-6782, Army Research Laboratory (US), Aberdeen Proving Ground, Maryland (2014)
5. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multirobot control: a review of human factors issues. *IEEE Trans. Human-Mach. Syst.* 44(1), 13–29 (2014)
6. Lewis, M.: Human interaction with multiple remote robots. *Rev. Human Factors Ergon.* 9(1), 131–174 (2013)
7. Barnes, M.J., Chen, J.Y.C., Wright, J., Stowers, K.: Human agent teaming for effective multi-robot management: Effects of transparency (in press)
8. Chen, J.Y.C., Barnes, M.J.: Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics* 55, 1043–1058 (2012)
9. Chen, J.Y.C., Barnes, M.J.: Supervisory control of multiple robots; effects of imperfect automation and individual differences. *Hum. Factors* 54(2), 157–174 (2012)
10. Wright, J.L., Chen, J.Y.C., Quinn, S.A., Barnes, M.J.: The effects of level of autonomy on human-agent teaming for multi-robot control and local security maintenance. Technical report, ARL-TR-6724, U.S. Army Research Laboratory, Aberdeen Proving Grounds, Maryland (2013)
11. Wright, J.L., Chen, J.Y.C., Barnes, M.J., Hancock, P.A.: The effect of agent reasoning transparency on automation bias: an analysis of performance and decision time (in press)
12. Meyer, J., Lee, J.: Trust, reliance, compliance. In: Lee, J., Kirlik, A. (eds.) *The oxford handbook of cognitive engineering*. Oxford University Press, Oxford (2013)
13. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* 46(1), 50–80 (2004)
14. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. *Hum. Factors* 37, 32–64 (1995)
15. Chen, J.Y.C., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.J.: Situation awareness-based agent transparency. Technical report ARL-TR_6905, Army Research Laboratory (US), Aberdeen, Maryland (2014)
16. U.S. Department of Defense—Research & Engineering Enterprise: Autonomy Research Pilot Initiative. <http://www.acq.osd.mil/chieftechologist/arpi.html>
17. Draper, M.: Realizing autonomy via intelligent adaptive hybrid control: adaptable autonomy for achieving UxV RSTA team decision superiority, Yearly report. Dayton, OH: US Air Force Research Laboratory (in press)
18. Mercado, J.E., Rupp, M., Chen, J.Y.C., Barber, D., Procci, K., Barnes, M.J.: Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors* (in press)
19. Stowers, K., Chen, J.Y.C., Kasdaglis, N., Newton, O., Rupp, M., Barnes, M.: Effects of situation awareness-based agent transparency information on human agent teaming for multi-UxV management (2011)
20. Schaefer, K.E.: *The Perception and Measurement of Human Robot Trust*. Doctoral Dissertation. University of Central Florida
21. Hancock, P.A., Billings, D.R., Schaefer, K.E.: Can you trust your robot? *Ergon. Des.* 19, 24–29 (2011)
22. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* 53, 517–527 (2011)

23. Schaefer, K.E., Chen, J.Y.C., Szalma, J.L., Hancock, P.A.: A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors* (in press)
24. Phillips, E., Schaefer, K.E., Billings, D.R., Jentsch, F., Hancock, P.A.: Human-animal teams as an analog for future human-robot teams: influencing design and fostering trust. *J. Human-Robot Interact.* (in press)
25. Schaefer, K.E., Evans, A.W., Hill, S.G.: Command and control in network-centric operations: trust and robot autonomy. In: 20th International Command and Control Research and Technology Symposium. Annapolis, MD (2015)
26. Schaefer, K.E., Brewer, R., Avery, E., Straub, E.R.: Matching theory and simulation design: incorporating the human into driverless vehicle simulations using RIVET. In: Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS). Washington, DC (in press)
27. Evans, A.W., Schaefer, K.E., Ratka, S., Briggs, K.L.: Soldier perceptions of robotic wingman platforms. In: Proceedings of the SPIE: Unmanned Systems Technology XVIII (in press)
28. Kelley, T.D.: Developing a psychologically inspired cognitive architecture for robotic control: the symbolic and subsymbolic robotic intelligence control system (SS-RICS). *Int. J. Adv. Rob. Syst.* **3**(3), 219–222 (2006)
29. Hanford, S.D., Janrathitikam, O., Long, L.N.: Control of mobile robots using the soar cognitive architecture. *J. Aerosp. Comput. Inf. Commun.* **6**(2), 69–91 (2009)
30. Mott, D., Poteet, S., Xue, P., Copestake, A.: Natural language fact extraction and domain reasoning using controlled English. DELPH-IN 2014, Portugal. <http://www.delph-in.net/2014/Mot:Pot:Xue:14.pdf> (2014)
31. Mott, D., Shemanski, D., Giammanco, C., Braines, D.: Collaborative human-machine analysis using a controlled natural language. Society for Photo-Optical Instrumentation Engineers—Defense, Security, and Sensing Symposium, MD (2015)
32. Mott, D., Giammanco, C.: The use of rationale in collaborative planning. In: Annual Conference of the International Technology Alliance (2008)
33. Mott, D., Xue, P., Giammanco, C.: A forensic reasoning agent using controlled English for problem solving. In: Annual Fall Meeting of the International Technology Alliance in Network and Information Sciences (2015)
34. Patel, J., Dorneich, M., Mott, D., Bahrami, A., Giammanco, C.: Improving coalition planning by making plans alive. *IEEE Intell. Syst.* **28**(1), 17–25 (2013)
35. Giammanco, C., Mott, D., McGowan, R.: Controlled English for Critical Thinking about the Civil-Military Domain. Annual Fall Meeting of the International Technology Alliance in Network and Information Sciences (2015)

Initial Performance Assessment of a Control Interface for Unmanned Ground Vehicle Operation Using a Simulation Platform

Leif T. Jensen, Teena M. Garrison, Daniel W. Carruth,
Cindy L. Bethel, Phillip J. Durst and Christopher T. Goodin

Abstract The successful navigation of Unmanned ground vehicles (UGV) is important as UGVs are being increasingly integrated into tactical and reconnaissance operations. Not only is there the possibility of winding environments but also the narrow passage of obstacles. This study investigated a participant's ability to navigate a maze environment incorporating narrow hallways, with two different user interfaces, using human in the loop simulation. Participants used a game controller and customized user interface to navigate a simulated UGV through a simulated maze environment. Results indicated that the video-plus-map interface displaying both video and LiDAR data required more time to complete compared to an interface displaying video-only data.

Keywords UGV · Human in the loop · Simulation · Human-robot interaction

L.T. Jensen (✉) · T.M. Garrison · D.W. Carruth
Center for Advanced Vehicular Systems, Mississippi State University,
Mississippi State, MS, USA
e-mail: leif@cavs.msstate.edu

T.M. Garrison
e-mail: teenag@cavs.msstate.edu

D.W. Carruth
e-mail: dwc2@cavs.msstate.edu

C.L. Bethel
Computer Science and Engineering, Mississippi State University,
Mississippi State, MS, USA
e-mail: cbethel@cse.msstate.edu

P.J. Durst · C.T. Goodin
U.S. Army Engineer Research and Development Center, Vicksburg, MS, USA
e-mail: Phillip.J.Durst@erd.c.dren.mil

C.T. Goodin
e-mail: Christopher.T.Goodin@erd.c.dren.mil

1 Introduction

Increasing integration of unmanned ground vehicles (UGV) into tactical and reconnaissance operations has altered the control requirements for navigating the UGV from the established domains of urban search and rescue and bomb disposal. While using UGVs, operators face the challenge of navigation and task performance, requiring a combination of cognitive, perceptual, and motor skills [1]. Operators must track the task while avoiding losses of mobility or incurring damage to the UGV platform [2]. However, in comparison to established domains, UGV inclusion in tactical operations also necessitates a more balanced approach between operation precision and speed than most previous applications, in which operator and bystander safety are paramount. In response to increased time pressure and broader control demands, an increasing number of sensors and information sources have been provided to operators. While additional sensors may support navigation and task performance, they may also place greater situation awareness (SA) demands on the operator.

With these challenges facing the operator, the ability of the user interface to allow the operator to maintain acceptable SA is increasingly important. The human-robot interaction (HRI) definition of SA had been adapted from the popular definition of SA by [3]. This definition focuses on the awareness the operator has of the robot's location, status, activities, and surroundings as well as the robot's interpretation of operator input in the current context [2, 4].

These problems have been the focus of several research projects using both real-world environments and simulation environments [5–8]. UGV navigation is difficult, thus producing interface features that allow the participant to successfully navigate and explore any given environment is essential. Video-centric interfaces are the most common type of interface and are used in most commercially available robots [5]. However, the placement of the camera on UGV platforms often does not allow the operator to see the sides of the UGV, including the tracks or wheels which extend several inches beyond the camera's field of view. This has been shown to cause a majority of collisions as the operator interprets what they can see in the video feed as being the bounds of the UGV [5]. In order to help in navigation, map-only interfaces or video-plus-map interfaces have been used to improve navigation and reduce collisions [6]. While the addition of maps and other sensors can prove helpful, they also have the potential to overload the user [8].

Test environments previously used, based off the National Institute of Standards and Technology (NIST) USAR test arena, do not indicate the incorporation of narrow hallways in their test environment [5]. These narrow passages help simulate obstacles, such as doorways, that a UGV may encounter. Passing through these obstacles present additional challenges to navigation that may not be sufficiently addressed with existing interfaces. This study was designed to test the effectiveness of two web-based video-centric user interfaces, using video-only and video-plus-map, ability to allow participants to pass through narrow hallways.

2 Methods

2.1 UGV and Test Environment

The experiment was conducted in a simulated environment, using a simulated UGV. This study used the Autonomous Navigation Virtual Environment Laboratory (ANVEL) simulation software package in conjunction with the ANVEL-ROS (Robot Operating System) plug-in [9]. Participants used a game controller and customized web-based user interface (see Figs. 2 and 3) to navigate a simulated UGV through a maze like environment. All data in the user interface was provided through simulation using the ANVEL platform. The UGV was modeled after the Jaguar V4 platform [10]. The Jaguar V4 is a tracked UGV with four articulated flippers that were set at approximately a 45° angle during participant's drives. Measurements of the UGV in this configuration are 30" long, 27.6" wide, and 7" high.

Participants were tasked with driving through a maze-like course with 13 90° turns (see Fig. 1). There was a single entrance and single exit in the course and participants were told to drive from the entrance until they exited the environment. The environment contained four narrow hallway sections of various size (Fig. 1a–d), which the participants were required to navigate through. The normal sections of the

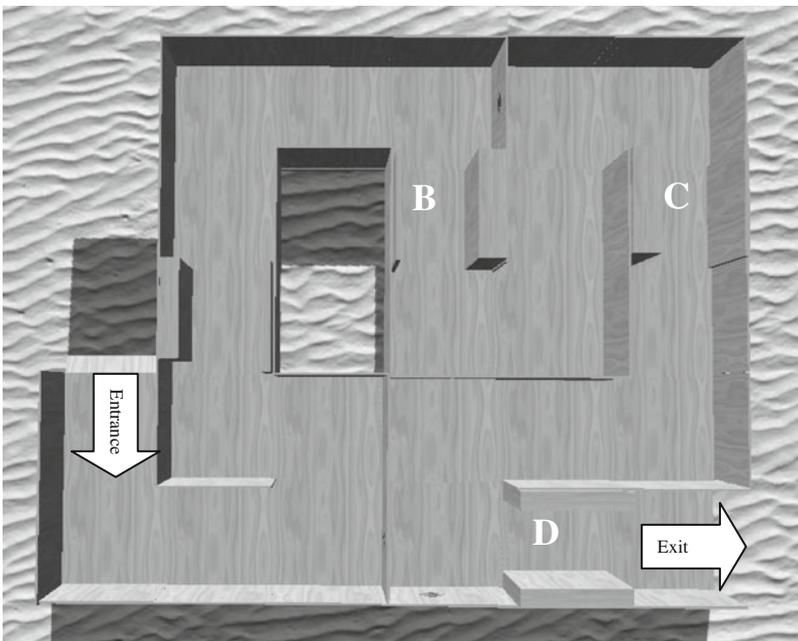


Fig. 1 Maze-like environment that was used for participants to drive the UGV through. Narrow sections: **a** (36"), **b** (34"), **c** (34"), and **d** (32")

course measured 48" wide. Narrow sections ranged from 32" to 36" wide providing the UGV between 4" and 8" of total clearance when passing through a hallway. This provided tight tolerances for the UGV's passage in these sections of the environment.

Participants' controlled the UGV using a Logitech gamepad controller. This controller has a directional pad on the left side, two analog sticks, two left and two right shoulder buttons, four buttons on the right, and in the middle of the controller a start, mode, back and vibration button. For this study input was limited to the right analog stick which was used to provide directional input. When the participant moved the right analog stick, forward and backward provided longitudinal movement of the UGV while right and left rotated the UGV.

2.2 User Interface

Two versions of a web-based user interface were used in this study. The user interfaces were custom implementations based on previously published interface designs [5]. The video-only interface (Fig. 2) consisted of two video feeds. A large forward-facing video feed was placed in the center of the interface. A smaller rear-facing video feed was placed in the upper right. A direction indicator pad

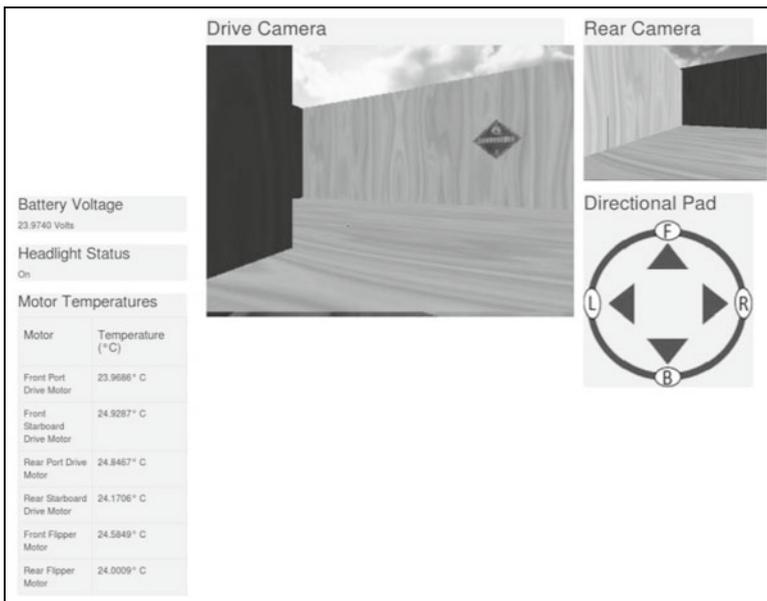


Fig. 2 Basic video-only user interface, showing main drive forward-facing video feed, and rearview video feed

changed arrow colors to indicate directional inputs received from the controller. Robot status indicators, such as battery status and motor temperatures, were placed on the left side of the main video display.

The video-plus-map user interface contains all of the same features as the video-only interface while including the addition of two LiDAR maps. The first LiDAR map gives a top down view of the robot, in a window of similar size to the rear facing camera, and is located in the upper left. The second LiDAR map is located below the main video screen and is a chase map positioned just behind the robot. Both of the LiDAR maps are updated with the sensor information as the robot proceeds through the environment.

2.3 *Study Procedures*

Twenty-four participants were recruited for the study. Of the twenty-four participants twelve were male and eleven female, one did not provide demographic information. Participants were given a brief introduction to the UGV, workstation, simulation environment, and study requirements. Participants were asked to fill out a motion sickness/simulator sickness questionnaire (MS/SSQ) to establish a baseline. After completing the appropriate forms, the participant was given the controller and instructed on the controls as well as familiarized with the interface layout. Each participant drove a familiarization and a trial run for each of the two user interfaces resulting in a total of 4 runs. Presentation order of the interfaces was counter-balanced. Participants filled out the MS/SSQ and additional surveys not reported here after completing the runs for each interface. After completing all 4 runs through the environment the participant completed a vehicle comparison, and user interface/user experience survey.

Of the twenty-four participants that were recruited for this study, three participants did not complete all four drives and were removed from the analysis. Of these three participants, one participant did not complete their drives due to technical difficulties, one was unable to complete the four drives in the allotted time (1 h), and one voluntarily withdrew. Two additional participants with data greater than two standard deviations from the mean on multiple dependent variables were also removed from the analysis. Results are presented for the remaining nineteen participants. Two primary measures were recorded from each participant's trial runs through the environment: number of collisions and completion time.

Collisions were counted via observation through video play back by a single individual. A collision was considered to have occurred any time a participant hit a wall or obstacle with any part of the UGV. If a participant hit a wall, which stopped forward progress on the UGV, then stopped giving controller input indicated by the UGV tracks no longer trying to move, then gave a command to move the UGV in the blocked direction, this was counted as a second collision. Time started at the participant's first forward movement, time stopped when the participant exited the course.

3 Results

Overall average time to complete a run through the maze was 210.47 s (SD = 131.8 s). Results indicated that the video-plus-map interface required more time to complete the task ($M = 255.2$ s, $SD = 157.77$ s) compared to the video-only interface ($M = 165.73$ s, $SD = 74.55$ s, $F[1, 18] = 6.299$, $p < 0.05$, Cohen's $f = 0.591$) (Figs. 3 and 4).

Not only was the completion time greater using the video-plus-map interface but the number of collisions was also greater ($M = 57.89$, $SD = 42.72$) compared to the video-only interface ($M = 30.84$, $SD = 12.53$, $F[1, 18] = 8.884$, $p < 0.05$, Cohen's $f = 0.702$) (Fig. 5).

The results of the user interface/user experience survey, consisting of questions designed for this study as well as the system usability questionnaire and semantic differential scale [5, 11], indicated that participants reported no significant differences in perception of the interfaces despite the poorer performance given with the video-plus-map interface (Table 1).

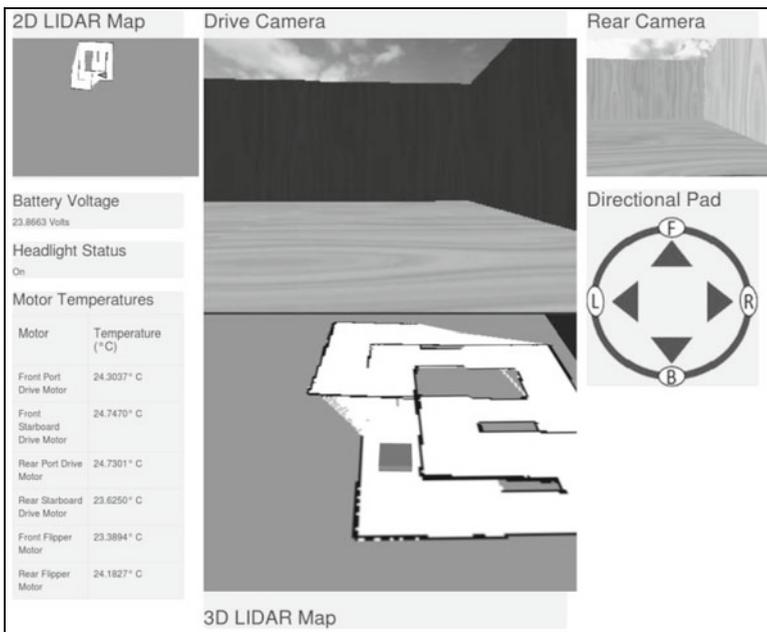


Fig. 3 Video-plus-map interface showing LiDAR map displays in addition to the main drive forward-facing video feed and rearview video feed

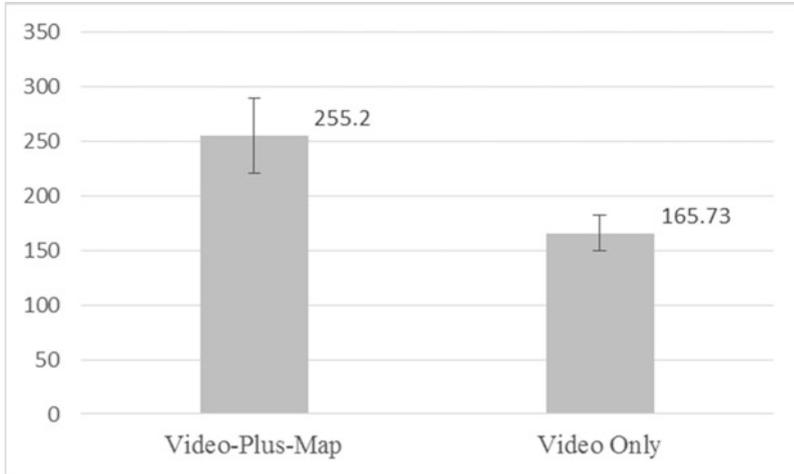


Fig. 4 Time to complete course for the video-plus-map interface is showing to be significantly larger than the video-only interface. Error bars show standard error

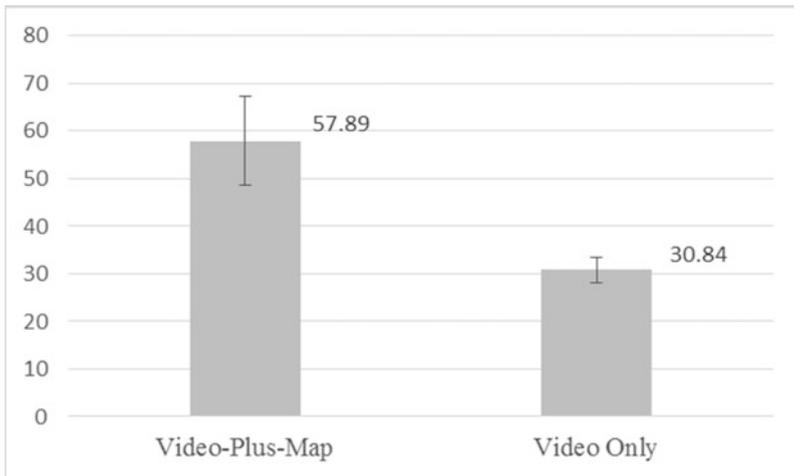


Fig. 5 Number of collisions for video-plus-map interface is significantly greater than in the video-only interface. Error bars show standard error

Table 1 User interface/user experience survey results showing no significant difference in UI perception

User experience survey	Mean	Standard deviation
Video-plus-map	2.3846	0.9892
Video-only	2.2802	0.7778

4 Conclusions

The results were opposite from what we expected to find, based on previous findings from published studies, as the video-plus-map interface with LiDAR map data, resulted in poorer performance than the video-only interface [5, 6]. There are several factors that may have contributed to the lack of performance in the video-plus-map interface when compared with the video-only interface. First the LiDAR chase map, in the video-plus-map interface, sometimes provided slightly inaccurate data, indicating that the UGV was closer or farther away from a wall or obstacle than it actually was. This potentially resulted in an increase in collisions if participants relied on the map interface. An interesting observation occurred when this happened; most of the participants realized that the LiDAR map was not completely accurate, verbally commenting to that effect either during or after their run. However, successive collisions in the same location indicated participants still continued to rely heavily on the LiDAR map even after verbally commenting on its inaccuracy. The continued focus on the map could be due to participant's dislike of not being able to see the sides of the UGV, which was commented on both verbally and written on the user experience survey. Another possible contributor to the reduced performance with the video-plus-map interface is the use of narrow hallways in the environment. The narrows sections appear to cause a large number of collisions. Previous studies did not include varying narrow sections for participants to traverse [5, 6]. Because of the tight tolerances required to pass through these sections even small inaccuracies in the LiDAR map could hinder the participant's ability to pass without collision.

Future studies will explore these issues, comparing the results of using the same environment minus the narrow hallways. Also exploring further enhancements to the user interface such as various safe path indicators to reduce the number of collisions as well as decreasing participant's completion times.

Acknowledgments Material presented in this paper is a product of the CREATE-GV Element of the Computational Research and Engineering Acquisition Tools and Environments (CREATE) Program sponsored by the U.S. Department of Defense HPC Modernization Program Office. This effort was sponsored under contract number W912HZ-13-C-0037.

References

1. Lathan, C.E., Tracey, M.: The effects of operator spatial perception and sensory feedback on human-robot teleoperation performance. In: *Presence: Teleoperators & Virtual Environments*. Vol. 11(4), pp. 368–377, 10p (2002)
2. Drury, J.L., Scholtz, J., Yanco, H.: Awareness in human-robot interactions. In: *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 1, pp. 912–918 (2003)
3. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 32(2), pp. 97–101 (1988)

4. Yanco, H.A., Drury, J.: "Where am I?" Acquiring situation awareness using a remote robot platform. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 3, pp. 2835–2840 (2004)
5. Keyes, B., Micire, M., Drury, J.L., Yanco, H.A.: Improving human-robot interaction through interface evolution (2010)
6. Nielsen, C.W., Goodrich, M.A.: Comparing the usefulness of video and map information in navigation tasks. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction pp. 95–101 (2006)
7. Gomez, A.V., Chiara, S., Fabiano, B., Gabriele, R.: Spatial processes in mobile robot teleoperation. *Cogn. Process.* **10**, 338–341 (2009)
8. Kadous, M.W., Sheh, R.K. M., Sammut, C.: Effective user interface design for rescue robotics. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction pp. 250–257 (2006)
9. Hudson, C., Carruth, D.W., Bethel, C.L., Lalejini, A., Odom, B.: ANVEL-ROS: The integration of the robot operating system with a high-fidelity simulator. In: Proceedings of Ground Vehicle Systems Engineering and Technology Symposium (2015)
10. Jaguar V4 Platform Specification. Dr. Robot, Inc., http://jaguar.drrobot.com/specification_v4.asp
11. Fruhling, A., Sang, L.: Assessing the reliability, validity and adaptability of PSSUQ. In: Proceedings of AMCIS 2005 pp. 378 (2005)

Part II
Confronting Human Factors Challenges

Examining Human Factors Challenges of Sustainable Small Unmanned Aircraft System (sUAS) Operations

Clint R. Balog, Brent A. Terwilliger, Dennis A. Vincenzi
and David C. Ison

Abstract Small unmanned aircraft systems (sUAS) represent a significant instrument for improving task efficiency and effectiveness across numerous industries and operational environments. However, concern has grown regarding potentially irresponsible operation and public apprehension to potential privacy loss. These concerns, combined with unique sUAS human factors challenges, may lead to unwanted and dangerous results, including reduction of safety, property damage, and loss of life. Such challenges include lack of command, control, and communication (C3) standardization; detection, tracking, and managing operations; and human perceptual and cognitive issues. Issues and concerns could be significant barriers to permitting routine and sustainable operations in the National Airspace System (NAS), but by closely examining these factors may be possible to devise strategies to better support future application. This exploratory study seeks to provide a review of relevant exigent literature as well as condense findings into sets of recommendations and guidelines for human factors in sUAS adoption and use.

Keywords Small unmanned aircraft systems (sUAS) · Command, control, and communication (C3) · Human-machine interface (HMI) · UAS human factors

C.R. Balog (✉) · B.A. Terwilliger · D.A. Vincenzi · D.C. Ison
Embry-Riddle Aeronautical University, Worldwide Campus, Daytona Beach, FL, USA
e-mail: clint.balog@erau.edu

B.A. Terwilliger
e-mail: brent.terwilliger@erau.edu

D.A. Vincenzi
e-mail: dennis.vincenzi@erau.edu

D.C. Ison
e-mail: david.ison@erau.edu

1 Introduction

Small unmanned aircraft systems (sUAS) represent a significant instrument for improving efficiency and effectiveness of task performance in a variety of industries and operational environments. However, concern regarding potentially irresponsible operation, including disregard for federal and local regulations, laws, and guidance governing their safe use, is growing [1–3]. In addition, there has been significant public apprehension to potential personal privacy loss during sUAS operations [4–6]. These concerns, combined with unique sUAS human factors challenges, present situations that may lead to unwanted and dangerous results, including reduction of safety, property damage, and loss of life. Such challenges include lack of command, control, and communication (C3) standardization [7]; detecting, tracking, and managing operations in permissible airspace (B-E, and G) [8]; and human perceptual and cognitive issues [9]. These challenges could be significant barriers to permitting routine and sustainable unmanned aircraft operations in the National Airspace System (NAS), beyond requirements already specified in existing and proposed regulations for operation and certification of sUAS. By closely examining these issues, it may be possible to devise strategies to better support future beyond visual line of sight (BVLOS), cargo or external load transport, and emergency response applications and commercial operations. It will be through safe performance of these potential uses where true economic and societal gains will be achieved. Therefore, the human factors affecting their implementation and realization warrant further investigation and analysis. This exploratory study seeks to provide a review of relevant exigent literature as well as condense findings into sets of recommendations and guidelines for human factors in sUAS adoption and use.

2 sUAS-Related Issues and Concerns

sUAS operations are increasing worldwide at an explosive pace. The sUAS operational community is rapidly expanding, both in absolute numbers and in the spectrum of operators, resulting from ever-expanding sUAS accessibility, which itself is both the result of, and feeds, significantly decreased costs to both enter into and participate in sUAS operations [10, 11]. This presents many issues and concerns. Many of these derive from a self-feeding cycle of ever-expanding sUAS technological capability, which results in the continued innovation and expansion of sUAS applications, which in turn results in continually increasing conflicts surrounding such operations. This, then, leads to expanding regulatory challenges surrounding the use of sUAS and their associated technologies. All of this has caught most involved parties insufficiently prepared.

2.1 Growing Operational Community and sUAS Accessibility

Until now the largest component of the sUAS market has been military/civil market. But new technologies and regulations are resulting in a rapidly expanding commercial sUAS industry [11]. sUAS can range from hobbyist models that fly short distances for brief time periods on one charge and cost as low as \$200 US to commercial level configurations that fly for extended periods of time and cost upwards of \$10 K US [10]. As prolific as sUAS manufacturing has become in the U.S., it is China that is currently the leading manufacturer of sUAS [12]. Estimates for future market growth universally show this trend accelerating. Various research estimates the trend in sUAS sales to exceed between \$6B and \$8.4B worldwide by the 2018/2019 timeframe, and approaching \$14B US in 10 years. By 2019 it is estimated that the commercial sUAS sector will command the majority of the market, holding roughly five times that of the prosumer/hobby market and two to three times that of the military/civil market [10], [11]. These estimates have fueled significant investment in sUAS manufacturers, further fueling this growth cycle. The only apparent throttles to this remarkable growth are cost-effective technological progress and uncertainty regarding regulatory framework.

In an initial response to operational concerns stemming from this market growth, particularly in the prosumer/hobby sector, numerous public outreach campaigns have been developed to educate operators new to the market in safe sUAS operations. Perhaps among the most notable of these is Know Before You Fly, a joint consumer education campaign founded by the Association for Unmanned Vehicle Systems International (AUVSI), the Academy of Model Aeronautics (AMA), and the U.S. FAA [13]. The FAA has also established its own, independent educational outreach program called No Drone Zone [14]. Even commercial retailers who sell sUAS have begun to support educational outreach, including Best Buy with the Fly Responsibly section of its Drone Buying Guide [15].

2.2 Increasing Technological Capability

As a growing consumer technology, sUAS advancements are being pushed by both progression in complimentary technologies, as well as customer-driven development efforts. Improving performance characteristics, such as speed, range, and endurance, combined with enhanced functionality of C3, remote sensing and processing, as well as affordability and availability, are increasing the potential applications supported using these systems. Such expanding capabilities enable sUAS to fly further, faster, carrying greater loads, while also being simpler to operate due to intuitive human-machine-interfaces (HMI) featuring integrated autopilot and flight stability control, system state, and positioning [16, 17]. However, with increased operational ease the potential increases for users to

become overconfident in flight [18], without understanding or having gained experience with a degraded operational state (e.g., loss of stability assist from GPS signal loss) or upset inducing conditions. Such a combination could result in damage, injury, or death, if improper recovery is performed or attempted. In addition, lack of a standardized format for system control HMI necessitates unique operational training or practice for each specific platform or family of platforms manufactured by a vendor. Furthermore, with the wide availability of platforms, lack of current sUAS specific certification or training requirements, and proliferation of similar consumer options, the potential is high for improper, possibly illegal, operations, as well as user confusion or inappropriate control response in operation. As consumer sUAS use continues to grow, it will be imperative to ensure a base level of knowledge, skills, and abilities (KSAs) and guidance resources relating to their operation exists among the user community, especially for those with little training, knowledge, or experience in aviation settings [19].

2.3 Regulatory Change

Per the FAA Modernization and Reform Act of 2012, the agency was required by Congress to establish regulations for the commercial use of UAS by September 30, 2015. Initial response by the FAA was laggard and after several years of limited action and strict regulation the reigns on UAS operation have begun to loosen. Currently (as of February 2016), the FAA does not require specific approval for recreational or hobby use of UAS as long as it weighs under 55 lb (unless certified by a modeling community-based organization) and requires all such vehicles to be registered if they weigh more than 0.55 lbs. For those wanting to operate UAS for non-recreational purposes, specific approval is required. There are four ways to achieve this approval from the FAA, Special Airworthiness Certificates—Experimental Category (SAC-EC), Restricted Category UAS type and airworthiness certificates, to petition for an exemption with a Certificate of Waiver Authorization (COA), or to petition for a Section 333 exemption. Section 333 and COAs are also awarded with a nationwide blanket COA for operations below 200 feet and additional restrictions on operations around airports, restricted airspace, and over populated areas. The FAA states that the Section 333 exemption process takes approximately 120 days while a COA may take from 60 to 90 days [20].

As of December 21, 2015, all sUAS from 0.55 lbs to less than 55 lbs must be registered and the owner must be 13 years or older and a US citizen or legal permanent residents (those under 13 must have someone at least 13 register the sUAS). Additional registration requirements exist for those using their aircraft for commercial purposes, other than hobby and recreation, if UAS weigh more than 55 lbs, or if flying outside U.S. The registration is valid for three years and the registration number provided must be marked on the sUAS. This requirement also includes Section 333 aircraft. Additionally, Section 333 operators must hold an

FAA airman certificate. For-hire operations currently require a commercial pilot certificate [21].

In early 2015, the FAA unveiled their plan for the creation of Part 107 of Federal Regulations which would cover sUAS. This proposed regulation provides more flexibility as to how, where, and by who sUAS can be operated. This proposed regulation outlines operator requirements, platform certification, registration requirements, and operational limitations.

Also released in 2015 was a Presidential Directive that mandates Federal agencies to be more transparent about data collection from aerial surveillance to ameliorate public concern about invasion of privacy. It is expected that similar guidance to be provided in the new sUAS rules once released. Additional concerns to sUAS operators are state and local legislation. According to the Association of Unmanned Vehicle Systems International (AUVSI), 21 states have active or passed bills restricting UAS operations. Some of these actions have focused on privacy as well [22]. Even more confounding, some municipalities have added their own limitations and in some cases outright prohibiting UAS operations within city limits [23]. Thankfully for UAS operators, it is surmised that such restrictions will not be valid once Federal standards are in place, although it will likely take time for the local entities to realize this, most likely within a courtroom [24].

3 Human Factors Challenges

While sUAS operations share many human factors challenges with manned air vehicle operations and with larger UAS operations, they also engender some unique challenges founded in human cognitive performance and limitations. These result from sUAS size, relative technological complexity, and from the experience level of the population of operators. The latter is a result of the relatively low cost and availability of sUAS, which opens entry into the operational environment available to an increasingly broad spectrum of the population. This results in unique sUAS challenges associated with C3 functions, detection, tracking, and management of operations, and with HMI.

3.1 Command, Control, and Communication

The human element has always been a priority concern in all aspects of manned aviation, and unmanned systems are no different. The term unmanned system is a misnomer in itself, since there is a human operator inherent to the system, but not collocated in the vehicle. As the FAA and other entities begin to acknowledge the impending proliferation of UAS in the NAS, identifying the proper balance of human interaction and autonomous control in the C3 of UAS has become an issue of extreme importance [25]. Among sUAS, lack of standardization of control

stations and interfaces creates confusion, increases cost for training, and increases safety and control issues. For local visual line of sight devices (VLOS), control devices can vary from small, smartphone and tablet devices to handheld dedicated control units. For BVLOS, control devices become larger, more capable, and more complex due to the need to control and communicate with a vehicle which can no longer be seen.

Ground control stations (GCS) of UAS range from commercial off-the-shelf (COTS) controllers, to sophisticated specific-purpose interfaces housed in trailers or ground facilities [26]. A considerable challenge has been to design interfaces that provide salient information capable of maintaining situational awareness in UAS that are capable of BVLOS operation. The absence of rich perceptual cues available to pilots of conventional manned aircraft has resulted in debate concerning proper design of control stations, level of automation needed, transparency of automation time de-lays in control and communication, and challenges of maintaining pilot vigilance and engagement during extended periods of low workload resulting from high degrees of automation and autonomy embedded within command and control systems [26, 27].

3.2 Detection, Tracking, and Management of sUAS Operations

Air Traffic Control (ATC) personnel have described the task of integration of UAS into the NAS as “the hardest things we have ever done” [28, p. 1]. Tracking sUAS can be challenging. See and avoid principles have been employed in manned aviation for many years in ATC, but with the advent and proliferation of UAS in general and specifically sUAS, visual detection and tracking can be difficult, if not impossible. New technologies, including sense and avoidance, are available and must be implemented in all UAS from the beginning, if smooth and safe integration with manned aircraft in the NAS is the objective. Response to ATC instructions by UAS also present a challenge as communications are often delayed because messages must be relayed via operators on the ground before the necessary control inputs are made [28].

The goal of safely integrating UAS without segregating, delaying, or diverting other aircraft and other users of the system presents significant challenges for ATC [29]. Existing standards ensure safe operation by pilots actually on board the aircraft. Removing the pilot from the aircraft creates a series of performance considerations between manned and unmanned aircraft that need to be fully researched and understood to determine acceptability and potential impact on safe operations in the NAS [29].

Two technological areas present specific human factors challenges to UAS pilots and ATC: Sense and Avoid capability and C3 system performance requirements. Sense and avoid capability must provide for self-separation and collision avoidance

protection between UAS and other aircraft analogous to the current “see and avoid” operation of manned aircraft [29]. This coupled with potential limitations of C3 including datalink reliability and lag time in communications presents potential safety issues that need to be overcome before true integration into the NAS with manned aircraft can be implemented.

3.3 sUAS-Specific Human-Machine Interfaces

sUAS HMIs should be designed to take best advantage of human performance capabilities (primarily cognitive and biomechanical) while overcoming human performance limitations in order to promote the efficiency, effectiveness, and safety of flight operations. Cognitively this means HMI design and operation specifics should support human perceptual (primarily visual), memory, risk assessment, problem solving, and decision making processes. Control HMIs will vary between systems employing internal and external operators and dependent upon the level of automation. For those systems providing any level of manual control or for the ability of the pilot to override the automation a standardized flight control system should be developed. Whether that be a stick and rudder system, a cyclic and collective system, joysticks, or some other system, the key is to standardize that control system across sUAS platforms as much as possible [30, 31]. Without standardization the effectiveness, efficiency, and safety of operations are all reduced due to the additional cognitive processing demands a variety of non-standardized systems would place on operators [31]. To this end, certain control types should also be standardized for specific non-flight control functions?

Similarly, sUAS operations would benefit from standardization of operator displays, but doing so requires answering numerous questions. All but line-of-sight operations will rely on imagery developed from onboard sensor data for manual vehicle control. The quality of the display may be degraded due to datalink bandwidth limits and spatial resolution. Such degradations include poor spatial and temporal resolution, poor field of view (FOV), and low visual and data update rates [32]. These, in turn, will impact both vehicle control and detection of other air traffic, as well as safety of operations, through their impact on operator cognitive performance [31]. What, then, should be the standards developed for these operational parameters? Can tools such as augmented reality displays and synthetic visual systems be used to support cognitive processing and thus compensate for these degradations [30]? Unlike pilots of manned aircraft, sUAS operators are deprived of a range of sensory cues, such as ambient visual input, kinesthetic/vestibular information, and sound, important to cognitive processing related to the flight operations. It is therefore important for sUAS to provide displays and alarms that will compensate for this missing information and keep operators well informed of system status [31]. Perhaps multimodal display systems can successfully accomplish this. If so, what data should be available to the pilot and when? How, and to what extent, should this data and display presentation be

standardized across sUAS platforms? Should sUAS displays be similar to manned aircraft? How should payload and other non-flight operations data be displayed and integrated? Should certain information always be available to the operator while other data is menu-driven? These are among the questions that must be answered in order for sUAS operations to reap the benefits of operational efficiency, effectiveness, and safety derived from HMI standardization.

4 Technological and Procedural Advances

The increased level of civilian interest in sUAS use is contributing to knowledge discovery and improvement, challenging members of the stakeholder community and peripheral fields to extend capabilities, while increasing safety and effectiveness of the underlying technology. While many of the innovative advancement are directly tied to technological development, procedural changes and regulatory requirements are also significantly impacting how sUAS are employed. These efforts, including individual and collaborative pursuits among government, industry, and academia, provide the potential for substantial economic, educational, risk mitigation, and altruistic gains.

Variability among sUAS platform options has changed substantially from the introduction of vertical take off and landing (VTOL) designs, such a multirotors. While performance limited, compared to fixed-wing options (i.e., reduced speed, range, and endurance), these platforms afford expanded maneuverability and rapid reorientation. Such capabilities enhance the capture of imagery from an aerial perspective, especially for localized VLOS infrastructure inspection, aerial filming and photography, and exploratory investigation and intelligence gathering [33]. A substantial number of consumer multirotor platforms are available as COTS, produced by vendors able to leverage economies of scale and process efficiency in manufacturing [34]. Construction of sUAS and their constituent components has also been affected by the availability of novel new materials and production techniques, such as additive layer manufacturing (e.g., 3D printing). The development of new composites has led to the integration of lighter and stronger structures, beneficial thermal properties (e.g., de-icing), micro-actuation, electrical conductivity, and the possibility for enhanced detection (e.g., increased radar signature) and self-repair [35, 36]. The availability of consumer 3D printing equipment, with design media shared across among user communities and low-cost consumable material purchasable from local and online retailers, has also continued to grow. The affordability and increased access to such technology has resulted in substantially varied design options, conceptualized and immediately fabrication (i.e., rapid prototyping), by a diverse population ranging from casual hobbyists to experienced professionals [37].

A critical challenge of integrating sUAS technology into the NAS is ensuring “see and avoid” capability, which is mandated by current Federal regulations, can be maintained among all aircraft operators (i.e., pilots). By the inherent nature of the design sUAS pilots are unable to see, unless they are within VLOS of the aircraft [38]. Using sensing technology, a pilot can sense or detect, and subsequently avoid (DSA), potential obstacles, aircraft, and objects, given a sufficient sensor fidelity. This use of technology, to replicate and improve upon human perception limitations, provides a potential solution to improve avoidance response and overall operational safety within the NAS. Sensing technology is also being improved through development of lower weight and sized components that increase accuracy and fidelity (e.g., greater resolution, depth, and field of view) with decreased power consumption, further extending the operational endurance or usability of platforms. Increased fidelity and accuracy of captured data, using such sensors, translates to the creation of more precise models, databases, and information. Improved manufacturing techniques, coupled with economic competition, has decreased cost, while increasing accessibility and use. Current areas of sensing advancement include active onboard distance sensing (sonar), ground-based sensing (e.g., radar), data processing and categorization algorithms, optical sensing (machine-vision), and alternative localization to support operations in GPS denied environments [39].

Improvements to C3 infrastructure, including the HMI (command and control) and communication capabilities, can improve overall efficiency and effectiveness of the system. Intuitive graphical user interfaces (GUIs) and controls are simplifying operator interaction and programmability of autonomous functions, while common (universal) architecture in both defense, as well as consumer options, enable control of multiple types of UAS with expandable features mapped to platform type. Integration with touch-sensitive smart devices, such as phones as tablets, are replacing or augmenting discrete physical controls and information display. Novel interfaces, such as electroencephalography, support mapping control commands to thought patterns [40]. Multi-nodal (peer-to-peer) and cellular networking are providing new pathways for data distribution and exchange, with users potentially able to share high-fidelity or resolution imagery and telemetry among platforms, pilots, and ground-based assets. These networks also provide primary or secondary (redundant) interoperability for command and control, decreasing potential for lost-link. Other areas of communications R&D include enhanced security (encryption) for protection of data exchange and investigation into exotic topics, such as quantum and plasmonic technologies [41, 42]. The potential gains expected from such research include improved speed and encryption (reduced latency) and reduced power consumption.

Power and propulsion sources affect overall speed, endurance, and range of a platform, which can limit or support the end usability of the overall system. Specifically, the size of operational coverage area, payload capacity, speed of response, and time aloft (i.e., mission time) [34]. A number of R&D efforts are being conducted to extend and improve the power and propulsion of unmanned systems, with some advances coming from manned transportation (e.g., electric automobile) and portable computing industries. At the forefront is the development and use of

alternative or enhanced power storage and conversion (motor) options. Researchers have developed new or modified fuel cell formats (e.g., solid, lightweight non-compressed hydrogen and solid oxide fuel cell [SOFC]) to improve energy density over conventional options [43, 44, 45]. Increased storage potential and life of batteries are also being studied, with promising discoveries associated with the incorporation of silicon, graphene, and solid electrolyte to improve stability, life, and energy density [46, 47, 48]. Research to apply supplemental power delivery for operational recharge using improved solar cells and laser energy transmission also indicate potential to reduce weight and increase efficiency [49].

R&D advances are not isolated to technological subsystem improvements; a number of projects are being conducted to better understand how to safely incorporate, manage, and apply sUAS within the NAS. Topics include airspace integration, BVLOS operations, airworthiness testing, training and certification, risk reduction, and maintenance. The Federal government, including the FAA and the National Space and Aeronautics Administration (NASA), has sponsored initiatives, such as the Pathfinder program, UAS Traffic Management (UTM), and individual projects conducted through the Test Sites and Center of Excellence to gather data and develop results that will be used to guide future regulatory requirements [50–53]. It is expected that future research ventures, which are currently conducted under COAs (public entities) and Section 333 Grants of Waiver, will continue to increase with improved access to airspace under proposed sUAS operational rules [54].

5 Conclusions and Recommendations

This exploratory study examined the human factors challenges of sustainable sUAS operations. Specifically, this research explored the ways that task efficiency and effectiveness can be augmented among a variety of operational environments and circumstances. Many of the issues identified in the implementation of sUAS are similar to issues present in manned flight operations, including human factors issues, level of automation and autonomy, and interface design. The primary challenges to sUAS are lack of standardization among C3 and HMI, in addition to limitations on detecting, tracking, and managing operations. These are known challenges that have been present and acknowledged for at least the last decade, and yet little has been accomplished to resolve these issues. With the rapid expansion and availability of sUAS, it has been identified that limitations must be mitigated to safely and effectively integrate these systems into the airspace operational environment. Clearly, sense and avoid as well as C3 issues must be resolved in order to achieve this goal.

Moreover, standardizing sUAS controls and operator displays will augment safety and inter-utilization in all types of operations particularly in light of the wide range of users either currently or potentially operating these devices. A primary concern of the FAA is and has always been the safe implementation of UAS into the NAS. This focus has not changed and will only proliferate in the future as UAS

become more viable as commercial alternatives for current applications. Further research and development activities in power, propulsion, construction, and application will also bolster the industry. With all of these efforts and advances coupled together, sUAS have the ability to provide extremely beneficial utilities to a range of industries and entities, as well as to overall mankind.

References

1. Jansen, B.: FAA: 181,000 drones registered since database began Dec. 21. USA Today (2016), <http://www.usatoday.com/story/news/2016/01/06/faa-181000-drones-registered-since-dec-21/78360534/>
2. Lowry, J.: Aviation task force recommends registration of even smaller drones. PBS News hour (2015), <http://www.pbs.org/newshour/rundown/aviation-task-force-recommends-registration-of-even-smaller-drones/>
3. Weiner, M.: Rogue drones: FAA plans crackdown after spike in risky flights. The post standard (2015), http://www.syracuse.com/politics/index.ssf/2015/08/drones_gone_wild_faa_plans_crackdown_afer_spike_in_risky_flights.html
4. Miller, S., Witt, N.: The wild west of commercial drones—why 2015 could be a pivotal ear in California. Public Law J. **38**(2), 1–8 (2015)
5. Sterbenz, C.: Should We Freak Out About Drones Looking In Our Windows?. Business Insider (2014), <http://www.businessinsider.com/privacy-issues-with-commercial-drones-2014-9>
6. Miller, P.C.: Aviation Attorney: UAS Privacy Issues Present Legal Challenges. UAS Magazine (2015), <http://www.uasmagazine.com/articles/1141/aviation-attorney-uas-privacy-issues-present-legal-challenges>
7. Kerczewski, B.: Spectrum for UAS Control and Non-Payload Communications. 2013 ICNS Conference (2013)
8. Pomerleau, M.: Who'll Keep Track of All Those Drones in Flight?. GCN (2015), <https://gcn.com/articles/2015/08/31/drones-utm.aspx>
9. National Research Council: Autonomy Research for Civil Aviation: Toward a New Era of Flight. The National Academies Press, Washington (2014)
10. Canis, B.: Unmanned Aircraft Systems (UAS): Commercial Outlook for a New Industry. Congressional Research Service, Washington (2015)
11. Small Unmanned Aerial Systems Market Exceeds US\$8.4 Billion by 2019, Dominated by the Commercial Sector and Driven by Commercial Applications. ABI Research. (2015). <https://www.abiresearch.com/press/small-unmanned-arial-systems-market-exceeds-us84-b/>
12. Global Industry Insight: Small Unmanned Aerial Systems Market Size, Share, Development, Growth and Demand Forecast to 2020. P&S Market Research. (2015) <https://www.psmarketresearch.com/market-analysis/small-unmanned-aerial-systems-market>
13. Know Before You Fly/ About, <http://knowbeforeyoufly.org/about-us/>
14. No Drone Zone, https://www.faa.gov/uas/no_drone_zone/
15. Drone Buying Guide/Fly Responsibly, <http://www.bestbuy.com/site/buying-guides/drones-buying-guide/pcmcat381100050003.c?id=pcmcat381100050003>
16. Floreano, D., Wood, R.J.: Science, technology and the future of small autonomous drones. Nature **531**, 460–466 (2015). doi:10.1038/nature14542
17. Cai, G., Dias, J., Seneviranate, L.: A survey of small-scale unmanned aerial vehicles: recent advances and future development trends. Unmanned Syst. **2**(2), 1–25 (2014). doi:10.1142/S2301385014300017

18. Roos, C.: Human error in operating mini RPAS—causes, effects and solutions. Paper presented at the RPAS Civops 2013 conference in Brussels, BE (2013). http://uvs-international.org/docs/rpas-civops-13/White-Papers/27_NLR_Netherlands_Christopher-Roos_WP.pdf
19. Terwilliger, B.A., Ison, D.C., Vincenzi, D.A., Liu, D.: Advancement and application of unmanned aerial system human-machine-interface (HMI) technology. In: Yamamoto, S. (ed.), *Human Interface and the Management of Information: Information and Knowledge in Applications and Services*, vol. 8522, pp. 273–283 (2014)
20. Unmanned Aircraft Systems (UAS) Frequently Asked Questions, <https://www.faa.gov/uas/faq/#qn19>
21. Section 333 Frequently Asked Questions, https://www.faa.gov/uas/legislative_programs/section_333/333_faqs/
22. State UAS Legislation Interactive Map, <http://www.auvsilink.org/it/TestMap/test.html>
23. Breitenbach, S. States Rush to Regulate Drones Ahead of Federal Guidelines. Pew Charitable Trusts (2015), <http://www.pewtrusts.org/en/research-and-analysis/blogs/state-line/2015/09/10/states-rush-to-regulate-drones-ahead-of-federal-guidelines>
24. Kang, C. FAA Drone Laws Start to Clash With Stricter Local Rules. *The New York Times* (2015), http://www.nytimes.com/2015/12/28/technology/faq-drone-laws-start-to-clash-with-stricter-local-rules.html?_r=0
25. Djapic, V., Galdorisi, G., Pels, J., Rodas, M., Volner, R.: Optimizing the human element in the age of autonomy. In: Proceedings of the Association for Unmanned Vehicle Systems International (AUVSI), Atlanta, GA, 4–7 May 2015
26. Hobbs, A., Shively, R.: Human factors guidelines for UAS in the national airspace system. In: Proceedings of the Association for Unmanned Vehicle Systems International (AUVSI), Washington, DC, 12–15 Aug (2013)
27. Cummings, M., Mastracchio, C., Thornburg, K., Mkrtchyan, A.: Boredom and dis-traction in multiple unmanned vehicle supervisory control. *Interact. Comput.* **25**(1), 34–47 (2013)
28. US Air Traffic Controllers Discuss Challenges, Solutions to UAS Integration. Flightglobal, 2015. Retrieved from: <https://www.flightglobal.com/news/articles/us-air-traffic-controllers-discuss-challenges-solutions-to-uas-integration-395238/>
29. Integration of Civil Unmanned Aircraft Systems (UAS) in the National Airspace System (NAS) Roadmap. U.S. Department of Transportation, Federal Aviation Administration. Retrieved from: http://www.faa.gov/uas/media/uas_roadmap_2013.pdf
30. McCarly, J., Wickens, C. (n.d.). Human Factors Implications of UAVs in the National Airspace, <http://www.tc.faa.gov/logistics/grants/pdf/2004/04-G-032.pdf>
31. Sternberg, R., Sternberg, K.: *Cognitive Psychology*, 7th edn. Cengage Learning, Boston (2017)
32. Van Erp, J.: Controlling unmanned vehicles: the human factors solution. RTO Meetings Proceedings 44 (RTO-MP-44), B8.1-B8.12 (2000)
33. Watts, A.C., Ambrosia, V.G., Hinkley, E.A.: Unmanned aircraft systems in remote sensing and scientific research: classification and considerations of use. *Remote Sens.* **4**(6), 1671–1692 (2012). doi:[10.3390/rs4061671](https://doi.org/10.3390/rs4061671)
34. Terwilliger, B., Vincenzi, D., Ison, D.C., Smith, T.D.: Assessment of unmanned aircraft platform performance using modeling and simulation (paper no. 15006). In: Proceedings of the 2015 Interservice/Industry Training, Simulation, and Education Conference (IITSEC), Arlington, VA: National Training and Simulation Association (2015)
35. Kopsaftopoulos, F., Nardari, R., Li, Y.H., Wang, P., Ye, B., Change, F.K.: Experimental identification of structural dynamics and aeroelastic properties of a self-sensing smart composite wing. In: Proceedings of the 10th International Workshop on Structural Health Monitoring, Stanford, CA (2015)
36. Material witness – the clever components transforming UAVS (2012). London, UK: Kable Intelligence Limited
37. Ratto, M., Ree, R.: Materializing information: 3D printing and social change. *First Monday*, **17**(7) (2012). doi:<http://dx.doi.org/10.5210/fm.v17i7.3968>

38. Federal Aviation Administration: Fact Sheet: Unmanned Aircraft Systems (UAS). Author, Washington (2015). http://www.faa.gov/uas/faq/media/1009_UAS_Fact_Sheet.pdf
39. John A.: Volpe National Transportation Systems Center: Unmanned Aircraft System (UAS) Service Demand 2015–2035, Literature Review & Projections of Future Usage. Cambridge, MA: author (2013). http://ntl.bts.gov/lib/48000/48200/48226/UAS_Service_Demand.pdf
40. Vincenzi, D., Terwilliger, B., Ison, D.C.: Unmanned aerial system (UAS) human-machine interfaces: new paradigms in command and control. Paper presented at the 6th international conference on applied human factors and ergonomics, Las Vegas, NV (2015)
41. Hoang, T.B., Akselrod, G.M., Argyropoulos, C., Huang, J., Smith, D.R., Mikkelsen, M.K.: Ultrafast spontaneous emission source using plasmonic nanoantennas. *Nat. Commun.* **6** (2015). doi:10.1038/ncomms8788
42. Leuthold, J., Hoessbacher, C., Muehlbrandt, S., Melikyan, A., Kohl, M., Koos, C., Freude, W., Dolores-Calzadilla, V., Smit, M., Suarez, I., Martínez-Pastor, J., Fitrakis, E.P., Tomkos, I.: Plasmonic communications: light on a wire. *Opt. Photonics News*, **24**(5), 28–35 (2013). doi:10.1364/OPN.24.5.000028
43. González-Espasandín, O., Leo, T.J., Navarro-Arevalo, E.: Fuel cells: a real option for unmanned aerial vehicles propulsion. *Sci. World J.* **2014** (2014)
44. Boeing: Boeing Delivers Reversible Fuel Cell-based Energy Storage System to U.S. Navy. Author, Chicago (2016). <http://boeing.mediaroom.com/2016-02-08-Boeing-Delivers-Reversible-Fuel-Cell-based-Energy-Storage-System-to-U-S-Navy>
45. New Solid Hydrogen-on-Demand Fuel Cell from HES Energy Systems Flies ST Aerospace UAV for Record 6 Hours. *Business Wire* (2016). <http://www.businesswire.com/news/home/20160216006857/en/Solid-Hydrogen-on-Demand-Fuel-Cell-HES-Energy-Systems>
46. Burrows, L.: A Metal that Behaves Like Water. Harvard, John A. Paulson School of Engineering and Applied Sciences, Cambridge. <https://www.seas.harvard.edu/news/2016/02/metal-that-behaves-like-water>
47. Mearian, L.: Samsung, MIT say their solid-state batteries could last a lifetime. *Computer world* (2015). <http://www.computerworld.com/article/2973483/sustainable-it/samsung-mit-say-their-solid-state-batteries-could-last-a-lifetime.html>
48. Liang, B., Liu, Y., Xu, Y.: Silicon-based materials as high capacity anodes for next generation lithium ion batteries. *J. Power Sour.* **267**, 469–490 (2014). doi:10.1016/j.jpowsour.2014.05.096
49. Laser Powers Lockheed Martin’s Stalker UAS For 48 Hours. *PR Newswire* (2012). <http://www.prnewswire.com/news-releases/laser-powers-lockheed-martins-stalker-uas-for-48-hours-162060735.html>
50. Test Sites. https://www.faa.gov/uas/legislative_programs/test_sites/
51. Bailey, A.: FAA awards Embry-Riddle, ASSURE UAS research funds. *UAS Magazine* (2015) <http://www.uasmagazine.com/articles/1293/faa-awards-embry-riddle-assure-uas-research-funds>
52. Association for Unmanned Vehicle Systems International: AUVSI underscores priorities for FAA reauthorization to advance the commercial UAS industry. Washington, DC: author (2016). <http://www.auvsi.org/browse/blogs/blogviewer?BlogKey=92a78695-90c7-4910-8506-5bb2c05d0c89>
53. Kopardekar, P.: Safely Enabling Low-Altitude Airspace Operations: Unmanned Aerial System Traffic Management (UTM). National Aeronautics and Space Administration, Washington (n.d.)
54. Department of Transportation, Federal Aviation Administration, Notice of Proposed Rule making [Docket No.: FAA-2015-0150; Notice No. 15-01], https://www.faa.gov/regulations_policies/rulemaking/recently_published/media/2120-AJ60_NPRM_2-15-2015_joint_signature.pdf

Mission Capabilities—Based Testing for Maintenance Trainers: Ensuring Trainers Support Human Performance

Alma M. Sorensen, Mark E. Morris and Pedro Geliga

Abstract Historically, Unmanned Aerial Systems (UAS) Training Device (TD) testing emphasized meeting performance specification with an insignificant level of attention paid towards mission capabilities-based requirements. Recent research focusing on effectiveness and efficiency of UAS TDs in supporting learning of critical skills has brought mission capabilities-based requirements testing to the forefront. This paper provides a review of the process and results of the MQ-8B Fire Scout Avionics Maintenance Trainer (AMT) MCT event. The Fire Scout AMT MCT event produced qualitative and quantitative data associated with the AMT's ability to support teaching of tasks/Learning Objectives (LOs) specific to maintenance. The multi-competency test team collected two types of data; one type measuring the capability of TD to provide critical attributes to support each task/LO and the other measuring the capability of TD to facilitate task/LO completion.

Keywords Training · Training devices · Mission capabilities · Maintenance

1 Introduction

Throughout history Naval aviation has been at the forefront dominating the air environment and remaining a critical component of military strategy and success. Within recent years, naval aviation has evolved from manned air systems to Unmanned Aerial Systems (UAS). With the steady and increasing reliance on UAS

A.M. Sorensen (✉) · M.E. Morris · P. Geliga
Naval Air Warfare Center Training Systems Division (NAWCTSD),
12211 Science Drive, Orlando, FL 32826, USA
e-mail: alma.sorensen@navy.mil

M.E. Morris
e-mail: mark.e.morris@navy.mil

P. Geliga
e-mail: pedro.geliga@navy.mil

to extend our military's eyes and improve targeting of aircraft and helicopters [1] UAS training is critical to ensure the United States' military power maintains a competitive edge. Due to the complexity of UAS, the military is turning to the acquisition of full fidelity simulators, training devices (TD), and part task trainers to ensure mission readiness. This training paradigm adds a complexity to the training environment resulting in the need to change the Navy evaluates full fidelity simulators, TDs, and part task trainers. Historically, emphasis has been placed on ensuring a TD meets performance specification (e.g., a light turns red when it is supposed turn red), while an insignificant level of attention has been granted towards evaluating the mission capabilities-based requirements (e.g., the light turning red during the executing of a troubleshooting task lead a student to complete the procedure correctly). Within recent years, initiatives to ensure mission capabilities-based requirements are met have been brought to the forefront as research focuses on how effective and efficient the systems are in supporting learning of critical skills. To accomplish this, the Department of Defense (DoD) has begun to change its system of evaluation to better support capabilities-based testing (Davis 2002). One area of focus includes conducting mission capabilities-based testing (MCT) of TDs to determine if they support learning and enable transfer and retention of learning. Successful completion of MCT assists in TD certification and allows for identifying potential deficiencies and future upgrade recommendations. To date, the focus has been on conducting MCT on aircraft operators TDs. However, MCT of a TD for aircraft maintenance is equally critical. This paper describes the recent successful completion of the MQ-8B Fire Scout Avionics Maintenance Trainer (AMT) MCT event and provides initial results of MCT data specific to an aircraft maintenance TD.

2 MCT Process Utilized

The six-phased process employed for test planning, execution, and report. Figure 1 provides a graphical illustration of the process discussed in the following sections.



Fig. 1 MCT process

2.1 Phase 1. Prepare Test Plan

Phase 1 consisted of writing a test plan. The test plan documented the scope, methods, and limitations for validating the capabilities of the AMT to support the MQ-8B Fire Scout Electronic Technician course. The test plan provided a systematic approach to planning and ultimately conducting an effective and safe test event. Prior to commencing any test event the Navy requires an approved Test Plan [2].

2.2 Phase 2. Prepare for Criticality Workshop

Phase 2 focused on preparing for the criticality workshop. First, instructional analysts reviewed training analysis data documented during the MQ-8B Fire Scout Maintenance Electronics Technician Front-End Analysis (FEA) (e.g., Media Report) to determine which Learning Objectives (LOs) would be designated to be taught, practiced, and/or assessed via the AMT. Once this was established, instructional analysts reviewed previous specification and regression-testing results to determine initial priority list for MCT. If a test procedure not completed during specification testing traced to a LO, then the LO received higher priority to ensure testing during MCT. After reviewing the training analysis data, Instructional analysts identified eight categories consisting of 22 simulator attributes for relevancy towards the AMT. Figure 2 identifies the baseline simulator attributes considered. The Integrated Product Team (IPT) down-selected the attributes based on functional requirements of the AMT. The IPT determined the following attributes (shown in bold font in Fig. 2) applied to the AMT MCT event: appearance, tactile feel, haptic

Physical Look & Feel	Function	Aero Models	Visual
<ul style="list-style-type: none"> • Appearance • Tactile Feel • Environment 	<ul style="list-style-type: none"> • Haptic Cues • System Response • System Delay 	<ul style="list-style-type: none"> • Own Aircraft Aerodynamic • Other Aircraft/Object Aerodynamic 	<ul style="list-style-type: none"> • Static • Active • Resolution • Depth • Light Level
Field of View	Auditory	Motion	Instructional Tools & Systems
<ul style="list-style-type: none"> • Horizontal • Vertical 	<ul style="list-style-type: none"> • Environment • Verbal Communications • Audible Signals • Sound Bearing 	<ul style="list-style-type: none"> • Aircraft Vibration • Aircraft Shake • Aircraft Surround Motion 	<ul style="list-style-type: none"> • IoS

Fig. 2 Categories and simulator attributes

cues, system response, system delay, audible signals, and Instructor Operating Station (IoS). Finally, instructional analysts derived a criticality scale using a Likert Rating Scale, where 0 = Not Applicable, 1 = Not Critical, 2 = Nice But NOT Critical, 3 = Critical, 4 = Highly Critical, and 5 = Absolutely Critical and created the corresponding data collection tool.

2.3 Phase 3. Conduct Criticality Workshop

Phase 3 involved conducting the criticality workshop. A virtual criticality workshop, organized and facilitated by members of the Fire Scout IPT, utilized Fire Scout fleet maintainers with prior deployment experience and included representation of the Fire Scout maintenance schoolhouse. Subject Matter Experts (SMEs) systematically rated each attribute for each LO using the established criticality scale mentioned Phase 2. Based upon the criticality rating received, instructional analysts reevaluated priority levels ensuring LO requiring multiple attributes received higher priority during the MCT event.

2.4 Phase 4. Prepare for MCT

Phase 4 concentrated on preparations for the MCT event. Preparing for the testing event included developing test cards for each LO that listed procedures copied from the Interactive Electronic Technical Manuals (IETMs); those manuals maintenance personnel use to perform preventive and corrective maintenance.. Simultaneously, instructional analysts developed two questionnaires using a Likert Rating Scale, where 1 = Completely Incapable, 2 = Marginally Incapable, 3 = Moderately Capable, 4 = Effectively Capable, and 5 = Fully Capable. The first questionnaire measured the TD's ability to replicate the critical attributes identified for each LO during the criticality attribute workshop. The second questionnaire, developed using a modified-Cooper Harper [3, 4] questionnaire, assessed the TDs overall capability to execute the LO. Instructional analysts developed a testing manual consisting of a test card and the two questionnaires for each LO.

2.5 Phase 5. Conduct MCT

Phase 5 involved conducting MCT. SMEs, consisting of both fleet maintainers (2) with deployment experience and schoolhouse instructors (3), executed each test procedure by system (e.g., Communications, Vibration Monitor, Navigation) in accordance with the test cards and IETMs. Upon completion of each test card, instructional analysts facilitated data collection by utilizing the two questionnaires

developed in Phase 4. Instructional analysts collected supplemental data regarding rationale for each Likert Scale rating collected requiring courseware specific instructor interventions related to any identified lack of TD capability. The data collected via the questionnaires also allowed the Test Director to determine and issues list related to the TD's capability and report, officially, within an Interim Summary Report (ISR) as part Phase 6.

2.6 Phase 6. Report Results

Instructional analysts reviewed the data collected during MCT. Instructional analysts provided the data, raw and analyzed, to the Test Director to support issue identification, lessons learned, and recommendations. The Test Director consolidated the issues, lessons learned, and recommendations into an ISR and promulgated it up through the NAVAIR Test Directorate and the MQ-8 Fire Scout program office for review and approval. Issues identified in the ISR are processed and published into Deficiency Reports (DRs) and reviewed, approved, and promulgated via a Report of Test Results (RTR). in accordance with Test Reporting Policy for Air Vehicles, Air Vehicle Weapons, and Air Vehicle Installed Systems NAVAIRINST 3905.1 [5].

3 Results

The following subsections discuss the results of the two data gathering events: the Criticality Attribute Workshop and the MCT event.

3.1 Criticality Attribute Workshop Results

The criticality attribute workshop focused on identifying which of seven attributes the TD should replicate for all 65 LOs to facilitate transfer of knowledge. The data collected during the criticality attribute workshop allowed analysts to prioritize the LOs for the MCT event. Table 1 illustrates the percentage of LOs by rating for criticality attribute (n = 8). Analysts collected data via consensus.

Table 1 identifies the three highest rated attributes critical to the TD facilitating and promoting knowledge of transfer are appearance, tactile feel, and haptic cues. The highest rated attributes, appearance, tactile feel, and haptic cues, received a 5 rating (Absolutely Critical) for 55 (82 %) of the LOs while 12 (18 %) of the LOs received a 3 (Critical) rating. The next highest rated attribute, Systems Response Interaction, received a 5 (Absolutely Critical) rating for 45 (67 %) of the LOs while 21 (31 %) of the LOs received a 2 (Nice But NOT Critical) rating. The third highest

Table 1 LOs percentage results for criticality attribute ratings

Attribute	Likert scale rating					
	5 (%)	4 (%)	3 (%)	2 (%)	1 (%)	0 (%)
Appearance (physical properties)	82	0	18	0	0	0
Tactile feel (touch sensation)	82	0	18	0	0	0
Haptic cues (kinesthetic response)	82	0	18	0	0	0
Systems Response Interaction	67	0	0	31	0	0
System delay	49	0	3	15	1	31
IoS	21	0	0	43	0	36
Audible signals	0	0	0	3	0	97

rated attribute, System Delay, received a 5 (Absolutely Critical) rating for 33 (49 %) of the LOs, a 3 (Critical) rating for 2 (3 %) of the LOs, a 2 (Nice But NOT Critical) rating for 10 (15 %) of the LOs, a 1 (Not Critical) for 1 (1 %) of the LOs, and a 0 (Not Applicable) rating for 21 (31 %) of the LOs. Next, the IoS attribute received a 5 (Critical) rating for 14 (21 %) of the LOs, a 3 (Critical) rating for 29 (43 %) of the LOs, and a 0 (Not Applicable) rating for 24 (36 %) of the LOs. The lowest rated attribute, Audible Signals, received a 2 (Nice But NOT Critical) rating for 2 (3 %) of the LOs while the rest of the LOs, 65 (97 %) received a 0 (Not Applicable) rating.

The attribute criticality results can be understood by considering the behavior of the LOs relevant to Fire Scout maintenance. The majority of the LOs, 21 (31 %) required the removal and installation of the AV’s components. Using the TD as a media to facilitating knowledge transfer of LOs focusing on removal and installation of components requires those components to have the physical properties, touch sensation, and kinesthetic responses of the real AV. The next highest grouping of LOs corresponds to troubleshooting or testing components on the AV. LOs requiring troubleshooting or testing each totaled 12 (18 %). Using the TD as a media to facilitate knowledge transfer of LOs focusing on troubleshooting and testing an AV’s components required accurate systems responses and system delay in receiving data. Combined, these three (3) types of LOs, removal and installation, troubleshooting, and testing, accounted for 45 (67 %) of the LOs, as shown in Table 2.

Instructional analysts updated the LOs testing prioritization to ensure testing of critical LOs occurred during the MCT event timeframe. The re-prioritization of LOs considered the frequency of ratings received during the criticality attribute workshop. LOs with four (4) or more attributes receiving a 5 (Absolutely Critical) rating received Priority I re-classification. LOs with three (3) attributes receiving a 5 (Absolutely Critical) rating received Priority II re-classification. LOs with fewer than three (3) attributes receiving a 5 (Absolutely Critical) rating received Priority III re-classification. Using the re-prioritized list of LOs, analysts conducted the MCT event. The following subsection discusses the results of the MCT event and the two types of data, Critical Attribute Results and TD Capability Results, collected.

Table 2 Breakdown between LO behavior and number of LOs/percentage

LO action	Number of LOs	Percentage (%)
Removal and installation	21	31
Troubleshoot	12	18
Test	12	18
Monitor	3	4
Adjust	1	1
Backup	1	1
Calibrate	1	1
Clean	1	1
Conduct	1	1
Configure	1	1
Power up	2	3
Utilize	1	1
Verify	2	3
Download	1	1
Shut down	2	3
Load	3	4
Operate	1	1
Perform	2	3

3.2 MCT Event—Critical Attribute Results

The critical attribute results focused on how well the TD represented the specific attribute in such a way that transfer and retention of knowledge occurred. At a high level, of the LOs tested, 31 (46 %) LOs received a rating of 5 (Fully Capable) for all attributes, seven (10 %) LOs received no 5 (Fully Capable) rating at all, and the mode rating was 3 (Moderately Capable) (n = 5). During testing analysts did not collect data on five LOs; the five LOs, removed due to their removal from the corresponding course are listed below.

- Clean Boresight Module
- Power up Air Data Subsystem
- Replace Coastal Battlefield Reconnaissance and Analysis (COBRA) Airborne Payload System (CAPS) Payload Housing Group (PHG)
- Replace COBRA CAPS Payload Support Mount

Because some attributes, during the criticality attribute workshop, received TD capability ratings below three (3), analysts did not collect data on all attributes for each LO. Taking into account the removal of the five (5) LOs identified above the maximum number of LOs tested equaled 62; for some attributes analysts collected no, or minimal, data. Table 3 identifies the number of LOs for which analysts

Table 3 Breakdown of attribute by LO

Attribute	Number of LOs
Appearance (physical properties)	62
Tactile feel (touch sensation)	62
Haptic Cues (kinesthetic response)	62
Systems response interaction	43
System delay	33
IoS	14
Audible signals	0

Table 4 Percent of LOs the TD adequately replicates by attribute

Attribute	Likert scale rating percentage				
	5 (%)	4 (%)	3 (%)	2 (%)	1 (%)
Appearance (physical properties)	89	2	5	2	3
Tactile feel (touch sensation)	89	2	5	2	3
Haptic Cues (kinesthetic response)	79	3	5	5	8
Systems response interaction	29	13	16	2	10
System delay	44	0	0	2	8
IoS	6	0	16	0	0
Audible signals	0	0	0	0	0

collected data, by attribute. Table 4 provides a breakdown of the TD's attributes representation capability by attribute.

As Table 4 indicates the appearance and tactile feel attributes received the highest ratings, followed closely by the haptic cues attribute. For the appearance and tactile feel attributes, data analysis yielded identical results. Of the 62 LOs tested, the appearance and tactile feel attributes received a 5 rating (Fully Capable) on 55 (89 %) LOs, a 4 (Effectively Capable) rating on one (2 %) LO, a 3 (Moderately Capable) rating on three (5 %) LOs, a 2 (Marginally Incapable) rating on one (2 %) LO, and a 1 (Completely Incapable) rating on two (3 %) LOs. Closely following appearance and tactile feel is the haptic cues attribute. For haptic cues, 49 (79 %) of the 62 LOs tested received a 5 (Fully Capable) rating, two (3 %) LOs received a 4 (Effectively Capable) rating, three (5 %) LOs received a 3 (Moderately Capable) rating, three (5 %) LOs received a 2 (Marginally Incapable) rating, and five (8 %) LOs received a 1 (Completely Incapable) rating. After haptic cues, system delay is the highest rated attribute. Per the results of the criticality attribute workshop and the removal of the above listed LOs, analysts only collected data for the system delay attribute for 33 LOs. Of the 33 LOs tested for system delay, 27 (44 %) received a 5 (Fully Capable) rating. Of the remaining eight LOs, one (2 %) LO received a 2 (Marginally Incapable) rating, and five (8 %) LOs received a 1 (Completely Incapable) rating. After the system delay attribute, the highest rated attribute is systems response interaction. Because of the results of the criticality

workshop and the removal of the aforementioned LOs, analysts only collected data associated with the systems response interaction attribute on 43 LOs. Of the 43 LOs tested, 18 (29 %) received a 5 (Fully Capable) rating, eight (13 %) LOs received a 4 (Effectively Capable), 10 (16 %) LOs received a 3 (Moderately Capable) rating, one (2 %) LO received a 2 (Marginally Incapable) rating, and six (10 %) LOs received a 1 (Completely Incapable) rating. After system response interaction, the next highest attribute is IoS. Because of the criticality workshop results analysts collected data associated with the IoS attribute on only 14 LOs. Of the 14 LOs tested, four (4) (6 %) LOs received a 5 (Fully Capable) rating and 10 (16 %) LOs received a 3 (Moderately Capable) rating. Based on the criticality workshop data analysis results, analysts did not collect attribute data on any LOs related to audible signals. Review of the data identifies the highest rated LOs are related to tasks requiring removal and installation of AV components and are closely followed by those LOs associated with troubleshooting and testing AV components. The critical attribute results indicated the TD is capable of supporting the execution of the LOs in a learning environment either fully as currently built or with minimal compensation. There are some LOs, more specifically associated with the Systems Response Interaction LOs, that require considerable compensation or workarounds, but few contain severe deficiencies or detractors from realistic execution and only two tasks are completely unsupported by the TD.

3.3 MCT Event—TD Capability Results

Analysts also evaluated the TD's capability to support overall LO execution. To do this, analysts used a modified-Cooper Harper scale to ask the SMEs if, using the TD, (1) they completed the task, (2) they complete portions of the task, (3) complete the task with workarounds, and/or (4) the TD contains the fidelity and functionality of a Non Fully Mission Capable (NFMC) MQ-8B Fire Scout. Table 5 provides a breakdown of the ratings received through execution of the modified-Cooper Harper questionnaire.

Of the 62 LOs tested, 34 (55 %) of the LOs received a 5 (Fully Capable) rating indicating the TD contains the fidelity and functionality of a NFMC MQ-8B Fire Scout and is fully capable of providing training specific to the LO with no departure from realism and requires no compensation to support LO execution. Five (8 %) of the 62 LOs received a 4 (Effectively Capable) rating indicating the TD contains the fidelity and functionality of a NFMC MQ-8B Fire Scout and is reasonably capable of providing training specific to the LO with minor, but annoying deficiencies, requiring minimal compensation to support LO execution. Five (8 %) of the 62 LOs tested received a 3 (Moderately Capable) rating indicating the task was completed but required workarounds and the TD is borderline capable of providing training specific to the LO with moderate deficiencies requiring considerable compensation to support LO execution. The list below identifies examples of the workarounds executed.

Table 5 Rating, definition, task complete, LOs, and percentage breakdown

Rating	Definition	Task completed	LOs	%
5—fully capable	Training device is fully capable of providing training specific to the LO with no departure from realism and requires no compensation to support LO execution	Yes, training device contains the fidelity and functionality of a NFMC MQ-8B Fire Scout	34	55
4—effectively capable	Training device is reasonable capable of providing training specific to the LO with minor, but annoying deficiencies, requiring minimal compensation to support LO execution	Yes, training device contains the fidelity and functionality of a NFMC MQ-8B Fire Scout	5	8
3—moderately capable	Training device is borderline capable of providing training specific to the LO with moderate deficiencies requiring considerable compensation to support LO execution	Yes, but with work arounds	5	8
2—marginally incapable	Training device is marginally incapable of providing training specific to the LO with severe deficiencies and detracts appreciable from realistic LO execution	No, only portions of task completed	16	26
1—completely incapable	Training device is completely incapable of providing training specific to the LO as presented and precludes LO execution	No	2	3

- **Lack of Consumables:** Instructors provided direction to student maintainers for steps requiring the use of materials not available for test execution (e.g., isopropyl alcohol, self-leveling sealant, electromagnetic interference gasket)
- **Lack of Equipment:** Instructors provided direction to student maintainers for steps requiring specific test equipment not available for test execution (e.g., test sets and torque wrench)
- **Lack of Functionality:** Instructors provided direction to student maintainers for steps related to a lack of functionality in AMT sub-systems not designed to provide the haptic cues, system responses, and systems delays representative of a fleet AV (e.g., corrode components, no power).

Of the 62 LOs tested, 16 (26 %) received a 2 (Marginally Incapable) rating indicating SMEs only completed portions of the LO and the TD is marginally incapable of providing training specific to the LO and has severe deficiencies which detracts appreciably from realistic LO execution. Two (3 %) of the 62 LOs tested received a 1 (Completely Incapable) rating indicating SMEs were unable to

complete the task and that the TD is completely incapable of providing training specific to the LO as presented and precludes LO execution. Overall, the majority (76 %) of tasks/LOs requiring removal and installation supported learning via the AMT whereas nearly half (41 %) of the tasks/LOs focusing on testing or troubleshooting were unable to fully support learning via the AMT.

4 Lessons Learned

Table 6 provides lessons learned identified in association with development of future MCT events.

Table 6 Lessons learned and explanation

Lesson learned	Explanation
Create test book for all future testing events	The test team created a comprehensive test book that included a specific test procedure followed by the correlating questionnaires that covered that test procedure. This increased efficiencies and ensured pertinent data did not get lost, omitted, or skipped
Use mission-based testing where applicable	Group multiple objectives and or test procedures in one test to increase efficiency and replicate actual task performance. For instance, performing a troubleshooting procedure may include performing a test that leads to a remove and replace
Have courseware available if possible	Reviewing courseware while conducting test procedures will ensure courseware aligns to training device capability and mitigates potential problems that may emerge while piloting or beginning a course
Have robust test cards	Test cards should include the following in addition to the procedure and procedures manual location: <ul style="list-style-type: none"> • Media type along with corresponding courseware location (section/page number) • A column for notes pertaining to alignment of courseware and steps • A column for notes identifying if an instructional workaround is needed • A courseware note block at the end of each test card for quick reference related to courseware issues/changes
Increase usability of attribute questionnaire	Optimize the questionnaire through: <ul style="list-style-type: none"> • Remove the rating definitions from the questionnaire • Change orientation to landscape for increased note taking space • Create laminated quick reference sheets for both the rating and attribute definitions
Conduct pre-test briefs to increase attribute rating reliability	Prior to conduct of each test, brief the participants on the relevant attributes so they are thinking about the properties data will be collected on
Tailor attribute definitions	Tailor the attribute definitions to the specific platform and the specific roles of the testers to increase reliability of ratings
Create a communication plan prior to testing event	The plan should include at a minimum: stakeholders to maintain contact with, frequency of contact, and form of communication

5 Conclusion

With increased complexity and demand for UAS use within the Fleet, a corresponding demand for TDs to train UAS maintainers has brought MCT to the forefront. The paradigm shift emphasizing MCT events focuses on ensuring both operator- and maintainer-based TDs meet system performance specifications and mission requirements. The successful completion of the MQ-8B Fire Scout AMT MCT event constitutes the first maintenance-based MCT event of its kind. The teamwork and processes incorporated build a bridge for similar efforts in the future. The high-level framework described in the paper provides a foundation for other maintenance-based UAS IPTs to conduct maintenance-based MCT events. The data collection workshop and data collection tools used for the MQ-8B Fire Scout AMT MCT event also provide a systematic approach for conducting maintenance-based MCT events for the U.S. Navy.

References

1. Bone, E., Bolkcom, C.: Unmanned aerial vehicles: background and issues for congress. Congressional Research Service—The Library of Congress (2003). Retrieved from: <http://fas.org/irp/crs/RL31872.pdf>
2. Department of the Navy. Project Test Plan Policu and Guide for Testing Air Vehicles, Air Vehicle Weapons, and Air Vehicle Installed Systems Naval Air System Command No. NAVAIRINST 3960.4B) (2005). Retrieved from: http://www.navair.navy.mil/nawcwg/vx31/documents/NAVAIRINST3960_4.pdf
3. Cummings, M.L., Myers, K., Scott, S.D. Proceedings from UVS Canada: Conference on Unmanned Vehicle Systems Canada '06: Modified Cooper Harper Evaluation Tool for Unmanned Vehicle Displays. Montebello, PQ, Canada (2006). Retrieved from: http://web.mit.edu/aerastro/labs/halab/papers/Cummings2006_uvs_FINAL.pdf
4. Donmez, B., Brzezinski, A. S., Graham, H., Cummings, M.L. Modified Cooper Harper Scales for Assessing Unmanned Vehicle Displays. Massachusetts Institute of Technology Humans and Automation Laboratory (2008). Retrieved from: http://web.mit.edu/aerastro/labs/halab/papers/Metrics_Report_Final.pdf
5. Department of the Navy.: Test Reporting Policy for Air Vehicles, Air Vehicle Weapons, and Air Vehicle Installed Systems (Naval Air System Command No. NAVAIRINST 3905.01) (2013)

Detecting Deictic Gestures for Control of Mobile Robots

Tobias Nowack, Stefan Lutherdt, Stefan Jehring, Yue Xiong, Sabine Wenzel and Peter Kurtz

Abstract For industrial environments esp. under conditions of “Industry 4.0” it is necessary to have a mobile and hands-free controlled interaction solution. Within this project a mobile robot system (for picking, lifting and transporting of small boxes) in logistic domains was created. It consists of a gesture detection and recognition system based on Microsoft Kinect™ and gesture detection algorithms. For implementing these algorithms several studies about the intuitive use, executing and understanding of mid-air-gestures were processed. The base of detection was to define, if a gesture is executed dynamically or statically and to derive a mathematical model for these different kinds of gestures. Fitting parameters to describe several gesture phases could be found and will be used for their robust recognition. A first prototype with an implementation of this technology also is shown in this paper.

Keywords Human-robot-interaction · Mid-air-gestures · Deictic gestures (Pointing) · Definition of gestures · Kinect 2™

T. Nowack (✉) · S. Jehring · Y. Xiong · S. Wenzel · P. Kurtz
Ergonomics Group, Technische Universität Ilmenau, Max-Planck-Ring 12,
98693 Ilmenau, Germany
e-mail: Tobias.Nowack@tu-ilmenau.de

S. Jehring
e-mail: Stefan.Jehring@tu-ilmenau.de

Y. Xiong
e-mail: Yue.Xiong@tu-ilmenau.de

S. Wenzel
e-mail: Sabine.Wenzel@tu-ilmenau.de

P. Kurtz
e-mail: Peter.Kurtz@tu-ilmenau.de

S. Lutherdt
Biomechatronics Group, Technische Universität Ilmenau, Max-Planck-Ring 12,
98693 Ilmenau, Germany
e-mail: Stefan.Lutherdt@tu-ilmenau.de

1 Introduction

Since motion capturing systems are available for about 200 €, the use of mid-air gestures will yet be interesting in industrial environments. For human-to-human interaction people use several different gestures like pointing to inform others where to search for anything or waving to call somebody with intention to force the other one to come over or even to follow. For interaction within technical systems like mobile robotic systems, it is necessary to have or create equivalent intuitive gestures to control these systems.

The first communication channel is of course speech, which is underlined by gestures to bring more emphasizes into it. At the same time these used gestures also will be additionally declared or precise the intention of gestures by the context information given by speech.

For phone or office applications, technical speech recognition is well developed and broadly used. But the use of speech recognition in industrial environments with noise levels up to 85 dB(A), which are allowed by German laws for production areas [1], is not applicable or only realizable with additional microphones (in front of the mouth or larynx).

Primary and most beneficial applications for human-robot-collaboration in the field of “Industry 4.0” are aided “manufacturing by hand”. These tasks necessarily needs flexible interaction between the human being and the technical system “robot” and should not be interfered by an additionally handling of microphones or other input devices.

2 Propaedeutics

All types of locomotion may be divided into three main parts [2]. For speech attended gestures Kendon [3] called this three main parts preparation, stroke and recovery. McNeill [4] has classified types of gesture movements during the stroke phase in the four major categories: iconic, metaphoric, deictic (pointing) and beat gestures. Metaphoric gestures will be a good choice for interacting purposes between human and machine if the involved staff can learn the special commands like a special language for their job. Diving staff or members of Special Forces use this kind of gestures for communication and controlling. For interacting with robots and real objects from the environment in collaborative tasks deictic gestures should be the preferred way. It prevents from undesirable failures and minimizes the learning effort of workers.

Kendon [3] describes gestures as speech attended, so gestures only attend additional information, the main content will presented on the acoustic channel. But if gestures should be used for controlling robots in noisy industrial environments, these main information have to be transferred only by gestures. Speech attended gestures are introduced by the speaker and the audience is already focused on this

main actor. For technical systems, in this case robots, the focus problem is similar to “the live mic problem” [5] of speech recognition. To start the “gesture control” a “wake up” and “identify the user” algorithm is needed. If control user and robot are paired, the main recognition algorithm e.g. pointing detection can work. With the cognitive performance of human beings, this wake up event could be detected implicitly. Also during other periods of gesture recognition humans use their cognitive performance and additional environmental (context) information to get a precise understanding of the command. E.g. if somebody searches a book and the communication partner points to bookshelf, then for orientation it is only necessary to know the rough direction (given by the arm and finger). The recipient is already able to detect the right one in detail by an intuitive knowing of the meaning of these gesture.

Another example for this intuitive identifying of gesture meaning in industrial environment is, if an apprentice watch to the master and know which tool is needed to complete the task successfully when the master point to the toolbox. As far as these context detection is not realized for technical systems, the gesture must contain all necessary information to specify the command.

Additional to the three main phases one up to three more phases are necessary to describe the complete gesture sequence. Figure 1 shows the two optional static moments pre- and post-stroke hold. Pavlovic [6] defined the preparation phase already as leaving the resting position and move the hand to the starting point. The recovery phase can be defined in the same behavior as leaving the gesture area and returning to the resting position.

When using the definition of the resting phase (home position) as hands on the hips, the preparation phase could be easily detected by leaving the home position (gesture start). Reentering the home position finishes the gesture (gesture end). The user presents the gesture above the hip, in front or beside the users body.

There are two different sensors available to detect the joints on the users body. For the mobile robotic system, an ASUS Xtion (similar technic like the Kinect 1™) with up to 15 frames per second (fps) and 15 joints, detected by OPEN NI (NITE) library, is used. With the new Microsoft Kinect 2™ sensor now 25 joints (see Fig. 2) can be re-corded with 30 fps.

To detect the user, the robot moves the ASUSsensor, mounted on a pan-tilt moving head. Because skeleton detection with the OPEN NI (NITE) library is very slow, a face detecting algorithm works on before and supports adjusting the camera

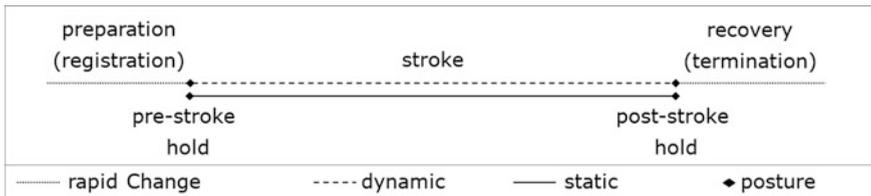
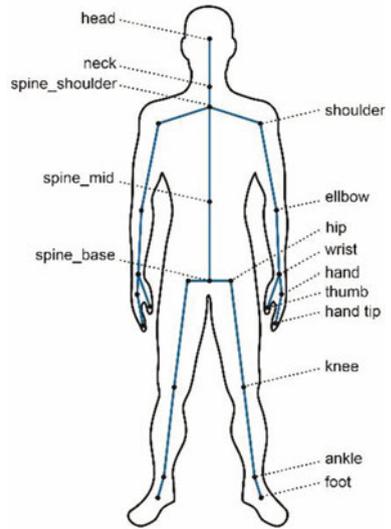


Fig. 1 Gesture phases [7]

Fig. 2 Joints detected by Microsoft Kinect 2™



system. To identify the user when more than one worker is in field of view the user lifts the arm on the height of or above the head (pre-stroke hold, see Fig. 8 left).

With the Kinect 2™ sensor the detection and representation as a skeleton is fast enough that without searching for users up to six persons in the field of view can be recognized already in the home position. An additional pre-stroke hold for identifying the control user is not implemented yet. Right now, the new sensor is only used for testing and upgrading the gesture identification algorithm.

Pavlovic [6] mentioned that “human gestures are a dynamic process, it is important to consider the temporal characteristics of gestures”. The moving speed during the different gesture phases will characterize the gesture. See in Fig. 1 rapid moves describe the preparation and recovery phase, whereas the stroke phase will be slower and more precise. In tests in 2015 79 % of the 90 participants preferred a static body posture to represent the command `SELECT` (pointing) [7]. Following the definition that the stroke phase is a dynamic process the post-stroke hold contained the information of a deictic gesture.

3 Gesture Description

As explained in Chap. 2 there are two general ideas how to describe the pointing gesture. There are a static definition, only recognized during the post-stroke hold and a dynamic definition, right now recognizing four parts of the gesture.

3.1 Static Gesture Description

Figure 3 shows the static definition (post-stroke hold) of pointing. After leaving the home position and register as user to control the robot, user will reach a static position of pointing. This body posture is defined as interaction area is reached (shoulder elevation $\alpha > 30^\circ$) and arm is stretched (elbow extension $\beta > 160^\circ$). The hold is characterized as a static period, minimum duration of 0.5 s.

During the period of static hold the difference of median filtered “shoulder angle” and “elbow angle” will not more than 5° each and median filtered “hand point” should not move more than 50 mm.

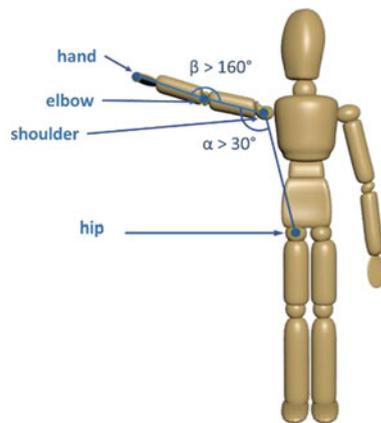
3.2 Dynamic Gesture Description

Defining a gesture as a dynamic process, it has to be discussed, what does exactly dynamic means? In mechanics, dynamics is concerning about the relationship between motion of bodies and the forces acting on the bodies. [8, 9]. There are two branches: kinematics and kinetics. Kinematics deals with observing motions and kinetics study motion and its relation to forces. The dynamic gesture description only deals with the kinematics.

To divide a gesture into different parts, descriptions for each part are needed. A threshold of an angle on a single frame describes some parts, like the static description; the curve shape during a sequence of frames will define other phases. The full gesture sequence contains (1) resting position, (2) start of gesture, (3) pointing, (4) recovery and (5) end of gesture.

The reference for all calculated angles is a vector between the spine-shoulder-joint and the spine-base-joint, detected with the Kinect 2TM sensor. For the resting position, a second vector was defined between the hand and the

Fig. 3 Static definition of pointing



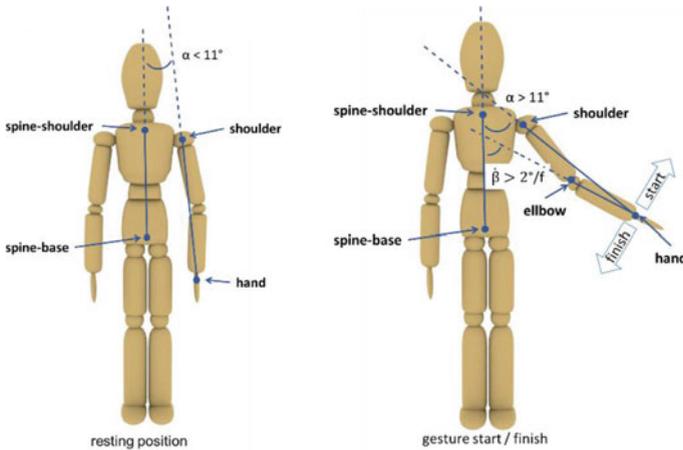


Fig. 4 Dynamic definition of a gesture

shoulder (see Fig. 4 left). If the hand is in the resting position both vectors should be parallel; for kinematics based on the human movement it makes sense to use a threshold value for angle α as less than 11° between the suggested vectors. With this vector definition all kind of arm position are possible, independent of the elbow angle. With about 10 experimentees and several experiments all threshold values are determined and an ongoing development will precise them.

A gesture starts in that moment angle α increases above 11° with an angular velocity β more than 2° per frame (see Fig. 4 right). The angle β is between the reference vector (torso) and the forearm, which is described by the hand and the elbow joints. The recovery phase finishes at the same moment, when angle α decreases under 11° , and reentering the resting position. Even though most of the probands had a great variety by executing the pointing gesture. It could be derived that all of them had used their forearm for this gesture (actually were reviewed data of about 10 experimentees). With this it is obvious to use this moving and positioning of forearm to detect (and describe) the pointing gesture even if it is executed in different ways like shown in Fig. 5.

The post-stroke hold contains the pointing part of the gesture (see Fig. 5). As explained in Chap. 2, this motionless phase contains the main information of a pointing gesture. Motionless means an angular velocity β less than 2° per frame kept holding at minimally five frames. As shown in Fig. 5 right this description fits both on pointing gestures performed with the stretched arm, similar to body posture works with the static description, as well as with a pointing gesture with bent elbow like shown in the Fig. 5 left.

It is necessary that in the chronological sequence (see Fig. 6) the recovery phase is followed directly by the pointing part (post-stroke hold). When working with such a short holding time (5 frames) it is possible to detect a pointing phase also if the user performs a wave gesture (command come over [7]) with a pre-stroke hold

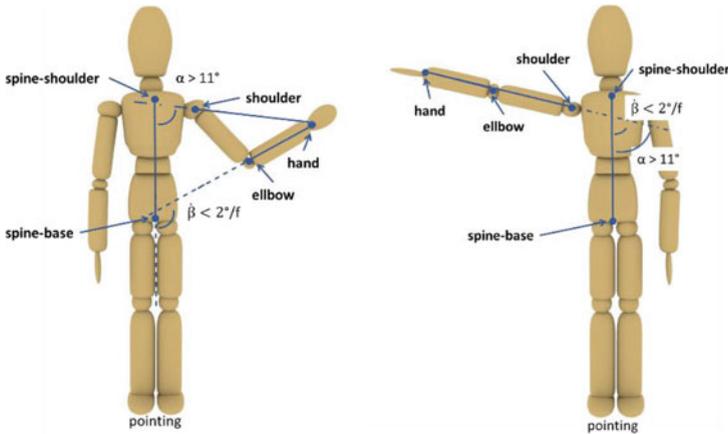


Fig. 5 Pointing (post-stroke hold)

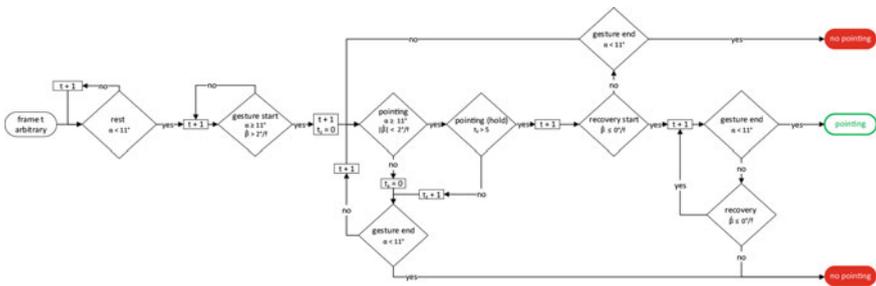


Fig. 6 Flow chart of dynamic gesture recognition sequence

similar to the pointing phase definition. During the recovery phase the hand joint leaves the position and moves with a monotonous decreasing of the elbow angle β until reaching the resting position.

After the occurrence of a complete sequence, the dynamic way of pointing can be confirmed.

3.3 Evaluation

A first group of six experimentees performed 122 gestures, 77 pointing gestures and 45 other, mainly wave gestures. These gestures had been recorded with a Kinect 2™ Sensor and evaluated in SciLab with algorithms according to the before introduced static and dynamic descriptions.

As shown in Fig. 7 the detection rate increases by using the dynamic definition from 54 % up to 68 %. The detection of pointing only increases by 5 % and the wrong detection of other gestures decreases by 9 %.

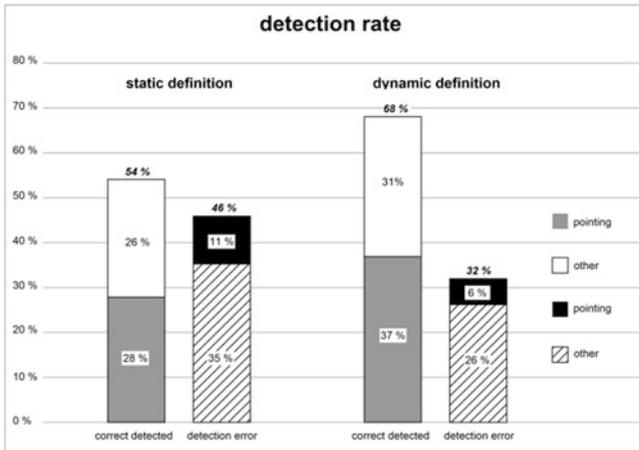


Fig. 7 Detection results of 122 gestures

4 Robot Control Concept

The static gesture recognition is used to control an assisting system for lifting and carrying goods, called KLARA. This system will support older or partially disabled workers to handle small containers up to 60 cm × 40 cm with a maximum weight of 50 kg. The main target group, elderly user with limited mobility in working environments should be able to command the system intuitively in three steps (see Fig. 8).

- Step 1: Register as the active user (register phase and pre-stroke hold).
- Step 2: Preselect an area with a container by using the pointing gesture (stroke phase and post-stroke hold).
- Step 3: Confirm the command on a touch screen.

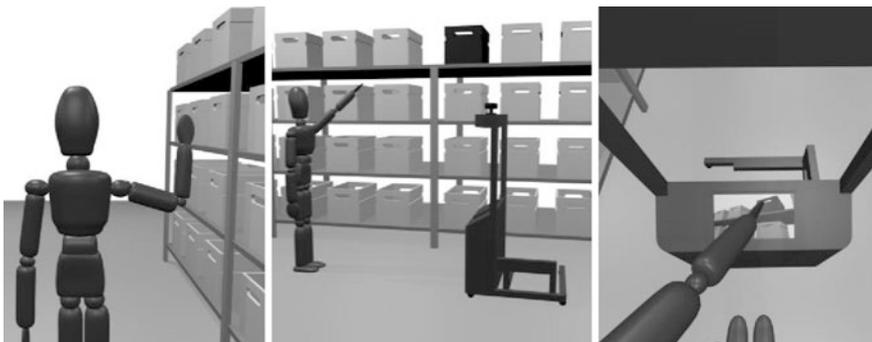


Fig. 8 Control sequence of the KLARA-robot-system: *left* register as an active user, *middle* preselect a container with a pointing gesture, *right* confirm the choice on touch screen

The KLARA-system has three control lights to support the users during the pointing sequence (Step 1 and 2). When the system is waiting for a command and searching the active user, a red light is on. Afterwards the system indicates that the active user was detected and is tracked. Now, indicated by a yellow light, the system signalizes that it is ready to recognize the pointing phase. During the post-stroke hold the system informs the user that the body posture of pointing is recognized and the system proves the period of holding an additional green light switched on. If the holding time is over the control light switch to green, the pan-tilt moving head turn the ASUS sensor in pointing direction and determine the selected container. Now the user can release the arm and should step forward to KLARA to confirm that this container on the touch screen is the chosen one.

After this confirmation, the system will move automatically forward and picks up the specified container autonomously.

5 Summary and Outlook

Gestures to control assisting systems for lifting and carrying small containers in industrial environments will allow a hands-free human-robot-interaction. Because gestures, especially pointing gestures, are common to enhance speech in human-to-human interaction the usage in human-robot-interaction can be intuitively for elderly user. However the context information in human-to-human interaction, for example the attended speech or other circumstances which could be recognized by cognitive skills, it have to be replaced by a technical description of body postures during the execution sequence.

The dynamic description including all gesture phases is the logical continuation from the static definition. Between gesture start and finish, also other gestures for example waving will be describable. For waving with fore- and upper arm, a sinus curve might characterize the angular movement of the elbow during the stroke phase, even when the pre- and post-stroke hold will be performed and detected similar to the post-stroke definition of pointing.

The next step to increase the detection rate for the pointing gesture will be a more precise identification of the threshold values in the definition. An important question in this case is, if the threshold values might be dynamically adaptable for each user. A detection rate of more than 80 % is achieved.

References

1. LärmVibrationsArbSchV: Verordnung zum Schutz der Beschäftigten vor Gefährdungen durch Lärm und Vibrationen (Lärm- und Vibrations- Arbeitsschutzverordnung - LärmVibrationsArbSchV), 06 March 2007 . http://bundesrecht.juris.de/bundesrecht/l_rmvibrationsarbschv/gesamt.pdf. Accessed March 2016

2. Meinel, K., Schnabel, G., Krug, J.: *Bewegungslehre – Sportmotorik Abriss einer Theorie der sportlichen Motorik unter pädagogischem Aspekt*, (11. überarb. und erw. Aufl. ed.), Meyer & Meyer, 2007
3. Kendon, A.: *Gesture Visible Action as Utterance*. Cambridge University Press, Cambridge (2004)
4. McNeill, D.: *Hand and Mind. What Gestures Reveal About Thought*. University of Chicago Press, Chicago (1992)
5. Walter, R., Bailly, G., Müller, J.: *StrikeAPose: Revealing mid-air gestures on public displays*. In: Wendy E. Mackay und A. Special Interest Group on Computer-Human Interaction (Hg.): *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. [S. 1.]: ACM, pp. 841–850 (2013)
6. Pavlovic, V.I., Sharma, R., Huang, T.S.: *Visual interpretation of hand gestures for human-computer interaction: a review*. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 677–695 (1997)
7. Nowack, T., Suzaly, N., Lutherdt, S., Schürger, K., Jehring, S., Witte, H., Kurtz, P.: *Phases of technical gesture recognition*. In: M. Kurosu (Ed.): *Human-Computer Interaction, Part II, HCII 2015*, LNCS 9170, pp. 130–139. Springer International Publishing Switzerland (2015)
8. *Analytical dynamics*, From Wikipedia, the free encyclopedia, https://en.wikipedia.org/wiki/Analytical_dynamics
9. *Technisches Taschenbuch*, Schaeffler Technologies GmbH & Co. KG (2014)

Effects of Time Pressure and Task Difficulty on Visual Search

Xiaoli Fan, Qianxiang Zhou, Fang Xie and Zhongqi Liu

Abstract The process of pilot constantly checking the information given by instruments was examined in this study to detect the effects of time pressure and task difficulty on visual searching. A software was designed to simulate visual detection tasks, in which time pressure and task difficulty were adjusted. Two-factor analysis of variance, simple main effect, and regression analyses were conducted on the accuracy and reaction time obtained. Results showed that both time pressure and task difficulty significantly affected accuracy. Moreover, an interaction was apparent between the two factors. In addition, task difficulty had a significant effect on reaction time, which had a linearly increasing relationship with the number of stimuli. By contrast, the effect of time pressure on reaction time was not so apparent under high reaction accuracy of 90 % or above. In the ergonomic design of a human-machine interface, a good matching between time pressure and task difficulty is key to yield excellent searching performance.

Keywords Visual search · Time pressure · Task difficulty · Reaction time · Accuracy

1 Introduction

In the future, tactical aircrafts are predicted to operate in a significantly more demanding environment than they do today. With the advancements in aviation science and technology, tactical aircrafts have been enhanced with advanced features, particularly the display interface of aircraft cockpits [1]. An aircraft cockpit is a highly complex human-machine interaction system, and its display interface is

X. Fan (✉) · Q. Zhou · Z. Liu
School of Biology and Medical Engineering, Beihang University, Beijing, China
e-mail: fanfan19851414@163.com

F. Xie
China North Vehicles Researching Institution, China North Industries Group Corporation,
Beijing, China

one of the most essential devices for pilot–aircraft interaction. Given the increased requirements for flight altitude, speed, and endurance time, as well as the automation of display systems, all tasks related to flight are now being concentrated on only one or a few pilots to accomplish. Such responsibilities cause tremendous physiological and psychological pressure on pilots. Consequently, ergonomics problems have become critical factors that restrict flight security [2, 3]. Data analysis of reasons for space flight failures indicated that 70–80 % of such cases were ascribed to the poor design of human-machine interfaces [4]. Cognitive compatibility is an important aspect in the design of the aircraft display interfaces. In other words, the structure of a human-machine interface should match the cognitive structures of its operators [5].

Vision is the key cognitive sensory channel of humans to interact with the world around them. Approximately 80 % of the perceived information is obtained through a person’s vision [4]. Information detection tasks are an essential part of a human-machine system, and these activities invariably involve visual search [6]. In visual detection tasks during flight, the central nervous system and visual organ of pilots often suffer from strain as a result of focusing too much attention on the aircraft’s operation process, which consists of continuous information gathering, analysis, judgment, and issuing commands, among others, ultimately causing visual and central nervous system fatigue [7]. Therefore, the work efficiency and flight safety of pilots and aircrafts largely depend on the matching degree between the visual display interface and the visual cognition characteristics of humans [8, 9]. Thus, the visual cognition of human operators should be examined to develop an optimized pilot–aircraft display interface [10, 11].

Aside from monitoring environmental changes, pilots must be aware of the situation of the aircraft itself while conducting missions. Numerous system parameters, such as performance, input, and work condition parameters, among others, are obtained from the display instruments. Pilots must monitor the target instruments by visually searching the instrument clusters constantly, and on the basis of which, they must render judgment and perform appropriate operations. However, the number of cockpit instruments has significantly increased with the improvement of aircraft performance; thus, the task difficulty of visual detection had also increased accordingly. Although electro-optic display systems have been widely applied and have played a significant role in reducing the number of display instruments, the task difficulty remains unresolved [12]. As such, pilots tend to feel pressure to act quickly, as time is a critical factor in the visual search process during missions. The pressure brought about by the time limit and the task difficulty largely affects the physiology and psychology of pilots, which may lead to a degraded search performance [13, 14]. The same situation exists in many other human-machine systems.

Existing ergonomic studies on aircraft display interfaces have mostly focused on coding information, such as character, symbol shape, size, color, background, luminance, and contrast; these works have helped improve the efficiency of identifying and improving the interface layout by visual area zoning [15–17]. Other works have explored the mental workload and situation awareness of operators [18, 19].

However, limited studies are available on visual search characteristics. As displays have increased in variety and complexity, the study of target visual detection tasks is becoming an increasingly popular research topic. The more accurate is the reaction of people to the target information in the time available, the more efficient the search would be. Therefore, an effective and efficient visual detection task should be designed to understand the factors involved in the visual search process and their relationships to the search performance.

This study developed a software to simulate the target visual detection task display in human-machine systems. The search performance of the subjects was then examined by setting different time pressure and task difficulty levels. These processes would provide scientific bases for the ergonomic design of a human-machine display interface, particularly for aircrafts.

2 Material and Method

2.1 Subject

A total of 10 college students from the Beijing University of Aeronautics and Astronautics, comprising six males (60 %) and four females (40 %), voluntarily participated in this experiment. Their ages ranged from 20 to 28 years old. They all have normal or corrected to normal visual acuity. Most of the subjects often spent between two and eight hours a day using personal computers for various purposes. None of them have had any experience with this type of experiment. All of the subjects were right-handed and were capable of flexibly using a mouse.

2.2 Apparatus and Software

The experiment for the visual detection task was conducted using a custom-made software developed on Visual Studio 2010. The software was run on a desktop computer with a 15-inch resolution (1280 × 1024 pixels) and luminance (68 cd/m²) display monitor (Fig. 1). Figure 1 shows that the middle gray region on the software interface is for the visual search picture, and the symbols represent different virtual instruments in the human-machine system. Among these symbols , denotes the target stimuli, whereas , , , , ,  and  represent different distractive stimuli. The parameter values of display time and number of stimuli, which includes both target stimuli and distractive stimuli, are both adjustable. The display time refers to the presentation time of each search picture. The visual search picture is divided into $n \times n$ small squares, with each square containing an instrument symbol. Therefore, the number of stimuli can be changed by adjusting the n -value. In the experiment, the different display times represented

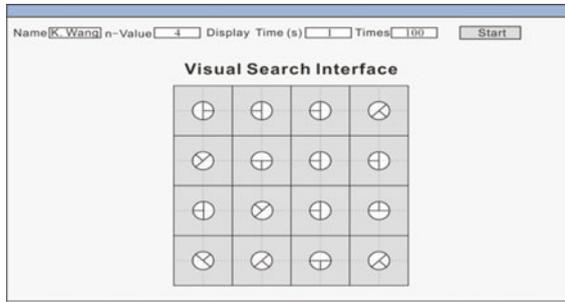


Fig. 1 Sample interface for visual detection task. Display time = 1 s. Number of stimuli = 16 (n-value = 4). Changing times of search picture = 100. All stimuli are refreshed and arranged randomly once the search picture changes. Each picture has at most only one target stimuli (appearance probability is 50 %), but two or more of the same distractive stimuli can appear simultaneously

different levels of time pressure, whereas the different numbers of stimuli represented different levels of task difficulty. The visual search picture changes once at a specified time interval. The symbols randomly appear in the different squares of the picture. Each search picture has two situations: “having a target” or “not having a target.” The probability of each situation is 50 %.

When the experiment begins and the search pictures appear, the subjects have to execute the following relevant actions:

- I. If the target is discovered, then the left button of the mouse should be pressed immediately.
- II. If the user is certain that there is no target by visual searching, then the right button of the mouse should be pressed immediately.
- III. If the user is unsure as to whether a target is present within the specified time, then no mouse button should be pressed.

After the experiment, the software automatically records the accuracy and reaction time.

2.3 Experiment Condition Control

In the visual field of anthropology, different space positions led to different search performances [20]. Therefore, to avoid any influence in the experiment, the visual search picture was placed in the optimum visual working area of the screen, and its size and position were kept constant. When different n-values were selected to divide the search picture, only the small square number was changed. In other words, the number and density of stimuli were changed, and the task difficulty changed with the change in n-value. The experiment was conducted under an

undisturbed environment with good illumination. To decrease the influence of fatigue in the experiment, the subjects can freely adjust their sitting posture, mouse placement, and screen angle.

3 Procedure

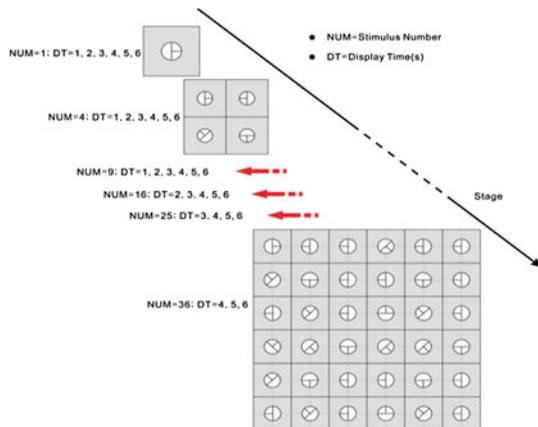
3.1 Pre-experiment

The pre-experiment adopted 6×6 a within-subjects design. To ensure experiment rationality and data volume, three subjects were randomly selected to participate in the pre-experiment. Different values of display time (1, 2, 3, 4, 5, 6/s) and number of stimuli (1, 4, 9, 16, 25, 36) were set. The valid range of display time and number of stimuli in this study were determined in consideration of the human factors, such as visual perception, mental workload, and so on. The actual experiment was divided into six stages because of its long duration (approximately 4 h in total); each stage consisted of several blocks that depend on the different values of display time and number of stimuli (Fig. 2). Each subject was given three days, with two stages per day, to complete the experiment. Such an arrangement avoids the effect of fatigue on search performance.

3.2 Actual Experiment

The subjects were required to familiarize themselves with the experimental process and train in advance. Before each block, the person conducting the experiment had to set the display time, n-value (number of stimuli = $n \times n$), and the number of

Fig. 2 Experimental design. The entire experiment is divided into six stages according to the different numbers of stimuli. Each stage consisted of several blocks that depend on the different display times. The entire experiment consists of 30 blocks, which require approximately 4 h to finish. All values of display time and the numbers of stimuli are based on the results of the pre-experiment



search pictures changed (each block in this study is set 100 times) in advance. To begin the block, the subjects had to stare at the center of the visual search region, where a red “+” symbol was displayed. They would then click “Start,” which was displayed on the screen, by pressing the mouse button. The stimuli were displayed on the center of the screen once the subject clicked “Start.” Memorizing the target symbol, the subjects were asked to scan the search picture by visual searching, and then react as quickly and as accurately as possible. After the first search picture disappeared, another one immediately appeared, with the stimuli being rearranged at random. To avoid lower motivation to engage in deep, deliberate, and systematic processing of information, the subjects under the time-pressure condition were told that they are on a tight schedule and reminded at intervals during the task to “hurry up” and “work harder”. When the block was finished, the accuracy and reaction time during this process were recorded by using a computer. The subjects can then make adjustments and rest.

3.3 Performance Measures

Search performance was investigated by assessing the accuracy and reaction time to complete a detection task. In this study, time pressure and task difficulty were considered the two main factors that affect the search performance. For statistical testing, two-factor ANOVA was performed using SPSS 19.0.

4 Results

4.1 Effects of Display Time and Number of Stimuli on Accuracy

The results shown in Table 1 were obtained by analyzing the data using two-factor ANOVA in which display time and number of stimuli are the two factors. This analysis indicates that the effects of both display time and number of stimuli on accuracy are significant (for display time: $F = 157.2$, $P < 0.05$; and for number of stimuli: $F = 140.166$, $P < 0.05$). This analysis also indicates that an interaction is apparent between display time and number of stimuli ($F = 18.069$, $P < 0.05$). Simple main effect analysis was then conducted by programming based on within-subject factor experiment. Table 2 shows that display time significantly differed at different levels of number of stimuli (all $P < 0.05$), whereas number of stimuli significantly differed at different levels of display time (all $P < 0.05$).

Figure 3 was constructed using number of stimuli as abscissa and accuracy as ordinate. For each given display time, the search accuracy shows a downtrend with the number of stimuli. However, the decrease in the amplitude of the curves differs

Table 1 Effect analysis of accuracy at different display times and number of stimuli

Source	Sum of squares	df	F	P
Display time	65414.3	5	157.2	0.0
Number of stimuli	58309.5	5	140.2	0.0
Display time * number of stimuli	28563.1	19	18.1	0.0

Table 2 Simple effect analysis between different display times and number of stimuli

Source	Sum of squares	df	F	P
Display time VS.1	729.93	5	12.0	0.0
Display time VS.4	29483.73	5	189.3	0.0
Display time VS.9	39520.00	5	199.9	0.0
Display time VS.16	59497.33	5	212.9	0.0
Display time VS.25	57287.53	5	210.5	0.0
Display time VS.36	54765.40	5	110.6	0.0
Number of stimuli VS.1s	59584.80	5	266.1	0.0
Number of stimuli VS.2s	73105.13	5	522.4	0.0
Number of stimuli VS.3s	416126.93	5	272.0	0.0
Number of stimuli VS.4s	16900.33	5	47.9	0.0
Number of stimuli VS.5s	9238.73	5	24.8	0.0
Number of stimuli VS.6s	3478.40	5	12.2	0.0

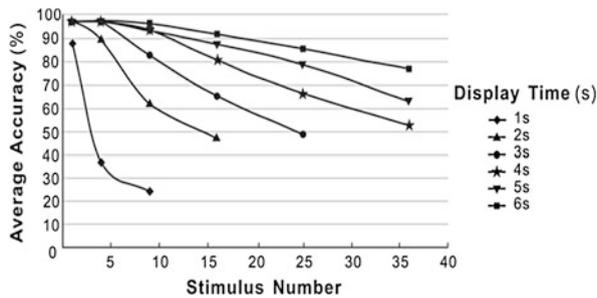


Fig. 3 Number of stimuli/accuracies at different display times. The general trend of accuracy decreases with the increase in the number of stimuli in visual searching. However, the decrease in the amplitude of the curves differs for different display times. A longer display time corresponds to a smaller effect of the number of stimuli on visual search accuracy

for different display times. By using nonlinear regression analysis, the functions of curve fitting were derived with high fitting coefficients, as shown in Table 3. All curves are expressed in quadratic equations. To a certain extent, there are $|y6'| < |y5'| < |y4'| < |y3'| < |y2'| < |y1'|$, which indicates that a longer display time corresponds to a smaller effect of the number of stimuli on visual search accuracy.

Table 3 Regression functions of reaction accuracy at different display times

Display time(s)	Optimal approximation function	R2
1	$y_1 = 108 - 25.62x + 1.81x^2$	1.00
2	$y_2 = 105.01 - 5.53x + 0.113x^2$	0.98
3	$y_3 = 102.3 - 2.31x + 0.003x^2$	0.998
4	$y_4 = 100.29 - 1.08x - 0.008x^2$	0.983
5	$y_5 = 98.28 - 0.37x - 0.017x^2$	0.996
6	$y_6 = 97.69 - 0.18x - 0.012x^2$	0.992

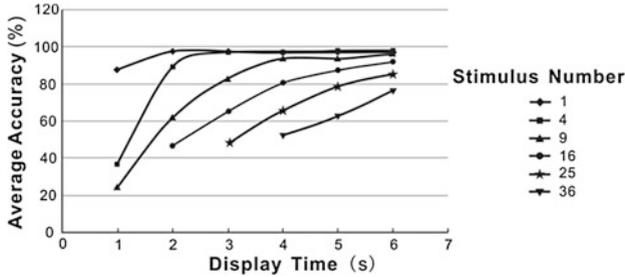


Fig. 4 Display time/accuracies at different number of stimuli. The curves show that accuracy tends to initially increase and then tends to stabilize. However, several small decreases in accuracy can be observed in the stable section (e.g., stimuli number = 1, display time = 3 s; number of stimuli = 9, display time = 5 s)

Moreover, this result also verifies the interaction between display time and number of stimuli.

Figure 4 shows the display time/accuracy curves for different numbers of stimuli. The curves show that accuracy tends to initially increase and then tends to stabilize. Several small decreases in values can be observed in the middle section. For example, in the single stimuli discrimination task with only one stimuli symbol appearing each time, the search accuracy is significantly low when the display time is 1 s. The reason is that the display time is too short, thereby causing a high level of time pressure. In addition, errors occur in the process because the subjects cannot follow the pace of the search picture switching in terms of using the eyes and implementing the operations. Accuracy is stable at approximately 98 % when the display time reaches 2 s or longer. Moreover, accuracy reaches the maximum when the display time is 3 s but is slightly lower when the display time continues to increase. The reason is that the presentation time of each search picture is too long; thus, the subjects spread their attention easily during the blocks. The same rules are applied in the multi-stimuli discrimination task. Table 4 shows the data of the maximum number of stimuli and optimum number of stimuli that can be discriminated at different display times under accuracy of 90 % or above. Figure 5 shows the comparisons between the display time and the maximum number of stimuli with a linear regression line, a correlation coefficient ($R^2 = 0.9365$), and a

Table 4 Relative number of stimuli threshold at different display times

Display time(s)	Maximum number of stimuli	Optimum number of stimuli
2	1	1
3	4	1
4	9	4
5	9	4
6	16	4

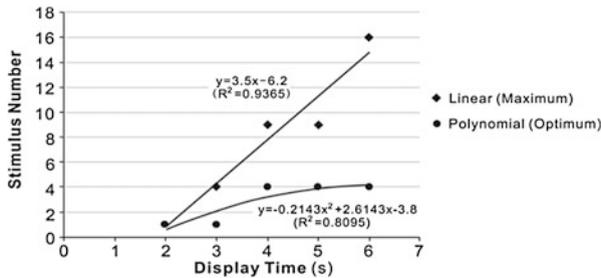


Fig. 5 Regression relationship between display time and maximum/optimum number of stimuli. Comparison between the display time and the maximum number of stimuli with a linear regression line ($R^2 = 0.9365$), and comparison between the display time and the optimum number of stimuli with a polynomial fitting curve ($R^2 = 0.8095$)

linear fitting equation ($y = -6.2 + 3.5x$). If $x = m$, the number of stimuli N only satisfies the condition $N \leq 3.5m - 6.2$, and a good search performance can be obtained. The polynomial fitting curve between the display time and the optimum number of stimuli is also presented in Fig. 5, with a correlation coefficient ($R^2 = 0.8095$), which suggests that the optimum number of stimuli is not directly proportional to the display time. Therefore, in the design of a human-machine interface, the best matching of display time and number of stimuli is necessary.

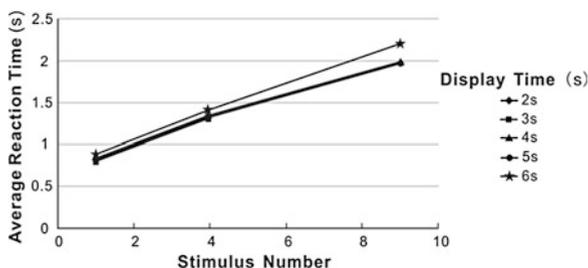
4.2 *Effects of Display Time and Number of Stimuli on Reaction Time*

Inherent connection and trade-off exist between reaction speed and accuracy [21]. The study of reaction speed is meaningless when reaction accuracy is too low. If the matching between the display time and the number of stimuli exceeds the normal scope of human visual cognition, then it will lead to low accuracy. Therefore, the influence of display time and number of stimuli on reaction time is analyzed in this study under accuracy of 90 % or above. Table 1 shows the related reaction time data. Two-factor ANOVA was conducted, and the results are shown in Table 5.

Table 5 Effect analysis of reaction time at different display times and number of stimuli

Source	Sum of squares	df	F	P
Display time(s)	0.313	4	1.4	0.226
Number of stimuli(s)	23.171	2	213.1	0.000
Display time * number of stimuli	0.118	5	0.4	0.823

Fig. 6 Number of stimuli/reaction time at different display times. Reaction time increases linearly with the increase in the number of stimuli



The main effect of display time on reaction time is insignificant ($F = 1.438$, $P > 0.05$), whereas the main effect of number of stimuli is significant ($F = 213.132$, $P < 0.05$). The interaction between display time and number of stimuli is insignificant ($F = 0.435$, $P > 0.05$).

With the use of the reaction time data in Table 1, the curves with display time and number of stimuli were designated as abscissa and accuracy as ordinate (Figs. 6 and 7). Figure 6 shows that reaction time increases linearly with the increase in number of stimuli. By conducting linear regression analysis, functions of curve fitting were derived with high fitting coefficients, as shown in Table 6. The four functions were considered comprehensively, with each additional stimuli extending the reaction time by 154 ms. Figure 7 shows that the mean reaction time remains generally stable despite the increase in display time in the visual search task with the same number of stimuli. However, the reaction time of several subjects was found to slightly increase with the increase in display time. The reason is that the

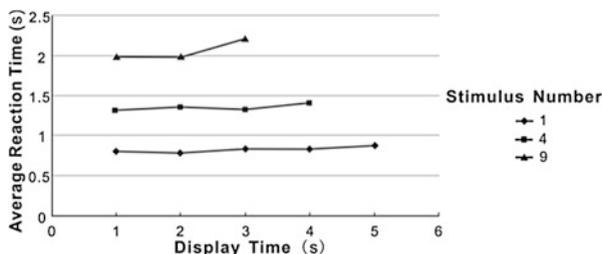


Fig. 7 Display times/reaction times at different number of stimuli. Mean reaction time remains generally stable despite the increase in display time in the visual search task with the same number of stimuli

Table 6 Regression function of reaction accuracy at different display times

Display time (s)	Optimal approximation function	R ²
3	$y = 0.621 + 0.171x$	1.000
4	$y = 0.73 + 0.143x$	0.995
5	$y = 0.721 + 0.141x$	0.996
6	$y = 0.732 + 0.164x$	0.999

display time was too long for the subjects, causing lower time pressure. The inattention of the subjects leads to longer reaction time.

5 Discussion

In this study, different display times represent different levels of time pressure, whereas different numbers of stimuli represent different task difficulties. The experimental results revealed that the effects of display time and number of stimuli on accuracy were both significant at the 0.05 level. Moreover, their interaction was also significant; that is, the two factors may work together on visual search accuracy. Simple main effect analysis revealed that the display time significantly differed at different levels of the number of stimuli; and that number of stimuli significantly differed at different levels of display time. The highest reaction accuracy in the search task would be obtained when the two factors reach the best matching.

Under the given time pressure, accuracy decreases as a result of the increase in the number of stimuli. The gradual development process can be described, as follows: (1) judge correctly and operate correctly; (2) judge correctly but operate wrongly or have no time to operate; (3) judge wrongly and the operation falls behind; (4) brain response cannot keep up with the changing search pictures and has no time to judge. For the given number of stimuli, accuracy increases as the time pressure decreases, and then tends to stabilize after reaching a certain point. Before that point, wrong judgments and operations easily resulted because of limitations in cognizing capacity, physiology, and emotion, among others, of the advanced nerve center caused by higher pressure [22]. When such point is reached, time pressure had no significant effect on accuracy. However, a slight rebound phenomenon was noted among several subjects. This phenomenon was due to their excessive relaxation, which caused wrong judgments and operations [23].

The current study analyzed the effects of display time and number of stimuli on reaction time under high accuracy because inherent connection and trade-off exist between reaction speed and accuracy. The results indicated that the main effect of display time on reaction time was insignificant, whereas the effect of the number of stimuli was significant. When the time pressure was given, the reaction time was linear with number of stimuli; this finding is inconsistent with the theory of log proposed by Merkel (1885) [24]. The two reasons for such difference are as follows:

(1) the stimuli material was different, which led to a different task difficulty; (2) the two situations of “having target” and “not having target” were considered one case without distinction. However, such consideration would not affect the visual search law in the current study.

6 Conclusion

In this study, the effects of time pressure and task difficulty on visual search were studied during a simulated visual detection task. As expected, task difficulty demonstrated a significant effect not only on search accuracy but also on reaction time. Time pressure exhibited a significant effect only on search accuracy. Moreover, their integrated effect on search accuracy is very significant. Therefore, the two factors should be comprehensively considered when designing a human-machine interface, particularly the dashboard design. Instrument number and cognitive difficulty should be decided reasonably based on the order of importance and urgency of instrument monitoring tasks and cognitive characteristics of humans. For instruments that need an immediate reaction, the dashboard should be designed with little search difficulty by reducing the number of instruments or other measures. Time pressure is relative with task difficulty; thus, the best matching of time pressure and task difficulty can produce good search performance.

Although several valuable findings involving visual search laws were obtained, several limitations were also noted in this study. First, the number of participants in the experiment was small because of the long course. Second, except for time pressure and task difficulty, other factors, such as target location, were not considered in the experiment, which also has effects on search performance. Third, the task difficulty was not only related to number of stimuli but also related to the shape and size characteristics of the stimuli [25]. Finally, in the actual human-machine systems, there are always more than one target in the visual detection tasks. In future studies on visual search laws, all of these factors should be considered to guide the ergonomic design of human-machine interfaces.

Acknowledgments This research was funded by the National Natural Science Fund (31170895), the National Defense Pre-research Fund (A0920132003), and the Human Factors Engineering Key Laboratory Fund Project (HF2013-K-06).

References

1. Haas, M.W.: Virtually-augmented interfaces for tactical aircraft. *Biol. Psychol.* **40**, 229–238 (1995)
2. Cao, P., Guo, H.: An ergonomical study on motionless and motional presentation of Chinese character on CRT display. *Space Med. Med. Eng.* **7**(1), 35–40 (1994)

3. Cao, L., Li, Y.: Sampling order for irregular geometric figure recognition. *Acta Psychol. Sinica* **35**(1), 44–49 (2002)
4. Jun, W., Wei, C.: Ergonomics evaluation of human-machine interface based on spatial vision characteristics. *Tactical Missile Technol.* **6**, 107–111 (2012)
5. Fuchs-Frohnhofen, P., Hartmann, E.A., Brandt, D., Weydandt, D.: Design human-machine interfaces to the user's mental models. *Control Eng. Practice* **4**(1), 13–18 (1996)
6. Pearson, R., van Schaik, P.: The effect of spatial layout of and link color in web pages on performance in a visual search task and an interactive search task. *Int. J. Hum. Comput. Stud.* **59**, 327–353 (2003)
7. Einke, D., Humphreys, G.W., Tweed, C.L.: Top-down guidance of visual search: a computational account. *Vis. Cogn.* **14**, 985–1005 (2006)
8. Zhou, Q., Cai, K., Li, J.: *Man-Machine Interface Design of Manned Spacecraft*. National Defence Industry Press, Beijing (2009)
9. Zhou, Q.: Research on human-centered design of man-machine interface in manned spacecraft. *Aerosp. Shanghai* **3**, 39–43 (2002)
10. Qu, Z., Zhou, Q.: Effects of digital information display time on human and machines monitoring performance. *Space Med. Med. Eng.* **6**(18), 191–195 (2005)
11. Bodrogi, P.: Chromaticity contrast in visual search on the multi-color user interface. *Displays* **24**, 39–48 (2003)
12. Wei, H., Zhuang, D., Wanyan, X.: An experimental analysis of situation awareness for cockpit display interface evaluation based on flight simulation. *Chin. J. Aeronaut.* **26**(4), 884–889 (2013)
13. Guo, X., Yu, R.: *The Effect of Speed, Time Stress and Number of Targets Upon Dynamic Visual Search*, Tsinghua University (2011)
14. Wang, Z., Zhang, K., Klein, R.M.: Inhibition of return in static but not necessarily in dynamic search. *Atten. Percept. Psychophys.* **72**, 76–85 (2010)
15. TS, Tullis: An evaluation of alphanumeric, graphic, and color information displays. *Hum. Factors* **23**(5), 541–550 (1981)
16. Spenkelink, G.P.J., Besuijen, J.: Chromaticity contrast, luminance contrast, and legibility of text. *J. SID* **4**(3), 135–144 (1996)
17. Knoblauch, K., Arditi, A., Szlyk, J.: Effect of chromatic and luminance contrast on reading. *J. Opt. Soc. Am. A* **8**(2), 428–439 (1991)
18. Perry, C.M., Sheik-Nainar, M.A., Segall, N., Ma, R., Kaber, D.B.: Effects of physical workload on cognitive task performance and situation Awareness. *Theor. Issues Ergon. Sci.* **9**(2), 95–113 (2008)
19. Wilson, G.F.: An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *Int. J. Aviat. Psychol.* **12**(1), 3–18 (2009)
20. Zhang, L., Zhuang, D.: Text and position coding of human-machine display interface. *J. Beijing Univ. Aeronaut. Astronaut.* **37**(2), 185–188 (2011)
21. Kail, R.: Sources of age differences in speed of processing. *Child Dev.* **57**, 969–987 (1986)
22. Rayner, K., Li, X.S., Williams, C.C., Cave, K.R., Well, A.D.: Eye movements during information processing tasks: individual differences and cultural effects. *Vision. Res.* **47**, 2714–2726 (2007)
23. Boot, W.R., Becic, E., Kramer, A.F.: Stable individual differences in search strategy: the effect of task demands and motivational factors on scanning strategy in visual search. *J. Vis.* **9**(3), 1–16 (2009)
24. Guo, X., Yang, Z.: *Basis of Experimental Psychology*, p. 152. Higher Education Press, Beijing (2011)
25. McIntire, J.P., Havig, P.R., Watamaniuk, S.N.J., Gilkey, R.H.: Visual search performance with 3-D auditory cues: effects of motion, target location and practice. *Hum. Factors* **52**(1), 41–53 (2010)

Part III
Human-Agent Teaming

Operator-Autonomy Teaming Interfaces to Support Multi-Unmanned Vehicle Missions

Gloria L. Calhoun, Heath A. Ruff, Kyle J. Behymer
and Elizabeth M. Mersch

Abstract Advances in automation technology are leading to the development of operational concepts in which a single operator teams with multiple autonomous vehicles. This requires the design and evaluation of interfaces that support operator-autonomy collaborations. This paper describes interfaces designed to support a base defense mission performed by a human operator and heterogeneous unmanned vehicles. Flexible operator-autonomy teamwork is facilitated with interfaces that highlight the tradeoffs of autonomy-generated plans, support allocation of assets to tasks, and communicate mission progress. The interfaces include glyphs and video gaming type icons that present information in a concise, integrated manner and multi-modal controls that augment an adaptable architecture to enable seamless transition across control levels, from manual to fully autonomous. Examples of prototype displays and controls are provided, as well as usability data collected from multi-task simulation evaluations.

Keywords Flexible automation · Adaptable automation · Unmanned systems · Autonomous vehicles · Operator-Autonomy · Interfaces · Multi-modal control · Display symbology · Human factors

G.L. Calhoun (✉)

Air Force Research Laboratory, 711 HPW/RHCI, Dayton, OH, USA

e-mail: gloria.calhoun@us.af.mil

H.A. Ruff · K.J. Behymer

Infoscitex, Dayton, OH, USA

e-mail: heath.ruff.ctr@us.af.mil

K.J. Behymer

e-mail: kyle.behymer.1.ctr@us.af.mil

E.M. Mersch

Wright State Research Institute, Dayton, OH, USA

e-mail: elizabeth.mersch.ctr@us.af.mil

© Springer International Publishing Switzerland 2017

P. Savage-Knepshield and J. Chen (eds.), *Advances in Human Factors in Robots and Unmanned Systems*, Advances in Intelligent Systems and Computing 499,
DOI 10.1007/978-3-319-41959-6_10

1 Introduction

Agility in tactical decision-making, mission management, and control is the key attribute for enabling human and heterogeneous unmanned vehicle (UV) teams to successfully manage the “fog of war” with its inherent complex, ambiguous, and time-pressured conditions. In support of an Assistant Secretary of Defense for Research and Engineering (ASD(R and E)) Autonomy Research Project Initiative (ARPI), a tri-service team led by the Air Force Research Laboratory (AFRL) is developing and evaluating an “Intelligent Multi-UV Planner with Adaptive Collaborative/Control Technologies” (IMPACT) system. Additionally, AFRL leads development of a new interface paradigm by which the operator teams with autonomous technologies. This effort involves designing intuitive human-autonomy interfaces which will enable: (a) operators to monitor and instruct autonomy in response to dynamic environments and missions and (b) the autonomy to make suggestions to the operator and provides rationale for generated plans. Facilitating operator-autonomy interaction is a key challenge for achieving trusted, bi-directional collaboration.

To guide the design of human-autonomy interfaces for multi-UV missions, defense mission scenarios for a military base were generated based on inputs from a cognitive task analysis (CTA) [1]. One UV scenario involved a single operator managing 4 air UAVs, 4 ground UGVs, and 4 surface USVs (watercraft) from a ground control station. (Each vehicle type in the simulation was programmed to emulate an existing vehicle (e.g., UAV: Boeing’s ScanEagle)). The operator’s responsibility was to work with the autonomy such that the UVs monitored the military base as well as responded to any potential threats. For a normal defense posture, each UV patrolled a portion of the base with the autonomy determining the specific routes, areas, and relevant parameters (e.g., UV speed). When a potential threat was detected, the operator changed the UVs’ patrol state to highly mobile, resulting in the autonomy modifying the routes to intensify defensive tactics (e.g., schedule UVs to patrol the base’s border in order to maintain the integrity of its perimeter). Threat detections could also prompt specific requests from a commander, communicated through chat to the operator, to position a sensor on a specific location (“Get eyes on the ammo dump”). This entailed re-tasking one or more UVs from the ongoing patrol (with the autonomy reallocating the remaining UVs on patrol segments for continued base defense).

It is important to note that the operator’s responsibility in these scenarios was to manage the movement and tasking of the UVs. While the operator was presented with simulated imagery from each UV’s sensor (as well as symbology of each sensor’s field-of-view on the map), the tasks to monitor and interpret the sensor imagery were assigned to a (simulated) remotely located sensor operator. Information about the sensor imagery assessment (from the sensor operator and/or commander) was communicated to the operator via chat and/or radio.

This paper provides an overview of the operator-autonomy interfaces developed to support the above mission scenario. These interfaces were implemented in a

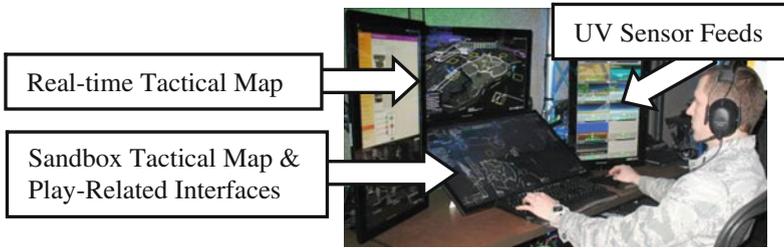


Fig. 1 Illustration of IMPACT simulation for single operator and multi-UV teaming

ground control station [2] that consists of four monitors (three: $26 \times 18 \frac{1}{2}$ in and one touch-sensitive: $29 \frac{17}{32} \times 17 \frac{3}{8}$ in), keyboard, mouse, headset, and footswitch (Fig. 1). To date, the research has focused on the center monitors (top: real time tactical map; bottom: sandbox tactical map) and right monitor (each UV's sensor imagery and state data). This IMPACT simulation is an instantiation of a Fusion framework [2] that employs a Sphinx speech recognition system, cooperative planners, intelligent agents, and autonomics monitoring services [1]. Before describing individual controls and displays, an explanation of the overarching design approach will be provided.

2 Flexible Goal-Based Control Architecture

Operations with multiple highly automated vehicles raise many questions concerning the balance of systems' autonomy with human interaction for optimal situation awareness and mission effectiveness [3]. Moreover, the application of autonomy can have negative impacts besides reduced operator situation awareness: for instance, inappropriate trust, unbalanced workload, decision biases, vigilance decrements, and complacency [4]. Thus, a control architecture that supports the flexible allocation of tasks and employs interfaces that promote operator-autonomy shared situation awareness is required. In this manner, the operator is more involved in explicitly invoking automation for certain tasks, UVs, mission events, or a unique combination of these dimensions. This extension of a delegation approach to provide finer-grain control lessens the unpredictability of changes in autonomy level and presumably the associated negative effects (e.g., "automation surprises") [5].

With flexible control architectures, the number of automated tasks, as well as their degree of automation varies at any given time to optimize the tradeoff between operator involvement and workload. This entails the instantiation of a wide spectrum of control methods, whereby some UVs are performing tasks using higher level automated control modes leaving the operator more time to dedicate attention to a particular vehicle that requires more precise and timely control inputs (e.g.,

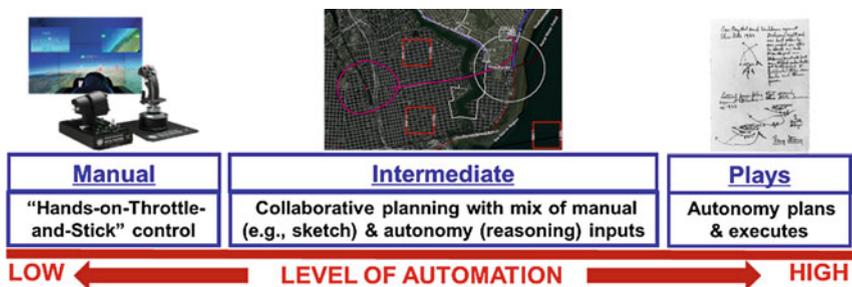


Fig. 2 Illustration of flexible control architecture for tasks and UVs with interface technologies supporting transition between any control level [9]

temporary manual control) [6]. Intermediate levels [7] that involve a mix of manual and autonomy inputs are also useful (Fig. 2). The interfaces must support seamless transitions between more detailed manual inputs for precisely commanding a specific UV’s movement to concise commands that initiate high level “plays.”

This paper will focus on our interface designs that support the more autonomous level involving plays. UV-associated plays are similar to a sports team’s plays such that quick commands result in rapid and efficient delegation as well as the capability to modify plays (i.e., sports “audibles”) as needed [7, 8]. In our instantiation of this adaptable automation approach, only a few commands are required for the majority of likely base-defense plays (i.e., tasks). An efficient goal-based approach is also accommodated whereby the operator simply has to specify key parameters: “what” (inspect with sensor) and “where” (location). All the other parameters can be determined by the autonomous team partner, for example “who” (which UV(s)) and “how” (route, speed, etc.). However, to better match a play to the specific requirements of a dynamically changing environment, the interfaces also support efficient inputs by which the operator can specify any or all play-related details. Moreover, a multi-modal control approach is supported such that the commands can be issued with push-to-talk speech commands and/or manual (touch or mouse) inputs.

3 Design of UV Play Interfaces

This section provides an overview of interfaces designed to support the delegation control aspect of our architecture, namely, the plays performed by highly autonomous UVs. This includes the interfaces by which the operator calls plays, specifies play details, views tradeoffs of autonomy-generated plans for plays, and monitors plays. Figure 3 illustrates the UV/play icon symbology employed across interfaces.

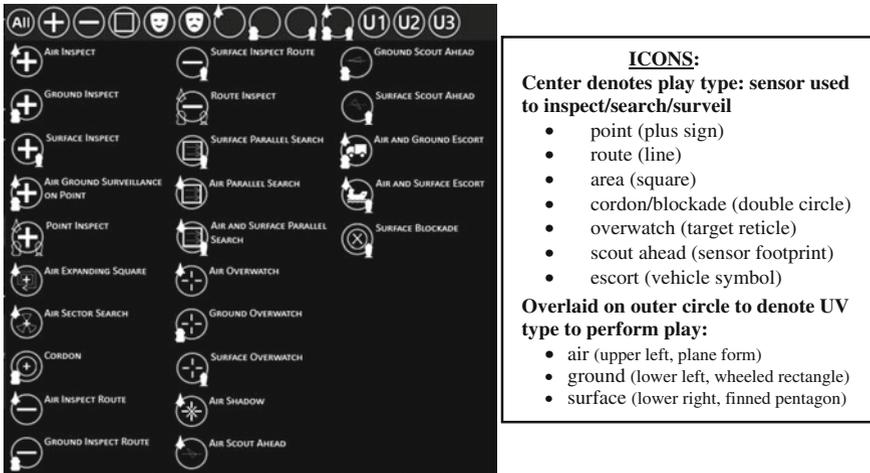


Fig. 3 Icons that specify plays and UV types

3.1 Sandbox Tactical Map Overview

The play-related interfaces overlay the map on the bottom monitor (Fig. 1). Each can be repositioned with the mouse to avoid obscuring map information. Figure 4 shows a normal full coverage patrol (NFCP) in effect, with the UVs monitoring the base and specific named areas of interest (NAIs). The NFCP is indicated by the white route lines and gray UV icons (air: diamond, ground: wheeled rectangle, and surface: finned pentagon) within the (yellow dashed) base’s area of responsibility. An exception is a UAV that has been pulled off NFCP to inspect Point Bravo (e.g., in response to the report of a suspicious watercraft). Plays are color-coded to be distinguished from patrol states (here, orange is used for the UV symbol and its route/loiter). To help illustrate specific play-related interfaces in the following sections, assume that the commander has concluded that the UAV’s imagery is inadequate and that the potential threat also needs to be viewed by a USV. Thus, the operator’s task is to call a play that positions a surface vehicle’s sensor on Point Bravo.

3.2 Calling Plays

There are three interfaces that support calling a play (versus manual control) that can position a USV to inspect Point Bravo. With one, the operator issues a speech command “Surface Inspect Point Bravo” and consent. In another, a Play radial menu appears when either a location (Fig. 5a) or UV symbol (Fig. 5b) is clicked or touched. This interface filters plays such that only the options relevant to that



Fig. 4 Overview of sandbox tactical map (*bottom*) shows eleven UVs on a normal full coverage patrol and one UAV inspecting a potential water threat

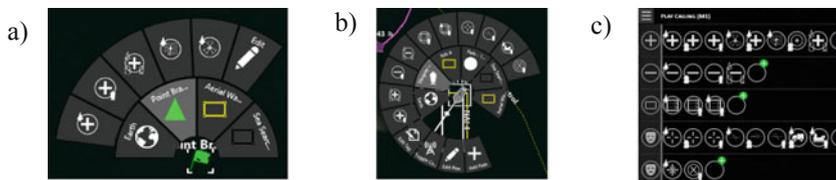


Fig. 5 Interfaces to specify play and UV type with mouse or touch

location or UV are presented (e.g., menu does not include ground plays for Point Bravo as it is located in the water). In a third Play Calling interface, all the plays listed in Fig. 3 are represented and the operator clicks or touches the desired play icon (Fig. 5c; upper left of Fig. 4). Icons can be arranged by either vehicle type or play type and in Fig. 5c are arranged by play type (point, route, area, and then ‘target plays’ (non-threat versus potential threats)) [10]. The icons that include outlined UV symbols provide the operator with the option of calling a play type and having the autonomy suggest which one or more UV types should perform it. The play symbols with no UV symbols enable the operator to edit which icons are presented in each row to customize the interface.

3.3 Reviewing Play Plans and Changing Play Parameters

Designation of a play type/location with speech or manual inputs results in changes on the sandbox to reflect the autonomy's generated proposed plan for the play. This includes dashed symbology on the map to indicate the proposed USV and route, as well as the addition of a two-page Play Workbook interface (Fig. 6). The information on the Play Workbook's left page reflects the play called and location specified. This information, together with the map symbology, depicts the proposed plan (UV and route) to accomplish the play. Note: the color of this new play (in this case blue) differs from the color of the existing UAV play and patrol. If the operator agrees with the plan, a consent input (speech command or checkmark selection) initiates the play and the dashed symbology becomes solid on the sandbox map and also appears on the real-time map. Thus, only three commands (what play, where, and consent) are required to initiate a play. Buttons on the left workbook page (or associated speech commands) also provide the operator other functionality (e.g., to call up more information on why the autonomy chose that plan). Moreover, the highlighted (selectable) icons on the right Workbook page indicate the factors that the autonomy prioritized in determining its recommended plan (e.g., "stopwatch" icon indicates that the plan was optimized on time). Each of these factors was determined via a CTA to provide constraints that the autonomy considers when generating a plan for a play. By selecting different icons, the operator can prompt the autonomy to re-plan. For instance, if the operator knows from the sensor operator that there is cloud cover, the operator can change that environment factor by selecting the "cloud" icon (or via speech), and the autonomy will generate a new play plan which has a more appropriate sensor. Tabs at the bottom of Page 2 call up

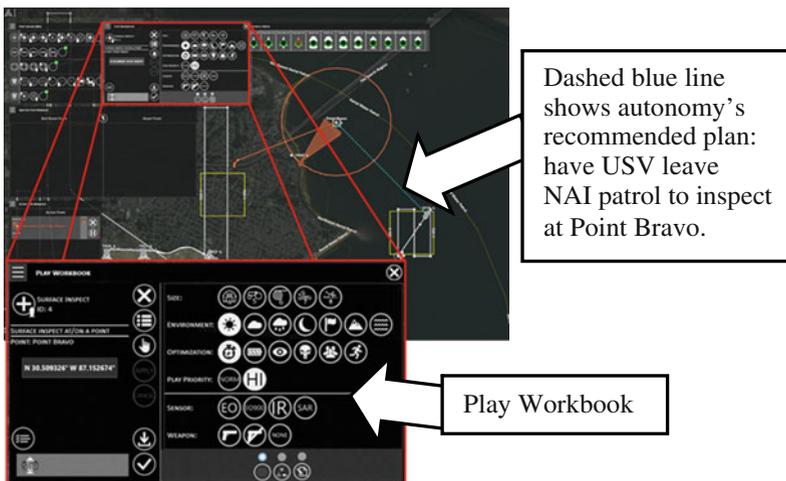


Fig. 6 Sandbox map showing autonomy's proposed UV and route to add a USV to inspect a suspicious watercraft at Point Bravo and close up of Play Workbook

other pages for the operator to specify which UV should perform the play or other play details (e.g., size of loiter).

3.4 Understanding Autonomy's Rationale for Proposed Play Plan

Besides noting the factors in the Play Workbook that are used by the autonomy to generate a proposed plan for a play, two other interfaces illustrated in Fig. 7 provide insight into the autonomy's reasoning. Figure 7a presents text that appears adjacent to the Workbook to provide further rationale, such as that the plan represents the best option for optimizing on time and how communication range was taken into account in UV choice. A plan comparison plot (Fig. 7b) illustrates the tradeoffs across play plans according to mission factors (see [11] for more detail).

3.5 Monitoring Play Execution

Once the operator makes a consent response to accept the autonomy's proposed plan for the play, the dashed route lines (Fig. 6) on the sandbox map turn solid (Fig. 8). The operator can monitor the play's progress by observing the movement of the UV symbols along routes. Additionally, other interfaces help monitor play progress.

The Active Play Table on the left (Fig. 9) lists all active plays (play type and UV) as well as the current patrol state (NFCP or highly mobile) at the bottom. Note: each play's symbol and text in the table is a distinct color that matches the play's route and UV symbol on each map. (For each multi-UV play, all associated UVs, routes, and tabular information would be the same color.) Rows in the Active Play map interface are also selectable, resulting in the Workbook displayed on the sandbox map to provide more information and serve as a guide for revising the play during

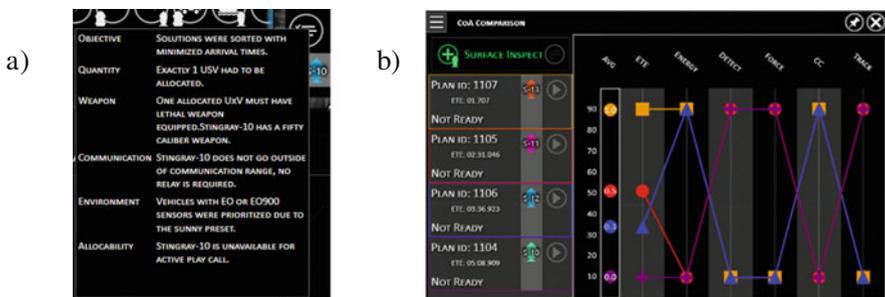


Fig. 7 Interfaces for operator to view autonomy's rationale for proposed play plan(s)



Fig. 8 Sandbox map showing both air and surface UVs on plays to inspect suspicious watercraft at Point Bravo and other UVs on normal full coverage patrol



Fig. 9 Close up of Active Play Table (left) and Play Quality Matrix (right)

its execution, if needed, with either manual Workbook selections or speech commands. More information on the patrol state is visible with the scrolling function highlighted in blue at the bottom. Functionality to cancel or pause the play is available to the right of the UV symbols.

The color coded vertical bar for each play (green, yellow, or red for normal, caution, or warning states) matches the most severe state indicated in the Play Quality matrix page available through the Workbook (on the right). Here the status of the play's progress is presented, as measured by autonomics monitoring services.

This example plots three parameters (estimated time enroute, fuel available, and UV speed). A normal state (green) is shown in the left two columns and deviations from that (yellow or red) and the respective direction of the more saturated color cell (far right column) indicates whether the parameter is above or below its expected operating range. (See [12] for more detail.)

3.6 Managing Multiple Plays

Considering 12 UVs are involved in the example scenario, it is likely that the operator will call many plays when there are multiple potential threats to the base. Moreover, for several base defense tactics, multiple UVs team on the same plays, such as a UAV scanning ahead to scout the route that a ground UV will take, or a UAV serving as a link for a surface UV that has a limited communication range. Multiple UVs are also involved if it is necessary to block an intersection of a road. The intelligent agents that reason among domain knowledge sources track each UV in order to determine the optimal UV assignment when generating proposed play plans. The intelligent agents' reasoning takes into account multiple factors including UV type, sensor capability, location, and whether or not the UV is already performing a play [11].

Given the complexity of multiple UVs and plays, an additional interface was implemented to supplement the Active Play interface. Figure 10 shows the Inactive Play interface that contains two tables. The "Not Ready" table contains plays that are called by the operator, but cannot be activated yet because they are waiting on a certain resource (e.g., specific sensor type or UV type). This interface also serves as a record of plays that the operator has called with the intent of activating them later in the mission. Once the needed resource is available for the play in the Not Ready Table, that play automatically moves to the "Ready" table. Here the play resides

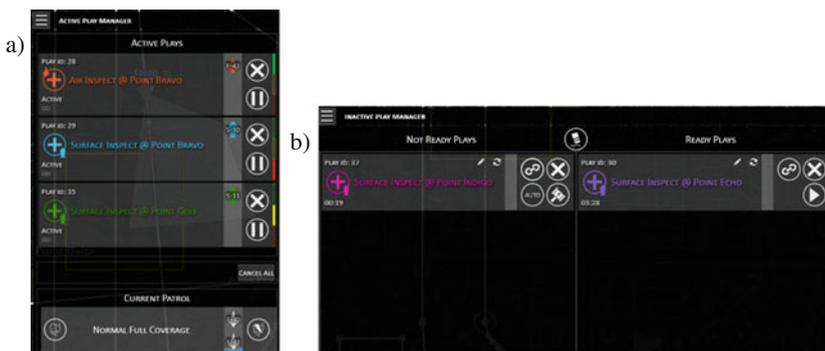


Fig. 10 Close up of Active Play Table (*left*) and Inactive (Not Ready and Ready) Play Tables (*right*)

until the operator either cancels the play or consents for it to begin. The Ready Table also includes any active plays that were paused by the operator. Similarly, plays in the Ready Table, whose UV asset is reassigned to a different play, would move to the Not Ready Table.

These management interfaces (Active, Not Ready and Ready Play Tables) provide the operator with buttons (to the right) to quickly pause, initiate and cancel plays, separate from their respective Play Workbooks. Selection of a row in the Inactive Play Table also calls up candidate plans for that inactive play. Moreover, there are methods by which the operator can chain plays, such that a play will become active, once another play has been completed. For certain plays deemed critical in the Not Ready Table that are waiting on a resource, the operator can designate it to automatically become active when an asset becomes available, without it first moving to the Ready Table to await operator input. This functionality was provided to alleviate the operator from constantly monitoring the Inactive Plays to initiate a critical play. These play management interfaces, with modification, can also support the presentation of autonomy suggested plays.

4 Evaluation of UV Play Interfaces

The interfaces described above are part of an ongoing evaluation. A summary of key results will be available for the presentation. For this evaluation, personnel familiar with base defense and/or UV operations are serving as experimental participants, performing the mission scenario described herein. Both quantitative and qualitative data are being collected for two (within-subjects) conditions: one that employs these operator-autonomy interfaces (and supporting autonomy components), as well as a condition where multi-UV navigation, mission planning, and loiter management are accomplished with only defined keyboard and mouse inputs (similar to what is described in [13]). The baseline condition has less autonomy support and does not employ the flexible delegation control architecture, including the interfaces illustrated in Figs. 5–10.

Preliminary data from the current evaluation suggests that the results will be aligned with an earlier evaluation that employed a less complex scenario (only 13 plays and 6 UVs) [1]. Moreover, the previous evaluation did not have the interfaces illustrated in Figs. 5a, b, and 7. The other interfaces were less refined too. Nevertheless, the first version of the IMPACT simulation was well received by the six participants who completed training, a 20-minute experimental trial, and extensive questionnaires in 3.25 h sessions. On a 5-point Likert scale (ranging from No Aid to Great Aid), five of the six participants strongly agreed that the approach for operator-autonomy teaming for UV defense missions has potential value and helps aid workload. Four participants rated it an aid to situation awareness as well. (All ratings were “4” or “5.”).

For this earlier study, the results were similar when the rating scale was applied to specific elements of the simulation: play calling, autonomy, feedback, and

control station. Ratings on the usability of IMPACT's approach to multi-UV operations were also positive. Ratings averaged 4.0 or higher (5-point scale) on an item "like to use this system frequently." Similar ratings showed that participants felt the system was easy to learn, easy to use, well integrated, and that they were confident in using the system without the need for support personnel. Participants in this earlier evaluation did, however, have many suggestions on how the system could be improved. Since the current IMPACT simulation has addressed the majority of the comments recorded from the initial evaluation, as well as added a map-based play calling radial menu, included interfaces for the operator to view the autonomy's rationale for proposed play plans, and expanded play management interfaces, it is expected that the results from the ongoing evaluation will also be favorable.

5 Summary

The results of this research and development demonstrate that operator-autonomy teamwork can be supported with delegation control interfaces that enable an operator to quickly call plays that determine the automatic actions of one or more UVs. This capability, as well as the ability to employ a wide spectrum of control and multiple control modalities, will help support future multi-UV operations. At one extreme of this flexible adaptable approach, the operator can quickly task UVs by simply specifying play type and location, and the autonomy determines all other task-related details. At the other, the operator can exert manual control or build plays from the ground up, specifying detailed parameters. Our approach facilitates efficient support of the operator's current desired level of control by making the parameters most likely to be adjusted more accessible. This also assists operator communication with autonomy on mission factors key to optimizing play parameters (e.g., target size and current visibility). Additionally, concise and integrated interfaces that highlight the tradeoffs of autonomy-generated plans, support allocation of assets to tasks, and communicate mission progress are important for shared operator-autonomy situation awareness.

The authors anticipate that the advantages of our approach for implementing adaptable automation for tasks involving heterogeneous UVs in a base mission defense scenario will generalize to other applications. Nevertheless, for any new application, the desired operator-autonomy interactions would need to be identified to serve as the basis of interface designs for control and display of plays and the provision of a wide spectrum of multi-modal control. Application of a flexible interaction approach should be useful for a variety of domains promoting operator-autonomy collaboration and teamwork to enable agile responses in dynamic environments.

Acknowledgments The authors thank Chad Breeden and Patrick Dudenhofer (previous interface-focused team members) who made significant contributions. Appreciation is also extended to other members of the IMPACT effort who contributed to operator-autonomy interface design: Dakota Evans, Allen Rowe, and Sarah Spriggs.

This research and development supports The Office of the Assistant Secretary of Defense for Research and Engineering (ASD(R&E)) Autonomy Research Pilot Initiative (ARPI) titled “Realizing Autonomy via Intelligent Adaptive Hybrid Control” that is developing an Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT).

References

1. Behymer, K.J., Patzek, M.J., Rothwell, C.D., Ruff, H.A.: Initial evaluation of the intelligent multi-uxv planner with adaptive collaborative/control technologies (IMPACT). Technical report, AFRL-RH-WP-TR-2016-TBD (in preparation)
2. Rowe, A.J., Spriggs, A., Hooper, D.: Fusion: a framework for human interaction with flexible-adaptive automation across multiple unmanned systems. In: Proceedings of the 18th International Symposium on Aviation Psychology, OH (2015)
3. Calhoun, G.L., Goodrich, M.A., Dougherty, J.R., Adams, J.A.: Human-autonomy collaboration and coordination toward multi-RPA missions. In: Cooke, N., Rowe, L., Bennett, W. (eds.) Remotely piloted aircraft: a human systems integration perspective (Chapter 5). Wiley (in press)
4. Sheridan, T.B., Parasuraman, R.: Human-automation interaction. In: Nickerson, R.S. (ed.) Reviews of human factors and ergonomics. Human Factors and Ergon Soc, 1(2), pp. 89–129, Santa Monica, CA (2006)
5. Calhoun, G., Draper, M.H., Ruff, H., Barry, T., Miller, C.A., Hamell, J.: future unmanned aerial systems control: feedback on a highly flexible operator-automation delegation interface concept. In: AIAA Infotech at Aerospace Conference, Garden Grove, CA, (AIAA-2012–2549), pp. 1–16 (2012)
6. Eggers, J.W., Draper, M.H.: Multi-uav control for tactical reconnaissance and close air support missions: operator perspectives and design challenges. In: Proceedings of the NATO RTO Human Factors and Medicine Panel Symposium HFM-135 held in Biarritz, France, NATO TRO: Neuilly-sur-Siene, CEDEX (9–11 October, 2006)
7. Calhoun, G., Draper, M., Miller, C., Ruff, H., Breeden, C., Hamell, J.: Adaptable Automation interface for multi-unmanned aerial systems control: preliminary usability evaluation. In: Proceedings of the Human Factors and Ergonomics Society, pp. 26–30 (2013)
8. Miller C.A., Funk, H., Goldman, R., Meisner, J., Wu, P.: Implications of adaptive versus adaptable UIs on decision making: Why automated adaptiveness is not always the right answer. In: Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas (2005)
9. Draper, M.H., Miller, C.A., Calhoun, G.L., Ruff, H., Hamell, J., Benton, J., Barry, T.: Multi-unmanned aerial vehicle systems control via flexible levels of interaction: an adaptable operator-automation interface concept demonstration. In: AIAA Infotech at Aerospace Conference, Boston, MA, (AIAA-2013–4803) (2013)
10. Mersch, E.M., Behymer, K.J., Calhoun, G.L., Ruff, H.A., Dewey, J.S.: Game-based delegation interface design for supervisory control of multiple unmanned vehicles. In: Human Factors and Ergonomics Annual Meeting (proposal submitted)
11. Behymer, J., Mersch, E.M., Ruff, H.A., Calhoun, G.L., Spriggs, S.E.: Unmanned vehicle plan comparison visualizations for effective human-autonomy teaming. In: 6th International Conference on Applied Human Factors and Ergonomics AHFE Conference (2014)

12. Behymer, K.J., Ruff, H.A., Mersch, E.M., Calhoun, G.L., Spriggs, S.E.: Visualization methods for communicating unmanned vehicle plan status. In: *International Symposium for Aviation Psychology*, OH (2015)
13. Calhoun, G.L., Draper, M.H.: Display and control concepts for multi-UAV applications. In Valavanis K.P., Vachtsevanos, G.J. (eds.) *Handbook of unmanned aerial vehicles* (Chapter 101, pp. 2443–2473), Heidelberg, Germany: Springer Science + Business Media Dordrecht (2014)

Shaping Trust Through Transparent Design: Theoretical and Experimental Guidelines

Joseph B. Lyons, Garrett G. Sadler, Kolina Koltai, Henri Battiste, Nhut T. Ho, Lauren C. Hoffmann, David Smith, Walter Johnson and Robert Shively

Abstract The current research discusses transparency as a means to enable trust of automated systems. Commercial pilots (N = 13) interacted with an automated aid for emergency landings. The automated aid provided decision support during a complex task where pilots were instructed to land several aircraft simultaneously. Three transparency conditions were used to examine the impact of transparency on pilot's trust of the tool. The conditions were: baseline (i.e., the existing tool interface), value (where the tool provided a numeric value for the likely success of a particular airport for that aircraft), and logic (where the tool provided the rationale for the recommendation). Trust was highest in the logic condition, which is con-

J.B. Lyons (✉)

Air Force Research Laboratory, Wright-Patterson AFB, Dayton 45433, OH, USA
e-mail: joseph.lyons.6@us.af.mil

G.G. Sadler · K. Koltai · H. Battiste · N.T. Ho · L.C. Hoffmann
NVH Human Systems Integration, Canoga Park, Los Angeles, CA, USA
e-mail: garrett.g.sadler@gmail.com

K. Koltai
e-mail: kolina.koltai@gmail.com

H. Battiste
e-mail: hbattiste@gmail.com

N.T. Ho
e-mail: nhut.ho.51@gmail.com

L.C. Hoffmann
e-mail: lauren.c.hoffmann@gmail.com

D. Smith · W. Johnson · R. Shively
NASA Ames Research Center, Moffett Field, Los Angeles, CA, USA
e-mail: david.smith@nasa.gov

W. Johnson
e-mail: walter.johnson@nasa.gov

R. Shively
e-mail: robert.shively@nasa.gov

© Springer International Publishing Switzerland 2017

P. Savage-Knepshild and J. Chen (eds.), *Advances in Human Factors in Robots and Unmanned Systems*, Advances in Intelligent Systems and Computing 499,
DOI 10.1007/978-3-319-41959-6_11

sistent with prior studies in this area. Implications for design are discussed in terms of promoting understanding of the rationale for automated recommendations.

Keywords Trust · Transparency · Automation

1 Introduction

Advanced technology has great promise to support improved task performance across a variety of domains. Yet, advances in technologies such as automation, while beneficial to performance in stable (high-reliability) states, can have detrimental effects when they fail [1]. One paradoxical reason why automation can be devastating is that humans may form inappropriate reliance strategies when working with automation [2, 3]. Thus, the issue of trust in automation has emerged as an important topic for human factors researchers [4, 5]. Trust is a critical process to understand because trust has implications for reliance behavior—i.e., using or “relying” on a system when that reliance matters most. The trust process as it relates to automation is complex because the factors that influence trust range from human-centric factors such as dispositional influences (e.g. predisposition to trust) and experiential influences (learned trust aspects), to situational features [see 5 for a recent review]. Failure to establish appropriate trust can result in performance errors due to over-trust in technology where a human places unwarranted reliance on a technology, or alternatively, humans can under-trust technology by failing to use technology when that reliance is warranted. One key for researchers is to identify the set of variables that influences the trust process and to provide humans with the appropriate information to drive appropriate reliance decisions. The current paper discusses one such influence, the role of transparency and its influence on the trust process by presenting experimental data related to different transparency manipulations in a high-fidelity, immersive commercial aviation task environment involving automation support to a pilot.

Transparency represents a method for establishing shared awareness and shared intent between humans and machines [6]. Transparency is essentially a way for the human and the machine to be on the same page with regard to goals, processes, tasks, division of labor within tasks, and overall intent-based approach toward the interaction. Lyons [6] outlines several dimensions of transparency: intent, environment, task, analytic, team, human state, and social intent. The intent dimension involves understanding the overall purpose of the technology and how well this purpose matches the expectations of the human partner. Human expectations can be driven by naming schemes, physical appearance or other symbols, as well as by descriptions of the technology and prior experiences with similar technologies. The environment component involves educating the human (either through training or real-time display features) about how the technology senses information in the environment. The task dimension involves communicating the technology’s limitations, capabilities, and task completion information to the human. The analytic

dimension involves sharing details about the rationale for behaviors taken or recommendations provided by the system as well as providing the human with an understanding of the programming of the technology (i.e., “how it works”). The team component involves understanding the division of labor between the human and the technology. The human state dimension involves communicating information about the human operator (e.g., stress, fatigue) to the technology. Finally, the social intent facet of transparency involves communicating information to the human regarding the planned pattern of interactions (e.g., style, timing, etiquette, etc.) between the human and the technology.

Previous research has found that transparency is beneficial to humans interacting with automated decision aids [7–9]. Transparency in these contexts has been shown to influence trust by conveying necessary information about the limitations, logic, or intent of a system. Transparency has also been explored in the context of automation for commercial aviation. Lyons and colleagues [9] systematically manipulated different levels of transparency in NASA’s Emergency Landing Planner tool (ELP) [10]. The ELP was designed as an automated aid to support rapid decisions for commercial pilots to support effective diversion decisions. They sought to examine the potential benefits of added rationale for recommendations provided by the ELP by creating two additional display features to augment the existing ELP infrastructure. The first feature, termed value, added a numeric value reflecting the calculated probability for a successful landing for that particular diversion airport on the first attempt (i.e., without requiring a “go-around”). This subtle (but in no way simple) calculation was believed to increase the credibility of an option and provide the pilots with a quick estimate on the feasibility of a landing option. However, to make the ELP more transparent to the pilots, the study authors added a second feature which explained the rationale, termed logic, for the recommendation. This added information communicated the reasons why this diversion airport was a good or bad option. Using a static, low-fidelity task scenario, the authors found that trust was rated highest when the pilots were given the highest level of transparency for the ELP (i.e., the logic condition). It was unknown however, whether these same benefits would transfer to a more realistic task environment characterized by complexity and time constraints. The same design principles used in the aforementioned study (i.e., value and logic transparency) were used as a template for the current study, though the current study used a high-fidelity task environment to examine the effects of transparency on trust.

2 Method

2.1 Participants

The participants were 13 commercial transport pilots experienced with glass-cockpit instruments and with flight management systems (FMS). Participants were recruited locally from the San Francisco Bay Area through the San Jose State

University Research Foundation in conjunction with the Human Systems Integration Division at NASA Ames Research Center. They all had over 10000 h of flight experience as line pilots with the exception of a single participant who had between 3001 to 5000 h of experience. All participants had real-world experience making diversions from their filed flight plans for a variety of reasons including bad weather, traffic issues, mechanical failure, and/or medical emergencies. Participants were either employed by their airlines as Captains (66.7 %) or as First Officers (33.3 %). Two-thirds of participants had prior military flying experience. A majority of pilots (75 %) indicated that they were either “somewhat familiar,” “familiar,” or “very familiar” with flying in the study’s simulated geographical area (Colorado-Utah-Wyoming).

2.2 *Experimental Design*

We used a within subject factorial design with three levels of transparency. The levels of the Transparency corresponded to providing the participant with no explanation (baseline) for the automation’s diversion recommendation, just success probability (value), and success probability plus explanation (logic) for the automation’s recommendation. This additive manipulation of transparency facets is consistent with similar methods in prior research [7].

Six experimental scenarios were constructed with six aircraft in each scenario, and presented to the participants in a singular fixed order. Each scenario was designed such that the best available landing options afforded a high success probability to three of the aircraft, but only a low success probability to the other three aircraft. The order in which the aircraft diversions occurred was experimentally prescribed for each of the six scenarios such that, when collapsed over participants, each scenario had an equal number of landings affording high and low Success Probability. Finally, the order of presentation of the transparency conditions was also counterbalanced. Each Transparency condition was presented in blocked fashion, with three blocks and two scenarios per block. This provided six potential block orderings, with each of these orderings given to at least two participants.

2.3 *Task/Apparatus*

A dynamic commercial simulation environment was used for the current study in which an operator at an advanced ground station monitored and produced diversions for aircraft. This study utilized a subset of the functionalities of the whole prototype ground station, specifically six principal components: a Traffic Situation Display (TSD), an Aircraft Control List (ACL), Automatic Terminal Information

Service (ATIS) broadcasts, FAA-issued approach plates and airport charts, and pop-up windows containing evaluations of specific diversions provided by the Autonomous Constrained Flight Planner (ACFP) recommender system, and the ACFP itself (see Fig. 1). The ACFP is a tool being designed to support flight path monitoring and re-routing for NASA’s Reduced Crew Operations (RCO) project [11], and which directly incorporated the ELP algorithm [10], served as the automated diversion recommendation aid during a complex landing scenario. Each of these diversions specified a runway at a specific airport, along with the route to that runway. The TSD provided participants with a visual display of the geographic area, convective weather cells, turbulence boxes, icons representing the locations of available airports, and information related to each aircraft’s current state: location, heading, altitude, and indicated airspeed. Using the ACL, participants were able to toggle focus between the six simulated aircraft in the TSD and look up the selected aircraft’s type (e.g., Boeing 747, Airbus A340, etc.). Local airport weather conditions were available to participants by requesting (from a menu accessed in the TSD) the ATIS broadcast for the corresponding landing site. Approach plate



Fig. 1 Example experimental ground station

information allowed participants to look up a schematic diagram for each available approach at a given airport in addition to legal requirements (e.g., weather ceiling minimums) necessary for the landing. Finally, the ACFP pop-up window interface provided participants with ACFP's recommendation for a landing site together with varying degrees and kinds of transparency information depending on the scenario's transparency condition (detailed below). In the scenario, participants were instructed to land all aircraft under their control, this resulted in the need to land 6 aircraft in each trial.

Following the examples set forth in [9] information presented to the participant in the ACFP window varied across scenarios using three hierarchical levels of transparency, identified here as baseline, value, and logic. In the baseline transparency condition (Fig. 2), participants were provided a recommendation from ACFP displaying the recommended landing site (airport and runway number), runway length (in feet), approach name/type, and distance to the landing site (in nautical miles). The value transparency condition (Fig. 2) included, in addition to the information presented in the baseline transparency condition, a "risk statement" that provided ACFP's evaluation of the probability of success for landing on the first attempt (e.g., "There is a 55 % chance that you will be able to successfully complete the approach and landing under current conditions"). It is important to note that a success probability of 55 % means that there is a 45 % chance of having to perform a "go-around" or follow-up attempts of the approach, not a 45 % chance of crashing. Finally, the logic transparency condition (Fig. 2) included all information presented in the low and medium conditions as well as statements to explain the ACFP's rationale behind its recommendation. These statements gave descriptions of relevant factors along the enroute, approach, and landing phases of flight that led to its determination for the recommendation See Figs. 3 and 4.



Fig. 2 Screen capture of the baseline transparency condition

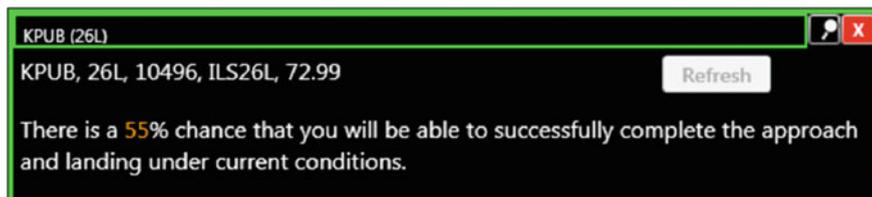


Fig. 3 Screen capture of the value transparency condition

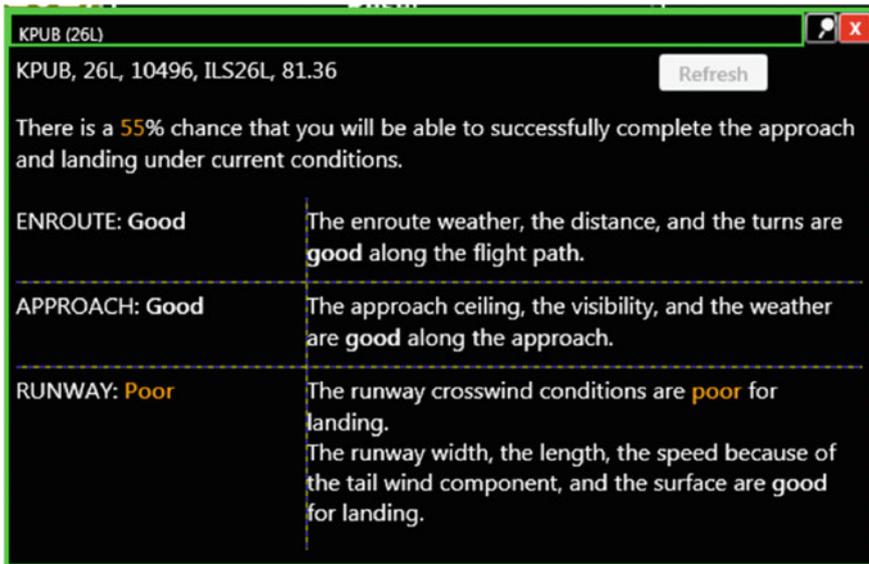


Fig. 4 Screen capture of the logic transparency condition

2.4 Measures

Trust was measured using a 7-item scale to gauge pilot’s intentions to be vulnerable to the ACFP [9]. Participants rated their agreement with the items using a 7-point Likert scale. Trust measures were taken after each transparency condition and the scale evidenced high reliability with alphas ranging from .88–.92. Example items included: “I think using the [ACFP] will lead to positive outcomes,” “I would feel comfortable relying the recommendations of the [ACFP] in the future,” and “when the task was hard I felt like I could depend on the [ACFP].”

3 Results

The order of transparency conditions was counterbalanced within a repeated measures design to maximize statistical power. To explore potential order effects of the transparency conditions over time, a repeated measures analysis was conducted. While there was no main effect of order, nor a main effect of time on trust (all p 's >0.05), there was a significant time by order interaction, $F(5, 7) = 12.44, p <0.05$. As depicted in Fig. 5, the interaction effect follows a quadratic trend such that participants tend to report higher trust when they interact with the logic form of transparency either early (e.g., 5) or later in the task (e.g., 3 and 1). It is also clear that participant interactions with the baseline transparency resulted in lower trust

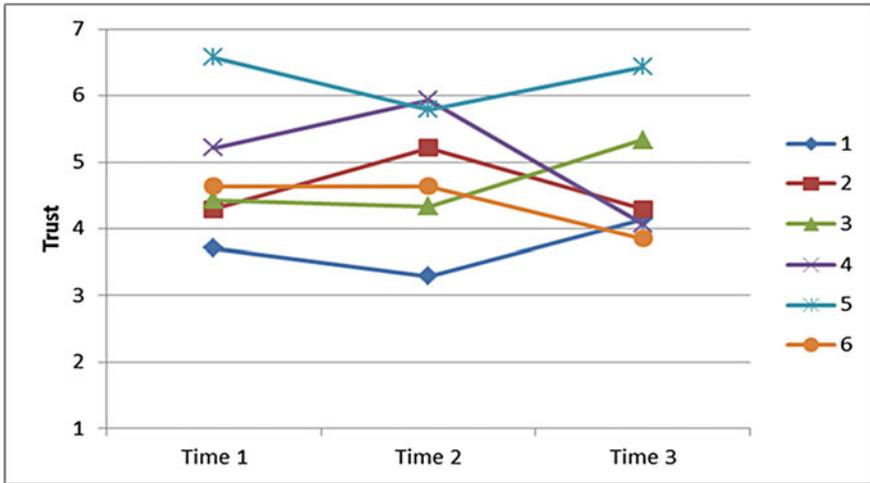


Fig. 5 Time by order interaction predicting trust

when it followed either the logic or value transparency (e.g., 4 and 6). Given that the order of the transparency conditions did have an influence on trust overtime, we used a repeated measures ANCOVA to examine the impact of transparency condition on trust while including order as a covariate. As shown in Fig. 6, trust was highest in the logic condition and lowest in the baseline condition. These

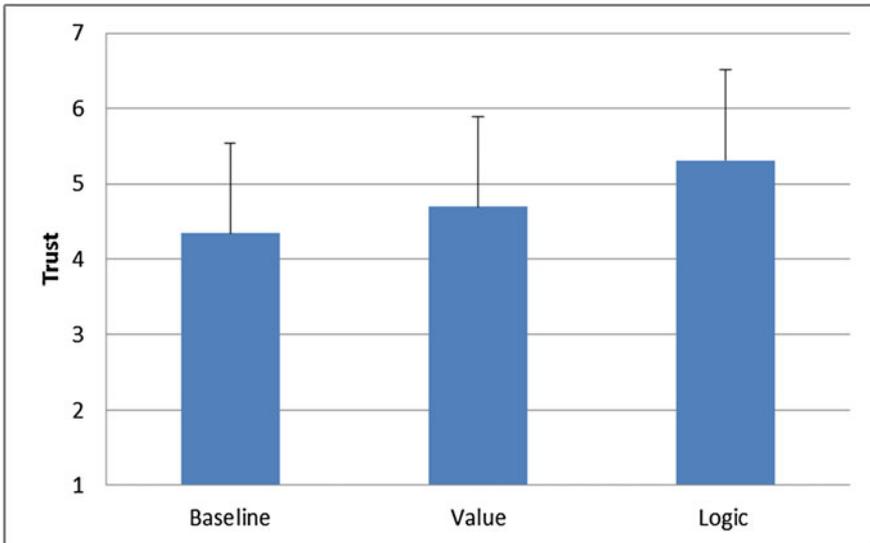


Fig. 6 Means for trust by condition

differences were reliable, $F(2, 22) = 4.39$, $p < 0.05$, demonstrating that trust was influenced by transparency condition and that the highest level of trust was associated with the logic-based form of transparency.

4 Discussion

Trust of automated systems remains a highly pertinent topic for researchers given the burgeoning nature of advanced technology. Technology offers the promise of improved performance and reduced workload for human users/managers of technology, yet these benefits will only be realized when the technology is designed in such a way as to foster appropriate reliance. One such method involves adding transparency features to automated systems. The present research explored the impact of transparency on trust using a high-fidelity simulation involving an automated aid in commercial aviation.

Consistent with prior research, the current study demonstrated that higher levels of transparency engender higher trust of automation. Specifically, the use of logic-based explanations for the recommendations was found to promote trust. This is consistent with a prior study that used similar transparency manipulations [9], however that previous study was conducted using low-fidelity methods. The current results naturally extend prior research by demonstrating the benefits of logic-based transparency in a high-fidelity task simulation using commercial pilots as the human operators. Clearly, when automated aids offer recommendations to humans they should include information related to the rationale or the key drivers of the recommendation, as this will help to foster trust in the automation. The rationale provided by the automation will help to reduce uncertainty on behalf of the human.

Future research should continue to explore the impact of transparency on the trust process. Future studies might consider a variety of different forms of transparency. The SA-based model of transparency highlights the importance of perception, comprehension, and projection and their additive effects [7]. Perhaps most importantly, Mercado and colleagues [7] found that higher levels of transparency modulated trust with no detriment to cognitive workload. This is critical as added information has the potential for overloading operators, which is counterproductive. Further, future research should explore an expanded view of transparency as outlined in [6].

References

1. Onnasch, L., Wickens, C.D., Li, H., Manzey, D.: Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Hum Factors* **56**, 476–488 (2014)
2. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum Factors* **46**, 50–80 (2004)

3. Lyons, J.B., Stokes, C.K.: Human-human reliance in the context of automation. *Hum Factors* **54**, 111–120 (2012)
4. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multirobot control: a review of the human factors issues. *IEEE Transactions on Human-Machine Systems*, 13–29 (2014)
5. Hoff, K.A., Bashir, M.: Trust in automation: integrating empirical evidence on factors that influence trust. *Hum Factors* **57**, 407–434 (2015)
6. Lyons, J.B.: Being transparent about transparency: a model for human-robot interaction. In: Sofge, D., Kruijff, G.J., Lawless, W.F. (eds.) *Trust and Autonomous Systems: papers from the AAAI spring symposium (Technical Report SS-13-07)*. AAAI Press, Menlo Park, CA (2013)
7. Mercado, J.E., Rupp, M.A., Chen, J.Y.C., Barnes, M.J., Barber, D., Procci, K.: Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors* (in press)
8. Wang, L., Jamieson, G.A., Hollands, J.G.: Trust and reliance on an automated combat identification system. *Hum Factors* **51**, 281–291 (2009)
9. Lyons, J.B., Koltai, K.S., Ho, N.T., Johnson, W.B., Smith, D.E., Shively, J.R.: Engineering trust in complex automated systems. *Ergon in Design* **24**, 13–17 (2016)
10. Meuleau, N., Plaunt, C., Smith, D., Smith, C.: Emergency landing planner for damaged aircraft. In: *Proceedings of the Scheduling and Planning Applications Workshop* (2008)
11. Brandt, S.L., Lachter, J., Battiste, V., Johnson, W.: Pilot situation awareness and its implications for single pilot operations: analysis of a human-in-the-loop study. *Procedia Manufacturing* **3**, 3017–3024 (2015)

A Framework for Human-Agent Social Systems: The Role of Non-technical Factors in Operation Success

Monika Lohani, Charlene Stokes, Natalia Dashan, Marissa McCoy, Christopher A. Bailey and Susan E. Rivers

Abstract We present a comprehensive framework that identifies a number of factors that impact human-agent team building, including human, agent, and environmental factors. This framework integrates existing empirical work in organization behavior, non-technical training, and human-agent interaction to support successful human-agent operations. We conclude by discussing implications and next steps to evaluate and expand our framework with the aim of guiding future attempts to create efficient human-agent teams and improve mission outcomes.

Keywords Social and emotional interaction · Human factors · Non-technical skills · Operation success · Human-Agent teams

M. Lohani (✉) · C. Stokes · N. Dashan · M. McCoy · C.A. Bailey · S.E. Rivers
Department of Psychology, Yale University, 340 Edwards Street,
New Haven, CT, USA
e-mail: monika.lohani@yale.edu

C. Stokes
e-mail: charlene.stokes@yale.edu

N. Dashan
e-mail: natalia.dashan@yale.edu

M. McCoy
e-mail: marissa.mccoy@yale.edu

C.A. Bailey
e-mail: christopher.a.bailey@yale.edu

S.E. Rivers
e-mail: susan.rivers@yale.edu

C. Stokes
Air Force Research Laboratory, Wright-Patterson Air Force Base,
Dayton, OH 45433, USA

1 Introduction

Social interactions are the basis of successful team building. Human-agent (H-A) interactions may share similar characteristics to interactions between humans in ways that impact the formation of team building between a human and an agent. We present a comprehensive framework for *human-agent social systems*, which is informed by organization behavior, non-technical training work, as well as human-systems interaction literature. We integrate existing empirical work that provide initial evidence to support the proposed components in our framework that are critical to accomplish successful H-A operations. Specifically, we discuss the effect of non-technical skills on H-A team effectiveness. We conclude by identifying the next steps to further evaluate and expand our framework with the aim of guiding future attempts to create efficient H-A teams and improve operation outcomes.

2 Human-Agent Team Effectiveness for Operation Success

An agent is a sophisticated computer system that is situated in some environment and can act autonomously and flexibly to solve a growing number of complex problems [1]. This general category may include, but is not limited to agents in virtual environments (virtual agents) and agents with integrated perception, cognition, and action (robotic agents). A H-A team refers to a set of two or more members who have each been assigned specific roles or functions to perform and they interact dynamically, interdependently, and adaptively toward a common and valued goal. H-A teams can perform different tasks and operate in various contexts, e.g., military, aviation, and search and rescue teams. Team-outcomes are the resulting consequences and end products of team work. These may depend on the team-goal specific to the mission and may include performance (e.g., successful search and rescue missions). Successful completion of team goals will result in team-operation success.

Which elements can influence H-A team operation success? Recent reviews [2, 3] found a number of factors that affect team building, including human, agent, and environmental factors can influence trust between a human and his/her agent teammate and impact operation success. Numerous models have also been proposed that focus on one or more elements that may influence H-A teams e.g., [4, 5]. Although these models have demonstrated the relevance of many factors that should be considered, the existing literature does not have existing framework that inform researchers to build H-A teams. In the context of H-A interaction, both human and team characteristics can influence team work. No framework to our knowledge provides a comprehensive framework that includes agent-related, human-related, and H-A team related factors as inputs to lead to H-A operation success.

In the current paper, we sketch a framework that integrates findings from humans interacting with computers, machines, virtual characters, and robots that could guide future research that aims to develop H-A teams. Of course, there are many overlaps across domains but also significant differences. With the growth of all these sub-fields, it is important to understand how these areas of research may converge. This is not to say that these areas are the same or that the findings from one are generally applicable to all others. On the other hand, these areas have an important overlap that can help inform each other that can help build an understanding of how humans interact with them.

3 A Framework for the Human-Agent Social Systems

We have developed a comprehensive framework that incorporates elements that may *collectively* influence how well H-A teams may perform together. Figure 1 presents our framework. We have used an input-process-outcome approach [6] for examining team effectiveness. *Team-inputs* are factors that allow and limit H-A team interactions. Inputs include individual team member characteristics (e.g., agent’s and human’s physical features, personality), individual member capabilities (e.g., agent’s and human’s technical expertise, task-management ability, situation awareness, decision-making ability), and team-level factors (e.g., H-A team’s rapport, communication, cooperation, team management). Contextual factors are indicated by the grey elliptical shape around H-A factor, which may also significantly influence and define team-level factors (e.g., setting the H-A team is working in, environmental complexity and uncertainties).

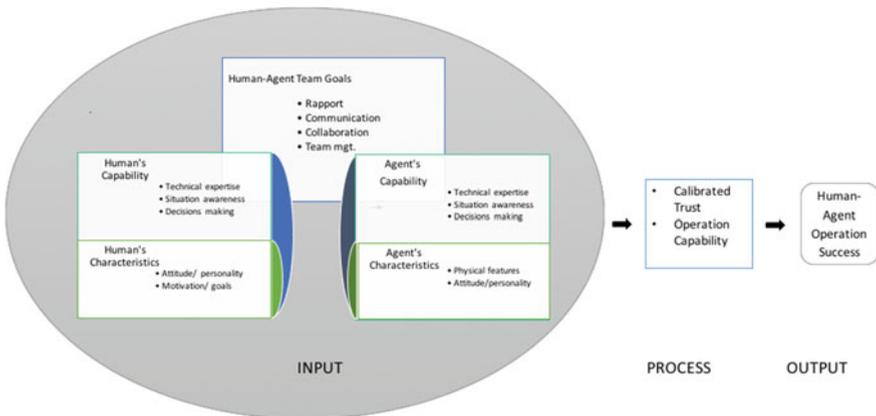


Fig. 1 A framework for human-agent social systems for operation success

Input-level factors integrate and determine team-processes (e.g., calibrated trust, reliance, and operation capabilities), which ultimately lead to output (e.g., H-A operation success). Team-outcomes are the resulting consequences and end products of team work. These may depend on the team-goal specific to the mission and may include performance (e.g., successful search and rescue mission). Successful completion of these goals will result in team-operation success.

When working together in operations with high stakes, a H-A team must work effectively even under highly complex, stressful, and time-sensitive conditions. But how do we create a team that works effectively? Which components are essential to consider when trying to build effective H-A teams? The goal of the current paper was to assemble a framework that can guide developers and researchers to build successful H-A teams. We incorporated existing evidence to support this framework and identified gaps that need to be addressed in future work. We presented a testable theoretical framework with factors that may influence H-A teams. This framework is a working model that will be improved by empirical testing with the ultimate goal of developing guidelines for developing efficient H-A team performance.

Table 1 provides details on each component of the framework. Both, the human and agent have characteristics and resources that contribute to team outcomes. While human-specific factors certainly are important to consider, these are beyond the scope of this paper. Much theoretical and empirical work in H-A research has explored design-based approaches to inform H-A interactions. For instance, in order for a human to perceive an agent as a peer (i.e., equal to oneself, similar in ability), the agent should express its motives and goals through complex behavioral and social characteristics that are similar to those of humans [7, 8]. For example, a humanoid robot that demonstrates gaze, eye contact, and utilitarian emotions is more likeable and facilitate team dynamics [9]. Many technical capabilities built into the agent have tremendously improved technical capabilities, situation awareness, and decision making [10]. These may include artificial cognition, speech recognition, natural language understanding & generation, dialogue management, emotion modeling, question response managers, speech generation, and non-verbal behavior. Unlike H-A team goals, much theoretical and empirical progress has been made with these components. Accordingly, in this paper, we specifically focus on how non-technical skills play a critical role in H-A team success after we briefly discuss agent's characteristics and capabilities.

4 Human-Agent Team Goals—Supporting Evidence

Non-technical skills include cognitive, social, and emotional abilities that are fundamental in team functioning. Non-technical skills, more so than technical abilities, have been recognized as leading contributors to accidents, errors, and low quality outcomes in human teams as well as H-A teams [11]. Evidence suggests that certain nontechnical aspects of performance can enhance or, if lacking, contribute to

Table 1 Components and underlying elements of human-agent social systems framework

	Component Element	Definition	Description of sub-components included
1	Human-Agent team goals		
1.1	Rapport	A flow in an interaction marked by harmony, synchrony, and connectedness	Behaviors indicating positive emotions (head nods or smiles, mutual attentiveness (e.g. mutual gaze), coordination (e.g., postural mimicry or synchronized movements))
1.2	Communication	Establishing shared understanding	Exchanging information; establishing roles and responsibilities; expressing vulnerability
1.3	Coordination	Coordinating task activities	Delegating tasks, mutual performance monitoring, minimizing idle time, giving implicit and explicit commands;
1.4	Team management	Interactions directed toward task accomplishment (planning/preparation); Prioritizing; Exercising leadership and followership; Engaging in adaptive and supportive behavior (team members step in and help when needed, dynamically reallocate workload, ask for assistance when needed)	Negotiating; using authority, assertiveness, critical judgment (when applicable); assessing capabilities; supporting others; tactical questioning; perspective tasking; conflict management; Coping with pressure
2	Agent’s capabilities		
2.1	Technical expertise	Task-dependent capabilities to support mission	Artificial cognition; Language generation and understanding
2.2	Situation awareness	Recognizing the need for an action, having common expectations, actively seeking information, and taking corrective actions	Gathering information; Anticipating future state; Calling for help
2.3	Decision making	Providing important input to make critical mission-related decisions	Identifying options; selecting and communicating options; implementing and reviewing decisions; balancing risks and benefits; Reevaluating

(continued)

Table 1 (continued)

	Component Element	Definition	Description of sub-components included
3	Agent's characteristics		
3.1	Physical design	Mission dependent design	Anthropomorphic features; speech qualities; posture shifts
3.2	Attitude/Personality	Adopting context-suitable attitudes	Agreeableness; Conscientiousness

deterioration of H-A team outcomes [2, 3]. Non-technical skills are critical in effective team building, emotion management, and performance [12–15].

Previous literature on team building processes among human team members has found that numerous factors are critical for team performance and effective teamwork, such as communication, coordination, decision-making, conflict management, and performance feedback, for e.g., [13–15]. After a thorough literature review, we have identified a few elements that have been consistently found to be critical in team work. Our framework incorporates elements of these non-technical skills as researched in human teams to H-A teams. In the sections below, we discuss empirical evidence supporting the importance of these elements in a H-A teams.

4.1 Rapport

Rapport is a feeling of mutual harmony, or being “in sync” with a teammate and it has been argued to “lead to communicative efficiency, better learning outcomes, improved acceptance of medical advice and successful negotiations.” [16]. This idea of rapport has been discussed both in the Human-Human interaction literature and in the Human-Robot interaction literature. Research has been conducted on the specifics of establishing rapport through the manipulation of these three elements: attention, positivity, and coordination. For example, rapport with virtual agents is many times broken due to the agent’s speech patterns. Furthermore, for rapport to be established, there has to be a “tight sense-act loop that has been traditionally lacking in embodied conversational agents” [17]. It is not enough for virtual agents to simply talk to their human counterparts. It is vital for the agent to appear to be responding to the human. In an experiment, it was shown that contingency, not just frequency of positive feedback was needed to create rapport [18]. It must seem as if the agent is reacting to you personally and quickly. This sort of feedback (“envelope feedback,” which includes head nods, gaze, and mutual beat gestures) has been shown even more important in building rapport than emotional feedback (which includes gestures such as smiling or showing puzzlement) [19]. Social cueing not only improves likeability of the agent, but enables agents to work as partners [7, 9, 10, 20].

4.2 *Communication*

Rapport alone is not enough to foster effective teamwork: humans-agent teams must not only act in synchrony, but they must also be able to exchange ideas with each other. *Communication* is different from rapport in that rapport is generally felt by the parties, but is difficult to observe, whereas communication is an observable exchange of information and sentiments. Encouraging peer-to-peer dialogue between humans and agents such that they communicate and coordinate with each other as human teams do may also influence performance [9]. This kind of ongoing dialogue engenders contextual and situational awareness between the H-A teams and may build a support system for both. Socially communicative agents relative to less social ones gain more acceptance as a conversational partner [20, 21].

According to [16], an agent's communication can be altered in several ways, one of which is through modeling of emotional state. If an agent is aware of another's emotional state (such as if somebody is frowning, or raises their voice in anger), then the agent will alter its communication to ask what is wrong and try to remedy the situation (if an agent is unable to perceive shifts in emotional states then it would not alter what it does or what it says). The role of emotion detection in communication is high. Some potential roles of emotions for artificial agents include: alarm mechanisms, action selection, adaptation, learning, motivation, social regulation, goal management, strategic processing, memory control, information integration, attention focus, and self-modelling [22]. Communication about intentions facilitates task performance [23], and will likely minimize information overload.

Several attempts are underway to improve social agent architecture by incorporating verbal and non-verbal cues to encourage social interactions and cooperation, social acceptance, and bidirectional intent recognition [9, 10, 24]. Transparency shown by the agents, even when the agent has low reliability and performance, can lead to better reliance highlighting the importance of non-technical factors, such as communication, in trust development [2, 3].

4.3 *Coordination*

Within a team, it is important for team members to coordinate who does what task and when. Even if there is a high degree of rapport between team members, a team can be unproductive if they fail to coordinate. *Coordination* involves activities such as giving commands, expressing intent to proceed with an action, fluidly timing the actions, incorporating unexpected requests and actions into the workflow, and doing the action. Without proper coordination, team members would likely perform the same tasks needlessly, spend too much time doing nothing, spend too much time on irrelevant tasks, or be confused.

A framework of coordination [25] between teammates suggests a series of algorithms based not on stages of expected behavior and subgoals, but on ensuring fluid coordination between teammates. This framework is based on key results from Human-teamwork studies and included algorithms to minimize idle time, preference implicit commands to the near future (and preference explicit commands to be completed immediately), and manage multiple agents [25]. These factors are important in creating workplace flexibility, establishing who will do which task, and in managing on-the-fly decisions. For better coordination, effective teammates communicate often as to the progress of the task, consider the consequences of their actions on others, and anticipate another's needs. They also consider other people's capabilities into their own action planning [26] and try to minimize everyone's idle time. The most effective team members also use both implicit and explicit commands when making requests because these two types of commands have different properties [8, 25]. Explicit commands nearly always elicit an immediate response, whereas implicit commands usually imply a more flexible time response. Thus, implicit commands require a lower switching cost than explicit commands by allowing the command to be smoothly incorporated into the original workflow. Previous research has shown that teams exhibit more implicit coordination as time-pressure increases, and that more implicit coordination leads to better performance outcomes. One way to be an effective implicit communicator is to periodically assess the situation [27].

4.4 Team Management

Within a team, rapport is needed to create synchrony and positive feelings, which lead to trust. Coordination is needed in order to complete immediate tasks. Something else is needed to ensure that everything runs smoothly and that small tasks aid in the completion of a larger goal. Problems can occur when either a human or agent partner miscommunicates regarding its goals and intentions, when external factors change and render the current goals useless, when mood changes, or when one partner has to unexpectedly pick up a task from another (either due to a situational change or an emotional change, such as lack of or breach of trust). Team management is the factor of effective teams that is needed to plan, anticipate conflict, solve problems, allow for performance feedback, and ensure team success.

One aspect of team management is negotiation. Sometimes the human and the virtual agent may disagree, and to resolve the conflict, it is essential that an agreed-on and effective conflict resolution system be in place. According to one framework, the negotiator should balance three different goals to be successful: solving problems, gaining trust, and managing the interaction [28]. Team negotiation can be represented by a sequence of negotiation stances: the agent who holds the stance, the action that this stance is about, the stance the agent holds toward the

action, the audience (a set of agents) that the agent has made the stance in front of, the reason for holding the stance, and the time at which the stance was made [29]. An agent or a human can accept, reject, explain, or counter propose a negotiation stance. There are numerous elements are essential for team management as listed in Table 1, such as critical judgment and conflict management, which we are unable to go into details because of space limitations. Recent work is exploring ways to include team management skills in H-A team setup. For instance, benefits of self-disclosure, expressivity, social bonds [30–35] can be used to create a dialog between the team such that the members can empathize, reassure, motivate, and support their team.

4.5 Conclusion

In sum, the H-A interactions could be framed to develop rapport, communication, collaboration, and team management skills. In systematic reviews, non-technical factors such as team collaboration and communication have a moderate to large effect on H-A trust [3]. Socioemotional skills can serve as a bonding mechanism for teams as they can promote trust. Adopting this team-building approach, we found that socioemotional interaction of a virtual agent as a teammate vs. a tool led to higher reports of the ability to cope with stress, acceptance of physiological assessment sensors, and it moderated the relationship between trust and reliance [30]. Another study showed that a robot can be programmed to provide different kinds of support to its human interaction partner in ways that effectively managed stress [33]. Past studies that have explored the effects of an agent's social abilities and social dialogue with a human increases acceptance and communication [8, 9, 30, 35]. Thus, limited, but promising studies illustrate that agents can provide non-technical support to their teammates.

5 Implications and Future Directions

Putting it together, our framework provides a comprehensive integration of agent-relevant, human-relevant, and H-A team-relevant factors that may contribute to successful team outcomes. We argue that traditional climate management and training techniques derived from the organization and education literature can be leveraged to foster a team atmosphere. The current framework is a working model that will evolve with new research. Based on the new evidence, many aspects may need to be added or revised. This framework was intentionally kept broad to maintain its applicability to most contexts in which H-A teams can work together. With further validation and improvement, this framework should help reduce the burden of humans and significantly improve H-A performance.

Much work is needed for an agent to be able to operate with a human teammate in complex and uncertain environments and perform in cooperation with human team members. Our framework provides key factors that are essential to develop future applications of agent technologies to be able to interactions with non-expert humans and play the role of a teammate and provide assistance to humans in high-stakes situations, such as warfighters in combat operations. Such teammate-support is very relevant for people exposed to high stress and high workload conditions, such as warfighters and intelligence analysts, where technologically savvy agents could provide technical expertise to help reduce the workload and also provide non-technical skills to improve stress coping by being a helpful teammate.

Acknowledgments This work is funded by the Air Force Research Laboratory.

References

1. Jennings, N.R., Wooldridge, M.: Applications of intelligent agents. In: Jennings, N.R., Wooldridge, M.J. (eds.) *Agent Technology*, pp. 3–28. Springer, Heidelberg (1998)
2. Schaefer, K.E., Billings, D.R., Szalma, J.L., Adams, J.K., Sanders, T.L., Chen, J.Y., Hancock, P.A.: A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction (ARL-TR-6984). MD. US Army Research Laboratory, Aberdeen Proving Ground (2014)
3. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction: human factors. *J. Hum. Factors Ergonomics Soc.* **53**, 517–527 (2011)
4. Trafton, J.G., Cassimatis, N.L., Bugajska, M.D., Brock, D.P., Mintz, F.E., Schultz, A.C.: Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **35**, 460–470 (2005)
5. Hoffman, G., Breazeal, C.: Cost-based anticipatory action selection for human-robot fluency. *IEEE Trans. Robot.* **23**, 952–961 (2007)
6. McGrath, J.E.: *Leadership Behavior: Some Requirements for Leadership Training*. U.S Civil Service Commission, Washington, D.C. (1962)
7. Breazeal, C.: Function meets style: insights from emotion theory applied to HRI. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* **34**, 187–194 (2004)
8. Breazeal, C., Kidd, C.D., Thomaz, A.L., Hoffman, G., Berlin, M.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: *IEEE/RSJ international conference on intelligent robots and systems*. pp. 708–713 (2005)
9. Breazeal, C.: Social interactions in HRI: the robot view. *IEEE Trans. Syst. Man Cybernet. C Appl. Rev.* **34**, 181–186 (2004)
10. Fong, T., Kunz, C., Hiatt, L.M., Bugajska, M.: The human-robot interaction operating system. In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction*. pp. 41–48 (2006)
11. Flin, R., O'Connor, P., Mearns, K.: Crew resource management: improving teamwork in high reliability industries. *Team Perform. Manag. Int. J.* **8**, 68–78 (2002)
12. Flin, R., Maran, N.: Identifying and training non-technical skills for teams in acute medicine. *Qual. Saf. Health Care* **13**, i80–i84 (2004)
13. Mathieu, J., Maynard, M.T., Rapp, T., Gilson, L.: Team effectiveness 1997–2007: a review of recent advancements and a glimpse into the future. *J. Manag.* **34**, 410–476 (2008)

14. Andersen, P.O., Jensen, M.K., Lippert, A., Østergaard, D.: Identifying non-technical skills and barriers for improvement of teamwork in cardiac arrest teams. *Resuscitation* **81**, 695–702 (2010)
15. Buljac-Samardzic, M., Dekker-van Doorn, C.M., Van Winjingaarden, J.D., Van Wijk K.P.: interventions to improve team effectiveness: a systematic review. *Health Policy*. **94**, 183–195 (2010)
16. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating rapport with virtual agents. In: Pelachaud, C., Martin, J.C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) *Intelligent Virtual Agents*. LNAI, vol. 4722, pp. 125–128. Springer, Heidelberg (2007)
17. Gratch, J., Okhmatovskaia, A., Laomthe, F., Marsella, S., Morales, M., Van der Werf, R.J., Morency, L.P.: Virtual rapport. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) *Intelligent Virtual Agents*. LNAI, vol. 4133, pp. 14–27. Springer, Heidelberg (2006)
18. Gratch, J., Wang, N., Okhmatovskaia, A., Laomthe, F., Morales, M., Van der Werf, R.J., Morency, L.P.: Can virtual humans be more engaging than real ones? In: Jacko, J.A. (eds.) *Human-Computer Interaction, HCI Intelligent Multimodal Interaction Environments*, LNCS, vol. 4552, pp. 286–297. Springer, Heidelberg (2007)
19. Cassel, J., Thorisson, K.R.: The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. *Appl. Artif. Intell.* **13**, 519–538 (1999)
20. Mutlu, B.: Designing embodied cues for dialog with robots. *AI Magazine*. **32**, 17–30 (2011)
21. Sidner, C., Rich, C., Shayganfar, M., Bickmore, T., Ring, L., Zhang, Z.: A robotic companion for social support of isolated older adults. In: *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction extended abstracts*. pp. 289–289 (2015)
22. Scheutz, M.: Evolution of affect and communication. In: Gökçay, D., Yildirim, G. (eds.) *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives*, pp. 75–93. IGI Global, Hersey (2010)
23. Harbers, M., Jonker, C., Van Riemsdijk, B.: Enhancing team performance through effective communication. In: *Proceedings of the 4th Annual Human-Agent-Robot Team Workshop*, pp. 1–2. Boston (2012)
24. Hayes, B., Scassellati, B.: Challenges in shared-environment human-robot collaboration. *Learning* **8**, 9–14 (2013)
25. Shah, J.A.: Fluid coordination of human-robot teams. Ph.D. Dissertation, Massachusetts Institute of Technology (2011)
26. Sebanz, N., Bekkering, H., Knoblich, G.: Joint action: bodies and minds moving together. *Trends Cogn. Sci.* **10**, 70–76 (2006)
27. Entin, E.E., Sefraty, D.: Adaptive team coordination. *Human Factors J. Human Factors Ergonomics Soc.* **41**, 312–325 (1999)
28. Traum, D., Marsella, S.C., Gratch, J., Lee, J., Hartholth, A.: Multi-party, multi-issue, multi-strategy negotiations for multi-modal virtual agents. In: Prendinger, H., Lester, J., Ishizuka, M. (eds.) *Intelligent Virtual Agents*. LNAI, vol. 5208, pp. 117–130. Springer, Heidelberg (2008)
29. Traum, D., Rickel, J., Gratch, J., Marsella, S.: Negotiation over tasks in hybrid human-agent teams for simulation-based training. In: *Proceedings of the second international joint conference on autonomous agents and multi-agent systems*, pp. 181–184. ACM, New York (2003)
30. Lohani, M., Stokes, C.K., McCoy, M., Bailey, C.A., Rivers, S.E.: Social interaction moderates human-robot trust-reliance relationship and improves stress coping. In: *Proceedings of ACM/IEEE international conference on human-robot interaction* (In Press)
31. Hoffman, G., Birnbaum, G.E., Vanunu, K., Sass, O., Reis, H.T.: Robot responsiveness to human disclosures affects social impression and appeal. In: *Proceedings of the 2014 ACM/IEEE international conference on human-robot interaction*, pp. 1–8 (2014)
32. Belpaeme, T., Baxter, P.E., Read, R., Wood, R., Cuaya’huitl, H., Kiefer, B., Humbert, R.: Multimodal child-robot interaction: building social bonds. *J. Human-Robot Interact.* **1**, 33–53 (2012)

33. Dang, T.H.H., Tapus, A.: Coping with stress using social robots as emotion-orientated tool: potential factors discovered from stress game experiment. In: Hermann, G., Pearson, M.J., Lenz, A., Bremner, P., Spiers, A., Leonards, U. (eds.) *Social Robots*. LNAI, vol. 8239, pp. 160–169. Springer, Heidelberg (2007)
34. Heerink, M., Kröse, B., Evers., V., Wielinga., B.: The influence of a robot's social abilities on acceptance by elderly users. In: *The 15th IEEE international symposium on robot and human interactive communication*. pp. 521–526 (2006)
35. Lohani., M., Stokes, C.K., McCoy, M., Bailey, C.A., Joshi, A., Rivers, S.E.: Perceived role of physiological sensors impacts trust and reliance on robots. In: *IEEE international symposium on robot and human interactive communication* (In Review)

Insights into Human-Agent Teaming: Intelligent Agent Transparency and Uncertainty

Kimberly Stowers, Nicholas Kasdaglis, Michael Rupp, Jessie Chen,
Daniel Barber and Michael Barnes

Abstract This paper discusses two studies testing the effects of agent transparency in joint cognitive systems involving supervisory control and decision-making. Specifically, we examine the impact of agent transparency on operator performance (decision accuracy), response time, perceived workload, perceived usability of the agent, and operator trust in the agent. Transparency has a positive impact on operator performance, usability, and trust, yet the depiction of uncertainty has potentially negative effects on usability and trust. Guidelines and considerations for displaying transparency in joint cognitive systems are discussed.

Keywords Transparency · Human factors · Human-Machine interaction · Systems engineering · Supervisory control · Unmanned vehicles

K. Stowers · N. Kasdaglis · M. Rupp · D. Barber
Institute for Simulation and Training, 3100 Technology Parkway,
Orlando, FL 32826, USA
e-mail: kstowers@ist.ucf.edu

N. Kasdaglis
e-mail: nkasdagl@ist.ucf.edu

M. Rupp
e-mail: mrupp@ist.ucf.edu

D. Barber
e-mail: dbarber@ist.ucf.edu

J. Chen (✉)
U.S. Army Research Laboratory, Orlando, FL 32826, USA
e-mail: yun-sheng.c.chen.civ@mail.mil

M. Barnes
U.S. Army Research Laboratory, Ft. Huachuca, AZ 85613, USA
e-mail: michael.j.barnes.civ@mail.mil

1 Background

As warfighting environments become more complex, operators will increasingly collaborate with intelligent agents (IAs) to manage teams of robotic systems [1]. However, increased autonomy may come with a cost: operators may have difficulty understanding IA rationale during mixed-initiative decision-making [2]. Mixed initiative decision-making is not only a requirement for effective operations in war-fighting environments, but also an inherent behavior of all Joint Cognitive Systems [3]. Therefore, operational environments demand not only autonomy and flexibility, but also collaborative interaction between system cognitive actors (human operators and IAs) to reach optimized performance. As a whole, the overall system should support operators' recognition of the state of the world, anticipation of the consequences of state changes in the world, and appropriate adaptation of system means and goals. This can aid in operators' general performance, as well as management of abnormal system-wide events [4, 5]. To facilitate this support, an operator's display should be designed with resilient operations in mind [6]—that is, buffering, to absorb the information processing deficits of a human cognition; flexibility, rather than brittleness, to adapt to dynamic events; and appropriate trade-off mechanisms that solve conflicting goal and/or competing resource allocations.

Additionally, collaborative work between IAs and human operators in a system presupposes a priori roles for effective organizational automation. Generally, these roles may exist along a continuum of supervisory control [7, 8]. While both IAs and humans process, reason, and communicate, such processing must be explicit to the human in order to avoid confusion, instill trust, and structure action [9, 10]. Thus, if an operator is to make informed decisions, a system display must make explicit what the IA knows, does not know, reasons, and projects about its operation context and its goals.

1.1 Situation Awareness in Mixed Initiative Decision-Making

The construct of situation awareness (SA) formalizes human interaction within a given context [11, 12]. Whether conceptualized as a processes or product, SA explicates situational human cognition for decision-making, as it represents an operator's awareness of the immediate situation, comprehension of the situation, and prediction of future possibilities [13, 14]. The most commonly relied upon model parses SA into three levels [11]: perception of the situation elements, comprehension of these elements, and projection as it relates to the perceiver and situation elements in the future. IAs possess a similar computational ability for sensing, reasoning, and projecting about their environment.

In this sense, both humans and IAs are data driven and concept driven. That is, human-cognition and computational-cognition are each concerned with a means and an end to accomplish a purpose. For example, both perceive or sense the environment and are able to utilize data in addition to planning and modifying that data to accomplish a goal. Separate, both are entities whose performance may improve through an external intervention. However, a paradigm that accounts for collaborative and coordinated human-agent interaction would allow for a unified cognitive system that integrates human and IA cognitive processes and outcomes [15, 16].

1.2 Transparency and Supervisory Control of Intelligent Agents

In addition to information sharing between an operator and an IA, as well as coordination of both of their respective activities, it has been suggested that collaborative work must acknowledge that each part of a system possesses partial and overlapping information relevant to the fulfillment of the overall system purpose [16]. To benefit from that information, a collaborative work system must provide a means for a transparent field of view of each agent's unique perspective. In this regard, increasing *transparency* in IA interfaces can improve operator performance [9], provided it gives an understandable representation of the mission environment and constraints, and the IA's knowledge, intent, and limitations [17, 18].

Not only will IA transparency increase operator SA by giving insight into the IA's current action and intent, its relevant knowledge of the state of the world and situational constraints, but it will additionally engender trust between the IA and the operator, who must rely on the IA's reasoning and projections to make decisions [9, 10]. Transparency specifically facilitates appropriate calibration of trust for the operator. Such calibrated trust should lead to appropriate reliance on the IA [19]. As opposed to under-reliance (IA disuse) or over-reliance (IA misuse), which impede overall system effectiveness [20], establishing appropriate reliance can increase overall performance in the human-machine system [19].

1.3 Situation Awareness-Based Agent Transparency

In an effort to meet the above needs and guide the design of transparency in IA, Chen et al. [10] proposed the Situation awareness-based Agent Transparency (SAT) model. By applying underlying theoretical assumptions inherent to the understanding of both SA and agent transparency, this model can facilitate effective mixed-initiative decision-making (Fig. 1). The model functions as a corollary to the three levels of individual SA [11], yet is particularly relevant to the domain of human-IA teaming.

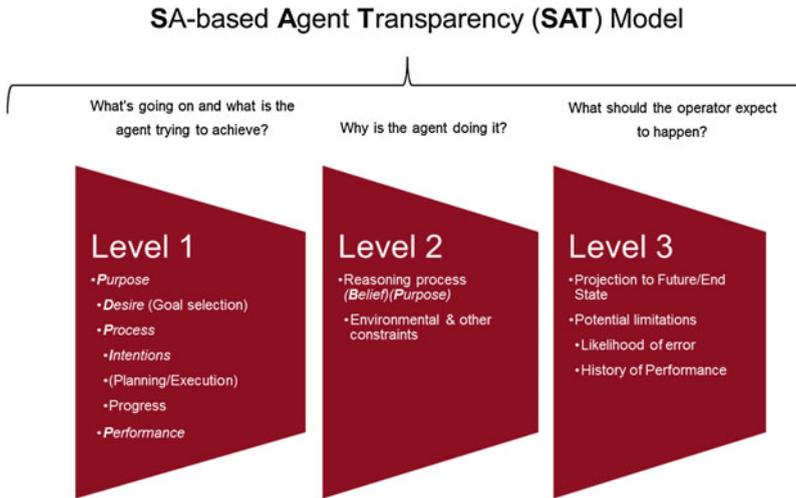


Fig. 1 SA-based agent transparency model diagram [10]

The SAT model provides a useful theoretical framework to guide design efforts of requisite IA display elements that support the operator's SA and facilitating appropriate trust calibration. While recognizing the danger in assigning human attributes to a computational agent, it is helpful to borrow such terms for clarity. Display requirements for human-supervisory control with IA thus correspond to the three levels of SA in humans. Each level seeks to provide the answer to three implicit questions of operator:

1. What is the agent trying to achieve?
2. Why is the agent doing it?
3. What should the operator expect to happen?

Implicit in the cognitive and computational process captured in Level 3 of the SAT model is the notion of uncertainty, specifically the fact that no future event can be absolutely known. Although an IA can make sense of the world, it does not necessarily know all parameters that may affect its actions. It is important for the IA to communicate this uncertainty as part of its interaction with the human for collaborative planning and decision-making. Thus, the IA must share its uncertainty concerning its reasoning and projections with the operator. For example, in order to make a suggestion, the IA often must "fill in the blank" regarding missing information—the IA must make an assumption. A transparent IA must then communicate the nature of that uncertainty and the assumption made by the IA to the operator.

This model is not solely a human model, nor is it a model only for the IA. Instead, it relates the IA cognitive process and products back to the human's supervisory purview. Level 1 communicates the IA's desires and intentions [21] as

they relate to its environmental, operational, and organizational context. As part of a goal-directed team, the IA examines its environment for data needed to algorithmically reason what actions are needed to achieve optimum system performance; such communicated reasoning is Level 2. Finally, the IA makes projection of the dynamic nature of the situation; Level 3 information provides the operator this insight.

1.4 Design of Transparency Displays for Heterogeneous UxV Management

The application of the above theoretical positions is particularly important in the study of multi-unmanned vehicle (UxV) management, where mixed-initiative decision making is integral to mission success. Increasingly, research is focusing on the development of IAs that can work with operators to manage teams of UxVs [1, 22]. One of those efforts is the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT) project currently funded by the U. S. Department of Defense's Autonomy Research Pilot Initiative [23, 24]. IMPACT is investigating issues associated with human-machine interaction in military contexts [24], and flexible "play-calling," such as that done in football [25, 26]. Such "play-calling," whereby a person chooses from a set of options or plans in a "playbook," can be applied in many warfighting contexts where warfighters are frequently required to make diplomatic decisions based on a limited set of options. It may be particularly useful in UxV management [25].

As part of this effort, and to explore the SAT model's utility for UxV management, the SAT model served as a guideline in the design of two separate IA interfaces evaluated in two consecutive studies (Figs. 2 and 3, referred to as Interface 1 and Interface 2, respectively). These interfaces were adapted from the U. S. Air Force Research Laboratory-developed IMPACT/Fusion interface [22, 27], and were further developed to convey three different conditions of SAT for each interface. These conditions and descriptions of their corresponding graphical displays are given in Table 1.

2 Study Design and Implementation

We examined the above interfaces separately in a pair of consecutive studies [28, 29]. Both studies sought to test a series of predictions regarding whether the aforementioned implementation of the SAT model was successful in facilitating UxV management. Specifically, we wanted to examine the impact of information sharing on several performance parameters critical to the success of multi-UxV management. For example, while additional transparency can improve performance



Fig. 2 Interface 1: displaying information according to the SAT model

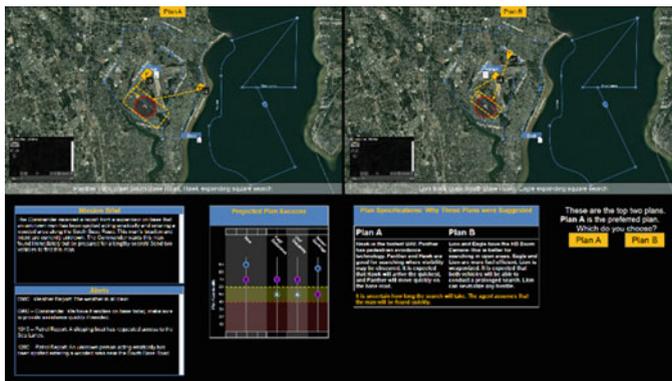


Fig. 3 Interface 2: displaying information according to the SAT model

Table 1 Manipulation of agent transparency using SAT levels to form experiment conditions. Each condition contains the components introduced in the prior condition

Interface 1		Interface 2	
SAT level	Display components	SAT level	Display components
L1	Map icons, plan details icon, and path show basic information	L1 + 2	Map icons, path, line graph, and text show basic information
L1 + 2	Pie graph and text add reasoning information to display	L1 + 2 + 3	Sliding points on line graph and extra text add reasoning and projection
L1 + 2 + 3	Opacity of sprocket pie graph varied and extra text add projections including uncertainty	L1 + 2 + 3 + U	Opacity of map icons and graph points varied, and extra text add assumptions and uncertainty

Table 2 Performance criteria according to IA and operator choice of plans

Performance criterion	Correct plan	IA suggestion	Operator choice
Proper IA use	A	A	A
Correct IA rejection	B	A	B

[10], it is important to consider the impact that this extra information has on response time and increased workload [30]. Furthermore, the usability of the interface may affect trust in machines [31]. Finally, as stated above, it has been suggested that additional transparency can improve trust on behalf of the human counterpart [9, 10].

To test these theoretical positions, we designed two studies, each with three conditions of transparency (see Table 1). During the corresponding experiments, participants took on the role of a UxV system operator whose task was to monitor and direct vehicles to carry out missions given to them by a simulated commander. Operators managed a team of six unmanned vehicles (UxVs): two unmanned aerial vehicles (UAVs), two unmanned ground vehicles (UGVs) and two unmanned surface vehicles (USVs), as well as an IA, which communicated plan options for completing the mission. To complete missions, operators needed to interpret their commander’s intent, understand vehicle and environmental constraints, and ultimately decide whether to follow the IA’s play-calling recommendation. The IA always suggested two options: Plan A as the most viable plan (which was its primary recommendation), and Plan B as the back-up plan. For 3 out of every 8 events, the IA’s recommendation was incorrect due to information it did not have access to—updated commander’s intent or other intelligence.

During each of these decisions, operators’ performance (based on the criteria in Table 2), and response time were monitored by the simulation. After each block of events, we surveyed participants for information including their perceived workload, perceived interface usability, and their trust in the IA.

2.1 Study 1: Interface 1

Results from study 1 [28] indicated that proper IA use and correct rejection were both significantly greater when participants were presented with SAT L1 + 2 + 3 and L1 + 2 compared to L1. The greatest rates of proper IA use (when the IA’s recommendation was correct) and correct rejection (when the IA’s recommendation was incorrect) were found in L1 + 2 + 3, suggesting that operators were more likely to make correct decisions when presented with all three levels of SAT information. We found no significant differences for response time or workload, indicating that operators did not take longer to complete each decision nor did operators experience more effort as the amount of information to support agent transparency increased.

We analyzed operator trust in the IA after the first block of interactions, and examined it across two contexts: the IA's analysis of the information, and the IA's ability to suggest and make decisions. There were no significant differences across SAT level for trust in the IA's ability to analyze information. However, we found that operator's trust in the IA's ability to suggest and make decisions significantly increased as transparency increased. Specifically, participants felt the IA made decisions that were more accurate when presented with L1 + 2 + 3 as compared to L1 + 2 or L1. We also found a significant effect of SAT level on the perceived usability of the IA, where the IA was perceived to be the most usable when presented with L1 + 2 + 3.

While this study differentiated basic information, reasoning, and future projections according to the SAT model, we only examined uncertainty as part of SAT Level 3 information and not on its own. Thus, the role of uncertainty in affecting operator decision making remained unclear. Study 2 filled this gap by setting up different conditions whereby the final condition parsed out uncertainty from other Level 3 information (see Table 1).

2.2 *Study 2: Interface 2*

Results from study 2 [29] indicated that proper IA use and correct rejection were both significantly greater when SAT L1 + 2 + 3 + U was presented compared to L1 + 2. The greatest rates of proper IA use and correct rejection were found with L1 + 2 + 3 + U, suggesting that operators were more likely to make correct decisions when they were presented with all three levels of transparency, as well as uncertainty. As was the case in study 1, no significant difference was found for workload, indicating that operators did not experience more effort as the amount of information to support agent transparency increased. However, unlike study 1, there was a significant difference in response time between L1 + 2 and L1 + 2 + 3 + U, with L1 + 2 + 3 + U taking the longest for participants to complete. This was not unexpected, as an increase in information on the display should naturally take longer to process.

Contrary to study 1, in which we only analyzed trust after a single interaction with the interface, for study 2 we analyzed operator trust as it developed over time while also controlling for the effect of pre-existing implicit associations [32]. There was a significant difference across SAT level for trust in both the IA's ability to analyze information and the IA's ability to suggest and make decisions. Specifically, participants trusted the IA's ability to analyze information most when presented with L1 + 2 + 3 + U, while they trusted the IA's ability to suggest decisions most when presented with L1 + 2 + 3. We also found a significant effect of SAT level on the perceived usability of the IA, where the IA was perceived to be the most usable when displaying L1 + 2 + 3 and the least usable when displaying L1 + 2 + 3 + U. This perception of usability is somewhat consistent with the participants' trust in the IA's ability to make decisions, where their trust and

perceived usability peaked at L1 + 2 + 3 and tapered off when uncertainty was added to the interface. This finding adds further support to the idea that usability impacts trust [31]. It also raises several questions about the display of uncertainty [33], which will be discussed next.

3 Discussion

Overall, we found evidence supporting the use of the SAT model to improve operator performance, increase trust in the IA, and increase perceived usability of the system, while minimizing potential costs of workload. Displaying SAT L1 + 2 + 3 information provided the most benefits to operators' trust and perception of the agent's usability, while displaying L1 + 2 + 3 + U provided the most benefits to operators' performance. Due to these findings, we recommend that similar automated decision aid systems incorporate information displays that provide the operator with both information regarding the reasoning of each decision provided, as well as displays of possible future states and sources of potential uncertainties that might affect their decisions. Our results suggest a number of aspects that designers of human-machine systems should consider.

First, how can we utilize SAT-based displays to improve decision-making performance while minimizing the impact on response time? It makes sense that response time may increase when more information is presented. Increased response time may not always be problematic, but in time-critical tasks, milliseconds may make the difference between success and failure. In such contexts, it is important to design interfaces that communicate vital information to the user while minimizing the amount of processing required to make a decision. If we aim to make truly flexible machines that can adapt to the environment [6], we must also consider how this flexibility applies to the display of SAT-based information.

Next, how can we best display uncertainty in a way that is both useful and usable to the operator? Such optimization of the interface may have effects on not just usability, but trust and performance, as well. Our interfaces displayed uncertainty both graphically and in text, but we did not statistically differentiate the usability of each of these components. Further analysis of such component parts may yield information and best practices about the display of uncertainty. For example, it is possible that intuitive graphical representations will be perceived as more usable—and may even result in lower processing time—than textual representations. Furthermore, trust and perceived usability may change when the IA presents its uncertainty in different ways. For example, reporting percentages of certainty (e.g. “80 % probable”) may lead to drastically different perceptions and outcomes in the operator than more ambiguous graphical representations, and these potential differences must be considered [34].

Finally, are the results we found here generalizable? We argue that the outcomes examined here depend on the task and the context. This position is supported by prior research and theoretical discussions positing that both task and environment

influence overall human-machine performance [35]. The studies presented in this paper examined the management of UxVs in a military environment. As such, our findings are most applicable to similar mixed initiative decision-making tasks. What remains to be seen is the validity of these parameters in entirely different contexts.

Future studies should examine the display of SAT-based information in new contexts, and thus refine our understanding of the usefulness of agent transparency in human-machine interaction. Furthermore, future studies should more thoroughly examine the role of uncertainty as a key to achieving appropriate levels of transparency. While the display of uncertainty may have tradeoffs, it should not be eliminated from displays [34]. It is wholly necessary, as are the other facets of transparency, to the successful performance of overall human-machine systems.

Acknowledgments This research was supported by the U.S. Department of Defense Autonomy Research Pilot Initiative, under the Intelligent Multi-UxV Planner With Adaptive Collaborative/Control Technologies (IMPACT) project. We wish to thank Joseph Mercado, Katelyn Procci, Isacc Yi, Erica Valiente, Shan Lakhmani, and Jonathan Harris for their contribution to this project. We would also like to thank Gloria Calhoun and Mark Draper for their input.

References

1. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multi-robot control: a review of human factors issues. *IEEE Trans. Hum. Mach. Syst.* **44**, 13–29 (2014)
2. Linegang, M.P., Stoner, H.A., Patterson, M.J., Seppelt, B.D., Hoffman, J.D., Crittendon, Z.B., Lee, J.D.: Human-automation collaboration in dynamic mission planning: a challenge requiring an ecological approach. In: *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, pp. 2482–2486. SAGE Publications, California (2006)
3. Hollnagel, E., Woods, D.D.: *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. CRC Press, Boca Raton (2005)
4. Boy, G.A.: *Cognitive Function Analysis*. Ablex Publishing Corporation, Stamford (1998)
5. Kasdaglis, N., Newton, O., Lakhmani, S.: system state awareness: a human centered design approach to awareness in a complex world. In: *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting*, pp. 305–309, SAGE Publications, California (2014)
6. Hollnagel, E., Woods, D., Leveson, N.: *Resilience Engineering: Concepts and Precepts*. Ashgate, Surrey (2006)
7. Boy, G.: Theories of human cognition: to better understand the co-adaptation of people and technology. In: Kiel, LD (ed.) *Knowledge Management, Organizational Intelligence and Learning, and Complexity*, vol. 3, Developed under the Auspices of the UNESCO, pp. 204–238. Eolss Publishers Co Ltd, Oxford (2009)
8. Parasuraman, R., Cosenzo, K.A., De, Visser, E.: Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. *Mil. Psychol.* **21**, 270–297 (2009)
9. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* **46**, 50–80 (2004)
10. Chen, J.Y.C., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.: Situation awareness-based agent transparency (ARL-TR-6905). Technical Report, US Army Research Laboratory, Aberdeen Proving Ground (2014)

11. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. *Hum. Factors* **37**, 32–64 (1995)
12. Smith, K., Hancock, P.: Situation awareness is adaptive, externally directed consciousness. *Hum. Factors* **37**, 137–148 (1995)
13. Endsley, M.R., Jones, D.G.: *Designing for Situation Awareness: An Approach to User-Centered Design*. CRC Press, Boca Raton (2012)
14. Stanton, N.A., Chambers, P.R., Piggott, J.: Situational awareness and safety. *Saf. Sci.* **39**, 189–204 (2001)
15. Woods, D.D.: Cognitive technologies: the design of joint human-machine cognitive systems. *AI Mag.* **6**, 86–92 (1985)
16. Hollnagel, E., Woods, D.D.: *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. CRC Press, Boca Raton (2006)
17. Cook, M.B., Smallman, H.S.: Human factors of the confirmation bias in intelligence analysis: decision support from graphical evidence landscapes. *Hum. Factors* **50**, 745–754 (2008)
18. Neyedli, H.F., Hollands, J.G., Jamieson, G.A.: Beyond identity incorporating system reliability information into an automated combat identification system. *Hum. Factors* **53**, 338–355 (2011)
19. Gao, J., Lee, J.: Extending the decision field theory to model operator’s reliance on automation in supervisory control systems. *IEEE Trans. Syst. Man Cybern. A. Syst. Humans* **36**, 943–959 (2006)
20. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse. *Hum. Factors* **39**, 230–253 (1997)
21. Bratman, M.: *Intention, Plans, and Practical reason*. CSLI Publications, Stanford (1987)
22. Behymer, K.J., Mersch, E.M., Ruff, H.A., Calhoun, G.L., Spriggs, S.E.: unmanned vehicle plan comparison visualization for effective human-autonomy teaming. In: *Proceedings of 6th international conference on ahfe and the affiliated conference*, pp. 1022–1029. Elsevier B. V, Netherlands (2015)
23. U.S. Department of Defense—Research & Engineering Enterprise. *Autonomy Research Pilot Initiative*. <http://www.acq.osd.mil/chieftechologist/arpi.html>
24. Draper, M.: *Realizing autonomy via intelligent adaptive hybrid control: adaptable autonomy for achieving UxV RSTA team decision superiority—year 1 report*. US Air Force Research Laboratory, Dayton (in press)
25. Fern, L., Shively, R.J.: A comparison of varying levels of automation on the supervisory control of multiple UASs. In: *Proceedings of AUVSIs Unmanned Systems North America*, pp. 10–13. Curran Associates Inc., New York (2009)
26. Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. *Hum. Factors* **49**, 57–75 (2007)
27. Rowe, A., Spriggs, S., Hooper, D.: Fusion: a framework for human interaction with flexible-adaptive automation across multiple unmanned systems. In: *Proceedings of 18th symposium on aviation psychology*, pp. 464–469. Curran Associates Inc., New York (2015)
28. Mercado, J., Rupp, M., Chen, J., Barber, D., Procci, K., Barnes, M.: Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors* (in press)
29. Stowers, K., Chen, J.Y.C., Kasdaglis, N., Newton, O., Rupp, M., Barnes, M.: Effects of situation awareness-based agent transparency information on human agent teaming for multi-UxV management (in press)
30. Lyons, J.B., Havig, P.R.: Transparency in a human-machine context: approaches for fostering shared awareness/intent. In: Shumaker, R., Lackey, S. (eds.) *Virtual, Augmented and Mixed Reality: Designing and Developing Virtual and Augmented Environments*, pp. 181–190. Springer, Berlin (2014)
31. Wang, L., Jamieson, G.A., Hollands, J.G.: Trust and reliance on an automated combat identification system. *Hum. Factors* **51**, 281–291 (2009)
32. Merritt, S.M., Heimbaugh, H., LaChapell, J., Lee, D.: I trust it, but i don’t know why effects of implicit attitudes toward automation on trust in an automated system. *Hum. Factors* **55**, 520–534 (2012)

33. Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., Chen, J.: Intelligent agent transparency: the design and evaluation of an interface to facilitate human and artificial agent collaboration. In: Proceedings of the Human Factors and Ergonomics 60th Annual Meeting (submitted)
34. Endsley, Mica R.: Designing for Situation Awareness: An Approach to User-Centered Design. CRC Press, Boca Raton (2011)
35. Stowers, K., Oglesby, J., Leyva, K., Iwig, C., Shimono, M., Hughes, A., Salas, E.: A framework to guide the assessment of human-machine systems. Human Factors (submitted)

Displaying Information to Support Transparency for Autonomous Platforms

Anthony R. Selkowitz, Cintya A. Larios, Shan G. Lakhmani
and Jessie Y.C. Chen

Abstract The purpose of this paper is to summarize display design techniques that are best suited for displaying information to support transparency of communication in autonomous systems interfaces. The principles include Ecological Interface Design, integrated displays, and pre-attentive cuing. Examples of displays from two recent experiments investigating how transparency affects operator trust, situational awareness, and workload, are provided throughout the paper as an application of these techniques. Specifically, these interfaces were formatted using the Situation awareness-based Agent Transparency model as a method of formatting the information in displays for an autonomous robot—the Autonomous Squad Member (ASM). Overall, these methods were useful in creating usable interfaces for the ASM display.

Keywords Agent-Transparency · Human-Robot · Interaction · Ecological interface design · Cognitive engineering

1 Introduction

With the rise of autonomous agents on and off the battlefield, they are taking on more tasks. As the number of tasks increases, autonomous agents' complexity and independence also increases. If there is an increase in the independence of autonomous

A.R. Selkowitz (✉) · J.Y.C. Chen
U.S. Army Research Laboratory, Aberdeen Proving Ground, Adelphi, MD, USA
e-mail: anthony.r.selkowitz.ctr@mail.mil

J.Y.C. Chen
e-mail: yun-sheng.c.chen.civ@mail.mil

C.A. Larios · S.G. Lakhmani
Institute for Simulation and Training, Orlando, FL, USA
e-mail: clarios@ist.ucf.edu

S.G. Lakhmani
e-mail: slakhman@ist.ucf.edu

agents, then there is also an increase in the necessity that their operator, supervisor, or teammate have an accurate understanding of autonomous agent's actions, environment, reasoning, projections, and uncertainty calculations. Making this information accessible to the user is a concept known as agent transparency [1]. Providing the operator with information to support agent transparency supports the operator's ability to maintain an accurate understanding of their environment and the autonomous agent. This allows the operator to accurately calibrate their trust in the autonomous agent. Since providing information to support agent transparency is critical in human-agent interactions [2], the display of this information mediates the operator's ability to interpret the information [3]. A poorly designed display could obfuscate the meaning behind the information and render the information useless [4]. The purpose of this article is to examine display design techniques that are best suited for displaying information to support agent transparency in autonomous agents, as well as two studies that implemented these techniques. First, we will review the Situation awareness-based Agent Transparency model (SAT model), which is a framework to serve as a basis to guide the format of the information provided by the autonomous agent to promote agent transparency. Second, we review the design principles of Ecological Interface Design, pre-attentive cuing, and integrated displays, and their application to the SAT model-based Autonomous Squad Member (ASM) interface. Lastly, we review two studies that demonstrate the benefit of including SAT model-based information.

2 SAT Model

The Situation awareness-based Agent Transparency model (SAT model) is a model of transparency that supports the operator's situation awareness [5] of the autonomous agent and its environment [1]. The SAT model is founded on the principles of situation awareness [5], the Beliefs, Desires, and Intentions Agent framework [6] and the 3 P's (Purpose, Process, and Performance) proposed to increase system transparency by Lee and See [16]. Endsley [5] proposed that one's knowledge of the state of a dynamic environment comprises situational awareness (SA). Endsley [5] defined SA in three levels: (1) the perception of basic elements of the environment, (2) comprehension of the current situation, and (3) projection of the future status. The creators of the SAT model theorize that by supporting the operator's SA of the autonomous agent's actions, environment, reasoning, projections, and the uncertainty associated with these elements, the operator can properly calibrate trust in the autonomous agent and reduce workload associated with monitoring the autonomous agent [1]. As seen in Fig. 1 below, the SAT model comprises three levels of transparency that are meant to support different aspects of the operator's SA of the autonomous agent.

Level 1 of the SAT model supports the operator's understanding of the autonomous agent's goals and its perception of the status of its environment. For example, an autonomous agent could provide the operator with its current path



Fig. 1 The SAT model (Chen et al., 2014)

or resource levels. In Level 2, the autonomous agent displays information to support the operator’s understanding of why the agent is performing its actions and reasoning behind its projections. For example, the autonomous agent may alert the operator that it has chosen its new route because fuel is a priority and the new route is more fuel-efficient. For Level 3 of the SAT model, the autonomous agent supports the operator’s ability to make projections based on the autonomous agent’s current actions and reasoning. For example, the autonomous agent may indicate that its new route projects to save three units of fuel, but it is uncertain about the amount of time the new route will take.

The following sections will review the principles used to design two displays from recent studies examining the effects of presenting SAT model-based information on operator trust in the autonomous agent, situation awareness, and perceived workload. Study 1 examined the effects of displaying SAT model-based information while the operator monitored the ASM completing a route. Study 2 implemented SAT model-based information in a version of the scenario from Study 1 in which the operator had to monitor multiple displays. In Study 2, there was an emphasis on the operator’s ability to gain information at a glance.

3 Ecological Interface Design

Ecological Interface Design (EID) is a design methodology based on the principles of ecological psychology [7]. EID was developed by Vicente and Rasmussen [8] as a means to support the operator’s direct perception of the system they are operating by enabling the “transparency” of the interface. The primary function of EID is to promote the visualization of the work domain that the display represents. Traditionally, in EID, this is accomplished through performing an Abstraction Hierarchy detailing all aspects and functions that the system interacts with and the effects that the system has on the environments [8]. Bennett and Flach [9] recommend not only designing the interface to support the display of the environmental domain constraints, but also supporting the knowledge constraints

Fig. 2 Metaphorical representation of fuel in ASM Study 1



Fig. 3 Metaphorical representation of signal strength from ASM Study 2



imposed-on the operator using the system. This principle aligns well with the display of information to support agent transparency because knowing the information constraints that the operator receives is crucial to being able to interpret the information displayed. For example, if the autonomous agent displays that energy is a priority in its reasoning process and it does so by displaying an icon that represents energy (Fig. 2), then the operator must be able to interpret correctly that icon and its meaning. This leads to the support of metaphor-based representations in an interface.

A metaphor-based representation describes the use of a symbol to represent an abstract concept in the interface [9]. For instance, instead of displaying an actual picture of an antenna or information system to represent signal strength, the display may draw upon the user’s prior knowledge and utilize iconography that communicates the same information with symbology that is familiar to the user. For an example, please see the Fig. 3.

The use of metaphorical based representations is especially suited to the displays promoting agent transparency, because much of the information used to support agent transparency is abstract in nature. Representing abstract information with the use of metaphor-based iconography allows the user to process the information using information and relationships already known to the user. However, the concepts needed to be displayed may be too complex, do not have an easy to represent symbol to be used, or simply space does not allow a symbol to be distinguished. In this instance, a propositional based icon is quite suitable for autonomous agent interface design (Fig. 4).



Fig. 4 Propositional Iconography from ASM Study 2

Propositional-based representations includes icons using alphanumeric symbology to represent the different information that the system is presenting to the user [9]. For instance, a propositional-based symbol was used to represent the ASM’s current action in Study 2. In Study 2, the ASM’s current actions were complex and the symbology needed to represent different maneuvers that the ASM would perform. A maneuver such as Duck and Cover is not easily summed up in a single symbol, thus the propositional representation of the letter “D” was used to represent this information. More examples of the propositional symbology can be seen in Fig. 7, below. In addition to the principles of Ecological Interface Design, another display design principle well suited to the display of information is the use of integrated displays.

4 Integrated Displays

The use of integrated displays is based on the proximity compatibility principle developed by Wickens and Carswell [10]. The proximity compatibility principle proposes that proximally displaying pieces of information that are conceptually related to one another can enhance the operator’s ability to process the information [10]. Integrated displays are especially relevant to displays to support agent transparency, because often the displays have limited space requirements, such as automated global positioning systems. In other words, integrated displays not only reduce the operator’s cognitive processing in using the display, but also aid in designing displays with size constraints. Integrated displays will often use different perceptual properties to distinguish different types of information.

Shown in Fig. 5 is the icon representing the ASM in the integrated display from Study 2. Within the icon is the ASM’s current action. The arrow in front of it represents the ASM’s current heading and its position on the map represents the ASM’s current location. This display integrates the ASM’s position, heading, and current action. This allows the operator to gain multiple pieces of information about the ASM in one location.

Fig. 5 Integrated display from ASM Study 2



5 Pre-Attentive Cues and Ecological Psychology

Using displays that uphold direct perception, humans are able to effortlessly tune into visual information based on characteristics that are visually salient [7] allowing them to detect differences in stimuli based on these features. In particular, EID forms spatial metaphors or emergent features that show the interaction between familiar concepts and actions [6, 11]. Displays that are impoverished in natural visual cues may cause slow response times and errors for operators who need to monitor systems for long periods of time. Thus, displays must reinforce natural perceptual competences by having simple element and symbology arrangements so that the operator can process critical system relationships. Certain features, such as color, orientation, or shape, have been shown to be able to be perceived for a short time period without the effort of serial search [12]. This type of visualization hones into pre-attentive cue-based processing (pre-attentive cues) [13]. Examples of pre-attentive cues examined in unmanned automation and automated decision aids include opacity (i.e. shading), graphs, color, shape, lines, icon size, and spatial synchronization (i.e. egocentrism view). The use of pre-attentive cuing was extensive in both ASM Study 1 and Study 2. This was done to promote the ease of processing complex information in the operator. For example, as seen in Fig. 6, in Study 2, the concepts of size and opacity were used to denote the ASM's uncertainty about the location of the obstacle that was being encountered.

In Study 1, the concept of shape and color were used to dual code the differences between types of hazards and areas that provided benefits. For instance, if the area had terrain that provided an advantage to the squad, then it was displayed using a green triangle border (Fig. 7), but if the area was an ambush hazard, then it was displayed using a red circle based border as seen below (Fig. 8). The color indicated

Fig. 6 In the left image, the ASM is indicating that it is certain about the location of the obstacle, and in the image on the right, the ASM is indicating that it is uncertain about the location of the obstacle

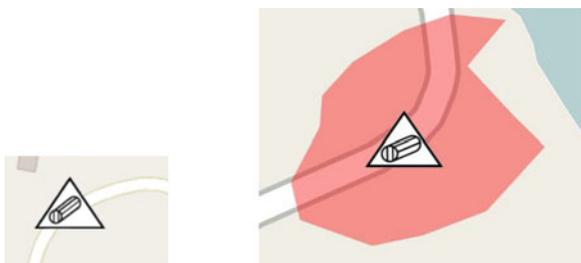


Fig. 7 Study 1 Beneficial Terrain icon



Fig. 8 Study 1 Ambush Icon

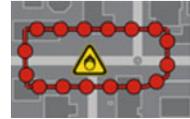


Fig. 9 Certain and uncertain projected signal loss



Fig. 10 All three bars indicated that the ASM was certain about severe signal loss in the near future



whether it was a hazard or beneficial area and the border indicated what type of hazard or beneficial area it was (Figs. 9 and 10).

In Study 2, the principle of opacity was used to denote uncertainty. In this case, it was the uncertainty that the ASM expressed when it was unsure of the projected severity of loss it would experience due to the obstacle it encountered. The lighter colored bars indicated that the ASM is uncertain whether it will experience that level of resource loss. In the interface, one bar indicated that the ASM projected that it would experience low resource loss, two bars indicated that the ASM projected that it would experience moderate resource loss, and three bars indicated that the ASM projected that it would experience severe resource loss. The dark colored bars were used to denote the ASM’s certainty of that level of resource loss.

5.1 SAT Model Based Studies

A recent study has shown the benefits of incorporating SAT model based information. Mercado and colleagues (in press) implemented SAT model-based information in the interface for an autonomous agent that coordinated and supervised several subordinate unmanned autonomous agents. The participants’ task was to choose between competing plans proposed by the autonomous agent supervising other autonomous agents. It was shown that when participants were presented with SAT model-based information, from all three SAT levels, their performance increased and they had more appropriate trust calibration than when presented with information from only Level 1 or Levels 1 and 2. In addition, when users were presented with more levels of SAT-based information, their workload did not increase. This is

important because it demonstrates that when information to support transparency is implemented using design techniques that do not overload the operator, the information can aid the operator monitoring the system in maintaining situation awareness.

Two further studies, conducted by the present authors and reviewed in this article, also found similar results when implementing SAT model-based information in the user interface for the ASM robot [14, 15]. The ASM is an autonomous ground-based unmanned vehicle, which follows its fellow squad members and provides them with information about its status. The squad members rely upon the ASM to inform them of its status. Additionally, the ASM acts as a robotic mule carrying the squad members' supplies. The ASM system is currently in development, so the studies reviewed in the present article were conducted with operators monitoring simulations of the ASM system in virtual environments.

5.2 Study 1

In Study 1, operators monitored the ASM as it completed a route. Along the route, the ASM had to reach numbered rally points in order. Although the goal was to take the shortest path possible, the ASM would deviate from the shortest possible route if it encountered an area with a hazard or an area that would provide benefit to its squad members. Examples of environments that would provide benefits include: an area that would provide cover or enhanced point of view, an area that would be easy to traverse, or an area that had good signal coverage. Periodically, the operator was asked questions to assess the operator's situation awareness of the ASM's resources and the ASM's route. Operators were presented with different levels of SAT model-based information according to the experimental condition to which they were assigned. An example of the interface, presenting the highest level of information, can be seen in Fig. 2. Operators with access to SAT Level 1 information were informed of the ASM's current heading and current resource levels. Operators with access to SAT Level 2 information had all of the information presented in SAT Level 1; in addition, they received information about the reasoning behind why the ASM would alter its heading. Operators who received SAT Level 3 information, had access to all of the previous levels of information, were informed about the ASM's uncertainty associated with its reasoning behind altering its route and in addition also received the ASM's predicted resource levels that were a result of it altering its route. In Fig. 11 the icons with the red borders represent hazards and the icons with the green borders represent beneficial areas. The blue line indicates the ASM's current route. The figure on the blue line represents the ASM's current position on its route. The circular green figure in the lower right hand corner of the interface represents the ASM's current resource status and the circular figure in the upper right hand corner represents the ASM's predicted resource usage. The difference in opacity of the hazards and beneficial areas represents the ASM's uncertainty of whether it will encounter that area. The greater the opacity represents



Fig. 11 Interface from ASM Study 1 displaying all 3 levels of SAT model-based information

that the ASM is certain that it will encounter the area and the lower opacity represents that it is uncertain. Overall, when compared to the other experimental conditions, it was shown that operators who received SAT Level 2 information but not Level 3 information had higher trust in the ASM, but not higher workload and did not rate the interface more highly in terms of usability. This supports the notion that by using the principles of Ecological Interface Design, pre-attentive cuing, and integrated displays, one can increase the information displayed on an interface, but not increase the workload encumbered by the operator. A detailed description of the study’s findings can be found in Selkowitz, Lakhmani, Chen, and Boyce [14].

5.3 Study 2

Although Study 2 was similar to Study 1, it expanded upon Study 1 by incorporating the aspects of dual tasking that would be present in the ASM’s operational environment. Study 2 had the operator take the role of a squad member whose function was to monitor the ASM and identify obstacles presented in the virtual environment. Similar to Study 1, the ASM and squad would proceed along a route to reach each rally point. During Study 2, in addition to monitoring the ASM, the operator monitored a second screen for obstacles that the squad and ASM would encounter. The ASM presented the operator with four different levels of information in a within-subjects experimental design. The operator monitored the ASM presenting Level 1, Level 1 + 2, Level 1 + 2+3 or Level 1 + 2+3 + uncertainty, SAT model-based information.

Level 1 incorporated SAT model Level 1 based information. This included the ASM’s current action, its understanding of the squad’s current action, the obstacle



Fig. 12 ASM Display from Study 2 with Level 1, 2, and 3 information without uncertainty



Fig. 13 ASM Display from Study 2 with Level 1, 2, and 3 information including uncertainty

in the environment that the squad and ASM were encountering, its current resource levels, its progress along the route, and the element of its environment that was currently influencing its actions. An example of this display is in Figs. 12 and 13, below.

Level 2 SAT model-based information detailed the ASM's reasoning behind its current action. The ASM's reasoning could be to conserve either energy, mechanical integrity, time, or signal strength. The ASM's current and previous actions determined which resources it would conserve. As shown in Figs.12 and 13. Level 2 information consisted of the Clock iconography with the green shield.

Level 3 SAT model-based information included the ASM's predictions of its projected time to the next rally point and the predicted loss of resources based on its current action, environment, and reasoning. For example, if the ASM and squad were encountering a sniper attack and the ASM decided that mechanical integrity was a priority, it may perform avoidance maneuvers to avoid the sniper fire, ensuring it would not receive damage. However, this course of action may cause the ASM to use extra energy during these maneuvers, which would result in future energy loss. In Fig. 3, below, the battery icon with red blocks next to it and the red box with the clock icon represented Level 3 SAT-model based information.

For Study 2, the display of the two types of Level 3 SAT model-based information, projections and displaying uncertainty, were separate conditions. In this condition (Level 1 + 2+3 + uncertainty), the ASM indicated its uncertainty about the location of obstacles and projections of resource expenditures. Examples of the display with uncertainty and without indicating uncertainty can be seen in Fig. 12 and Fig. 13 below

Preliminary results from Study 2 indicate that the additional levels of SAT model-based information did not have an impact on the operator's trust in the ASM, but Level 3 SAT model-based information did aid the operator in maintaining Level 3 situational awareness without increasing the operator's self-reported workload. A more detailed description of these results is available in Selkowitz, Lakhmani, Larios, and Chen (in press).

6 Conclusions

This article presented two recent displays to promote agent transparency in the Autonomous Squad Member system. Overall, the principles of display design discussed here proved to be useful in the development of the ASM displays. Future studies, and applications using autonomous agents, should consider the use of Ecological Interface Design, integrated displays, and pre-attentive cuing in their display design efforts.

One area that these display design principles, and the applications developed in the studies reviewed, would be applicable to is in autonomous automobiles. As manufacturers integrate more autonomous features in vehicles, it is imperative that the operators have an accurate understanding of the systems lest they fall into the perils of misuse and disuse of the system. Recent examples of autonomy implemented in vehicles include autonomous collision avoidance systems in which the vehicle will detect an upcoming collision and attempt to avoid it. Though this feature is only one-step to implementing full autonomy in automobiles, the

implementation of SAT model-based information, using the display principles outlined in this article, would still have merit. For instance, alerting the operator that the system is uncertain that the collision avoidance system is available due to uncertainty in its sensor system caused by weather conditions would allow the operator to know whether they can trust the system at that time. Displaying the current status of the system and its environment would constitute SAT model-Level 1 information, and displaying the reasoning behind its current status would constitute SAT model-Level 2 information. The above scenario is only one example of when one could implement SAT model-based information to improve the operator's understanding of autonomous systems.

Overall, this article presented the methodologies used to develop two novel interfaces for an autonomous agent, the ASM. The use of Ecological Interface Design, integrated displays, and pre-attentive cues produced usable interfaces. Study 1 results indicated an increase in operator trust, and Study 2 indicated an increase in situation awareness of the ASM, both without increasing the operator's perceived workload.

Acknowledgements This research was supported by U.S. Department of Defense's Autonomy Research Pilot Initiative. The authors wish to thank MaryAnne Fields, Daniel Barber, Erica Valiente, Andrew Watson, Kelvin Oie, and Susan Hill for their contribution to this project.

References

1. Chen, J.Y.C., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.: Situation awareness-based agent transparency. Technical Report: ARL-TR-6905, Aberdeen Proving Ground MD: US Army Research Laboratory (2014)
2. Lyons, J.B.: Being transparent about transparency: A model for human-robot interaction. In: AAAI Spring Symposium Series. Palo Alto, California (2013)
3. Kilgore, R., Voshell, M.: Increasing the transparency of unmanned systems: applications of ecological interface design. *Virtual, augmented and mixed reality. Applications of virtual and augmented reality*, pp. 378–389. Springer International Publishing, Heraklion (2014)
4. Lyons, J.B., Havig, P.R.: Transparency in a human-machine context: approaches for fostering shared awareness/intent. *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, pp. 181–190. Springer International Publishing, Heraklion (2014)
5. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. *Hum Factors: J Hum Factors Ergon Soc* **37**, 32–64 (1995)
6. Rao, A.S., Georgeff, M.P.: BDI agents: From theory to practice. *ICMAS. 95*, pp. 312–319. California, San Francisco (1995)
7. Gibson, J.J.: *The ecological approach to visual perception, classic edn.* Psychology Press, New York (2014)
8. Vicente, K.J., Rasmussen, J.: The ecology of human-machine systems II: mediating direct perception in complex work domains. *Ecol Psychol* **2**, 207–249 (1990)
9. Bennett, K.B., Flach, J.M.: *Display and interface design: subtle science, exact art.* CRC Press, Boca Raton (2011)

10. Wickens, C.D., Carswell, C.M.: The proximity compatibility principle: its psychological foundation and relevance to display design. *Hum Factors: J Hum Factors Ergon Soc* **37**, 473–494 (1995)
11. Bennett, K.B., Posey, S.M., Shattuck, L.G.: Ecological interface design for military command and control. *J Cognitive Eng and Decision Making* **2**, 349–385 (2008)
12. Kosara, R., Hauser, H., Gresh, D.L.: An interaction view on information visualization. In: *State-of-the-Art Report Proceedings of EUROGRAPHICS 2* pp. 123–137. Blackwell, Granada (2003)
13. Treisman, A.: Preattentive processing in vision. *Comp Vision, Graphics, and Image Process* **31**, 156–177 (1985)
14. Selkowitz, A.S., Lakhmani, S.G., Chen, J.Y., Boyce, M.W.: The Effects of Agent Transparency on Human Interaction with an Autonomous Robotic Agent. *Proceed Hum Factors Ergono Soc Annual Meeting*, 59, pp. 806–810. SAGE Publications, Thousand Oaks (2015)
15. Selkowitz, A.S., Lakhmani, S.G., Larios, C. N., Chen, J.Y.C.: Agent transparency and the autonomous squad member, to be presented at the 2016 annual meeting for the human factors and ergonomics society. Washington, DC (in press)
16. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum Factors: J Human Factors Ergon Soc* **46**, 50–80 (2004)

The Relevance of Theory to Human-Robot Teaming Research and Development

Grace Teo, Ryan Wohleber, Jinchao Lin
and Lauren Reinerman-Jones

Abstract In many disciplines and fields, theories help organize the body of knowledge in the field and provide direction for research. In turn, research findings contribute to theory building. The field of human-robot teaming (HRT) is a relatively new one, spanning only over the last two decades. Much of the research in this field has been driven by expediency rather than by theory, and relatively little effort has been invested in using HRT research to advance theory. As the field of HRT continues to expand rapidly, we find it increasingly necessary to relate theories to the research so that one can inform the other. As an initial effort, the current work will discuss and evaluate two broad research areas in human-robot teaming, and identify theories relevant to each area. The areas are (i) human-robot interfaces, and (ii) specific factors that enable teaming. In identifying the relevant theories for each area, we will describe how the theories were used and if findings supported the theories.

Keywords Theory · Human-robot teaming · Human-robot interface · Human-robot interaction · Robot capabilities

G. Teo (✉) · R. Wohleber · J. Lin · L. Reinerman-Jones
Institute for Simulation and Training, University of Central Florida,
Orlando, FL, USA
e-mail: gteo@ist.ucf.edu

R. Wohleber
e-mail: rwohleber@ist.ucf.edu

J. Lin
e-mail: jlin@ist.ucf.edu

L. Reinerman-Jones
e-mail: lreinerm@ist.ucf.edu

1 Introduction

As research in human-robot teaming (HRT) continues to advance, it is important to consider the role of theory in the field. In addition to the diverse contexts and domains in which HRT occurs, there is also the variability in robot design, communication mode, and functionality, which make it difficult to identify theories that can generalize across different human-robot teams. For this reason, many have questioned the relevance of theory in the field of human-robot teaming [1] and some researchers have resorted to using qualitative methods (e.g. [2]) and case-based usability research for direction in developing robots (e.g., [3]). While these approaches are essential for individual applications, they do less to consolidate and build on the field's knowledge base compared to theory-based approaches, which can guide subsequent research efforts. The present work begins with an overview of some of the applications of theory in HRT work and ends with a discussion of the role of theory in HRT and the implications of increasing that role.

2 Background

2.1 *The Role of Theory in Research*

Theory is a system of knowledge that depicts generalized relationships about how the world works, which enables predictions to be made. In most fields of study, theory provides a framework for organizing and guiding research. It frames observations and links a single study to a broader, common base of knowledge to which other researchers contribute. Gaps in understanding can be identified from such an organization of knowledge, which in turn, drive future research [4, 5].

There are different levels of theories corresponding to their level of abstraction. General theories are highly abstract and are almost unlimited in scope, while middle-range theories explain a less comprehensive set of phenomenon. Theories at a lower level of abstraction are generalized statements that account for a more restricted range of empirical observations with limited application [6]. It is possible that newer areas of research, such as human-robot teaming, will have a greater number of lower level theories. This is especially true since much of the research in HRT has been driven by expediency [7] and the need to understand the impact of the use and application of human-robot teams to increasingly more areas in our work and lives.

2.2 The Field of Human-Robot Teaming

There are several research areas related to the field of HRT, the closest being Human-Robot interaction (HRI). Domains of application for HRI research include the military, healthcare, manufacturing, and others. For instance, robots have been deployed in the military in bomb disposal as well as search and rescue missions. In healthcare, there are surgical assistant robots with high-definition 3D vision systems and dexterous robotic arms that help with surgeries [8], and service robots that allow caregivers to monitor and communicate with homebound patients [9]. Robots are also developed to facilitate rehabilitation regimens and provide therapy [10]. Finally, robots have been used in manufacturing because their ability to manipulate materials and objects with great speed and precision boosts productivity [11].

Apart from HRI, other areas related to HRT include automation, human-computer interaction (HCI), psychology, and neuroscience. Human-robot teaming research has benefited from research on automation (e.g., the unintended effects of inappropriate automation use), while HCI research has informed the design of interfaces that facilitate human-robot teaming (e.g., [12]). On the other hand, cognition, neuroscience, systems theory, control theory, and others have informed the design of architecture and mechanisms underlying human-robot teams (e.g., [13]).

2.3 Non-theoretical Development of Human-Robot Teaming

Non-theoretical techniques for developing robots and their interactions with humans are available and in common use. These techniques take a targeted approach typically informed by watching interactions and applying tried-and-true rules to the design of interfaces. For example, Clarkson and Arkin developed a heuristic evaluation (HE), or a set of guidelines for evaluators to use to identify human factors issues, for HRT. They based their HE on previous evaluations used for human-computer interface (HCI) and computer-supported cooperative work (CSCW) domains. They modified items from these earlier evaluations and added new items by brainstorming, consulting subject matter experts, and other informal techniques. They then validated the list by testing its performance, thus evaluating HRIs. In addition, Michaud et al. utilized focus groups and usability test scenarios in the development of a homecare tele-assistive mobile robot [14].

2.4 Applications of Theory in Human-Robot Teaming

Theory can be used to optimize or to enhance human-robot teaming. To optimize human-robot interactions, it is important to use theory to understand how humans

perceive, think, and act in relevant situations so that robots can be designed in such a way that increases the efficiency of interactions, while minimizing errors. This encompasses work on the interface for human-robot communication. It encompasses studies on the effects of robot appearance and form, input methods and modalities, displays, and adaptive interfaces and displays on human social behavior and cognitive processes.

Second, theory can be used to enhance human-robot teaming by augmenting the abilities of the robot. Specifically, theories of human perception, cognition, and action can be used to identify and implement advanced capabilities to facilitate human-robot teaming. This area relates to the social or teaming aspect of HRT, and includes capabilities and features of robots that specifically enables them to function as a teammate. Research is likely to model HRT after human teams as humans naturally team with other humans. Hence, to develop robots that can team, the robot would need to be more human-like. Studies also address research questions such as how to organize a human-robot team for various task and missions.

3 Theories Related to Human-Robot Teaming Research

We will review two broad areas of HRT research and will identify the relevant theories and discuss the role of theory within the area. The areas pertain to the research and development of (i) Human-Robot interfaces and (ii) capabilities that would help a robot team.

3.1 Theories Related to Research on Human-Robot Interfaces

Several theories that are used to optimize HRT are used in older fields such as HCI and their application is modified to address the unique challenges of HRT. Gillan and colleagues identified three areas that are of particular importance to HRT: situation awareness (SA), spatial cognition and mental maps, and task switching, which relates to executive functioning [15]. SA relates to how well the robot can support the human teammate's ability to perceive a situation, interpret it, and project a future state. Spatial cognition relates to the ability of a human teammate to ascertain the robots orientation and build a mental map of its environment. Task switching is important because the human teammate will need to perform relevant tasks and keep track of the robots state and location. Task switching ability is especially important when there are multiple robot teammates. A theoretical understanding of how humans switch tasks can help robot designers facilitate the calling of human teammate's attention appropriately.

The modalities utilized in the human-robot communication interface is also an important factor in HRT. Wickens' Multiple Resources Theory postulates that the human attentional capacity can be thought of as multiple "resource pools". These "pools" loosely correspond to encoding and response modalities, as well as stages of information processing. In a multi-tasking context, performance on the tasks undertaken simultaneously would be better if the tasks drew upon different resource pools than if the tasks required resources from the same resource pool [16]. The theory would predict that human-robot interfaces that enable tasks to be performed with various modalities would be more advantageous than interfaces with limited modalities. This has largely been supported by research, which found that interacting with robots that have multimodal interfaces can result in a reduction in human cognitive load especially when multiple tasks have to be performed concurrently (e.g., [17–21]).

Such predictions from theory have contributed to the recent focus on multimodal communications in robots, which encompasses visual displays, gestures, speech, non-speech audio and haptics [22]. Inclusion of speech and gesture detectors in interfaces can facilitate human robot teaming as these are modalities associated with natural language processing and do not require translational input devices like a keyboard or mouse. Such studies on robot interfaces with speech and gesture in teaming have motivated work in speech and gesture recognition and classification (e.g., [23]).

Some theories relate specifically to HRI research. The type of physical nature of a robot can influence the nature of the human-robot interaction. Humans may perceive robot behavior and interact with robots more effectively with those that have a human-like appearance than those more mechanical-looking. They may be more inclined to talk to a robot or smile at it if the robot has a human face or appears to understand speech [17]. In fact, research has shown that a robot's appearance affects the expectations humans have of its capabilities and function [24]. This finding suggests that humans are more likely to team with a robot that resembles a human. Furthermore, the more human-like the robot is in appearance, the more likely it would be accepted by the human as a teammate. However, as robot appearance becomes more and more human, the trend reverses, and the robot becomes repulsive because the appearance and functionality are incongruent. This is the "Uncanny Valley" in Mori's theory, named for the dip in the graph that plots level of acceptance against anthropomorphism.

Mori's Uncanny Valley theory was proposed directly from empirical research on human-robot interaction. The theory predicts that the robot's appearance can impact its practical application [25]. For instance, robots tending to trapped victims were perceived as "creepy" and not reassuring [26]. As a result of studies related to the Uncanny Valley theory, [27] argued that a robot with more human like appearance and behavior would be more acceptable and interact with people more effectively so long as the degree of robot anthropomorphism stops short of the uncanny valley. The quality of HRI and by extension performance in HRT, even with the most efficient and easy-to-use robot, can be jeopardized if the human teammate dislikes, distrusts, doubts, or resents the robot [28]. The United Theory of Acceptance and

Use of Technology (UTAUT) combines many of the above themes into a comprehensive look at factors that bear on acceptance. UTAUT identifies four factors that contribute to technology acceptance: effort expectancy, performance expectancy, social influence, and facilitating conditions [29]. The first two factors refer to traditional HCI considerations. Effort expectancy refers to the ease of use or usability of the robotic system and performance expectancy refers to the benefits of working with the robot. Social influence involves the approval of the human teammate's peers on his use of the robot teammate in a given situation. Finally, facilitating conditions refers to the extent that the human teammate believes that the organizational and technological infrastructure is in place to facilitate their use of the robot teammate.

3.2 Theories Related to Research on Characteristics that Help Robots Team

Human-robot teaming research departs from HCI and HRI research in that HRT seeks to develop robots with which humans can team. This necessarily entails having a human teammate collaborate with the robot to work towards a common goal. In such a situation, the robot is less of a tool and more of a partner or teammate. Much of the research in this area has drawn upon the factors affecting human-to-human relationships and human teams, as the premise is that humans are more likely to team with robots if robots possessed the characteristics that allow humans to team with other humans. These characteristics encompass (i) attributes of the robot that directly facilitate teaming, or (ii) factors that promote emergent characteristics that contribute to teaming.

Attributes that Directly Facilitate Teaming. There is a line of HRT research on the social-cognitive mechanism and processes required to design robots that can team. Several studies have proposed that robot teammates need to possess the human attribute of having theory of mind (ToM), which allows humans to cooperate and team with other humans (e.g., [30–34]. ToM involves inferring other's mental states (i.e., thoughts, beliefs, intents) from their behaviors such as speech, actions, facial expressions, and gestures [35]. The cognitive mechanisms implicated in ToM relate to Simulation Theory, which posit that we, humans, infer the other person's mental states by thinking as if we are the other person, i.e., we simulate the other's actions and stimuli they experience in our own minds, using our own cognitive mechanisms, to predict what they are thinking [36]. Together with ToM, Simulation theory has provided HRT researchers some direction in terms of the social-cognitive mechanisms that robots should possess. For example, robots should be able to discern where its human teammate's attention is directed by following his/her eye gaze. This notion has been investigated in a number of studies (e.g., [37–40]).

Another human characteristic that supports teaming is the ability to possess joint intent. The Joint Intention Theory postulates that teammates need a set of shared beliefs so that they can work together towards a common goal. The theory includes the concepts of *joint activity* and *joint commitment*. Joint activity results from the sharing of specific mental properties, while joint commitment is prioritizing the common goal above individual goals, as well as having a mutual belief about the status of the goal [41]). Ideas of the theory have been investigated in HRT research (e.g., [42–45], and have been implemented in models and frameworks such as STEAM (Shell for Teamwork) and GRATE* [46]. Another line of research related to the Joint Intention Theory is the work on shared mental models (SMMs) between humans and agents or robots. These studies include developing a research approach to measure and evaluate SMMs in human-robot teams [47], testing if SMM achieved in the planning phase can benefit teamwork in the execution phase [48], understanding how SMM can inform design of agents capable of teaming [49], and specifying requirements for a robot’s computational mental model of the task and teammate [50].

Factors that Promote Emergent Characteristics that Contribute to Teaming. A substantial amount of HRT research has been in trust in human-robot teams. In HRT, trust refers to the human’s “attitude that an agent will help achieve an individual’s (the human’s) goals in a situation characterized by uncertainty and vulnerability” [51]. Unlike communication capabilities and computational mental models, trust cannot be “built” into a robot, but is an emergent property of the human-robot relationship. The level of trust the human has in the robot can determine whether or not the human uses and relies on the robot [52]. A meta-analysis on trust in human-robot teams identified the following classes of factors [53]:

- Human-related
 - self-confidence [54]
 - inclination to trust [55]
 - expertise [56]
 - familiarity and understanding of robot functioning [57]
- Robot-related
 - robot reliability, predictability [57]
 - proximity [58]
 - robot personality [59]
 - anthropomorphism
- Environmental factors
 - culture [60]

The Uncertainty Reduction theory, which postulates that humans act to reduce uncertainty in their interactions, accounts for a few of these factors. For instance, the Uncertainty reduction theory is supported by the factor relating to being familiar

and having an understanding of robot functioning, i.e., trust in the robot is more likely when the human understands how the robot works, when the robot's mechanisms and algorithms are transparent to the human. Research in the effects of robot transparency as related to trust indicate that transparency is related to perceived predictability [61, 62], another factor that impacts trust. Robot reliability and predictability denotes a high degree of consistency in robot performance, which minimizes the uncertainty experienced by the human. Hence, the theory appears to be supported by the results of the meta-analysis.

4 Conclusion

The review in the current paper indicates that theory is still an important part of HRT research. In areas reviewed, there are middle abstraction level theories that can still inform direction of research and explain certain observations. Findings from HR teaming research can still be informative for theory building because as researchers reverse engineer human capabilities in robots, they should be able to test and contribute to theories of human social behavior and cognitive processes.

References

1. Talamadupula, K., et al.: Planning for human-robot teaming. In: ICAPS 2011 Workshop on Scheduling and Planning Applications (SPARK) (2011)
2. Sabelli, A.M., Kanda, T.: Robovie as a mascot: a qualitative study for long-term presence of robots in a shopping mall. *Int. J. Soc. Robot.* 1–11 (2015)
3. Boissy, P., et al.: Usability testing of a mobile robotic system for in-home telerehabilitation. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 1839–1842. IEEE (2011)
4. Neuman, W.L.: *Social research methods: qualitative and quantitative approaches*. Allyn & Bacon, Needham Heights (1997)
5. Parsons, T.: The role of theory in social research. *Am. Soc. Rev.* 3(1), 13–20 (1938)
6. Wacker, J.G.: A definition of theory: research guidelines for different theory-building research methods in operations management. *J. Oper. Manag.* 16(4), 361–385 (1998)
7. Steinfeld, A., Jenkins, O.C., Scassellati, B.: The oz of wizard: simulating the human for interaction research. In: *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 101–107. IEEE (2009)
8. Surgery, d.V.: Changing the experience of surgery. [cited 2016 03/09]; Available from: <http://www.davincisurgery.com/>
9. McNickle, M.: 10 medical robots that could change healthcare. *Information Week 2012*; Available from: <http://www.informationweek.com/mobile/10-medical-robots-that-could-change-healthcare/d/d-id/1107696>
10. Shamsuddin, S., et al.: Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO. In: *2012 IEEE 8th International Colloquium on Signal Processing and Its Applications (CSPA)*, pp. 188–193. IEEE (2012)
11. Lamb, R.: How have robots changed manufacturing? (2010) [cited 2016 03/09]; Available from: <http://science.howstuffworks.com/robots-changed-manufacturing.htm>

12. Doisy, G., Meyer, J., Edan, Y.: The impact of human-robot interface design on the use of a learning robot system. *IEEE Trans. Hum. Mach. Syst.* **44**(6), 788–795 (2014)
13. Vidyasagar, M.: System theory and robotics. *IEEE Control Syst. Mag.* **7**(2), 16–17 (1987)
14. Michaud, F., et al.: Exploratory design and evaluation of a homecare teleassistive mobile robotic system. *Mechatronics* **20**(7), 751–766 (2010)
15. Gillian, D., Riley, J., McDermott, P.: The cognitive psychology of human-robot interaction. *Hum. Robot Interact. Future Mil. Oper.* 53–66 (2010)
16. Wickens, C.D.: Multiple resources and mental workload. *Hum. Factors J. Hum. Factors Ergon. Soc.* **50**(3), 449–455 (2008)
17. Perzanowski, D., et al.: Building a multimodal human-robot interface. *IEEE Intell. Syst.* **16**(1), 16–21 (2001)
18. Rogalla, O., et al.: Using gesture and speech control for commanding a robot assistant. In: 11th IEEE International Workshop on Robot and Human Interactive Communication. Proceedings, pp. 454–459. IEEE (2002)
19. Stiefelhagen, R., et al.: Natural human-robot interaction using speech, head pose and gestures. In: 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 2004. Proceedings, pp. 2422–2427. IEEE (2004)
20. Salter, T., Dautenhahn, K., te Boekhorst, R.: Learning about natural human-robot interaction styles. *Robot. Auton. Syst.* **54**(2), 127–134 (2006)
21. Barber, D.J., et al.: Field assessment of multimodal communication for dismounted human-robot teams. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 921–925. SAGE Publications (2015)
22. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a survey. *Found. Trends Hum. Comput. Interact.* **1**(3), 203–275 (2007)
23. Harris, J., Barber, D.: Speech and gesture interfaces for squad-level human-robot teaming. In: SPIE Defense + Security, pp. 90840B-90840B-11. International Society for Optics and Photonics (2014)
24. Powers, A., Kiesler, S., Goetz, J.: Matching robot appearance and behavior to tasks to improve human-robot cooperation, vol. 105. Human-Computer Interaction Institute (2003)
25. Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* **19**(2), 98–100 (2012)
26. Murphy, R.R., Riddle, D., Rasmussen, E.: Robot-assisted medical reachback: a survey of how medical personnel expect to interact with rescue robots. In: 13th IEEE International Workshop on Robot and Human Interactive Communication, 2004. ROMAN 2004, pp. 301–306. IEEE (2004)
27. Minato, T., et al.: Development of an android robot for studying human-robot interaction. In: Innovations in Applied Artificial Intelligence, pp. 424–434. Springer, Berlin (2004)
28. Thompson, L.F., Gillan, D.J.: Social factors in human-robot-interaction. In: Human-robot-interactions in future military operations, pp. 67–81 (2010)
29. Venkatesh, V., et al.: User acceptance of information technology: Toward a unified view. *MIS Q.* 425–478 (2003)
30. Wiltshire, T.J., Barber, D., Fiore, S.M.: Towards modeling social-cognitive mechanisms in robots to facilitate human-robot teaming. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1278–1282. SAGE Publications (2013)
31. Streater, J., Bockelman Morrow, P., Fiore, S.: Making things that understand people: the beginnings of an interdisciplinary approach for engineering computational social intelligence. In: 56th Annual Meeting of the Human Factors and Ergonomics Society, Boston, MA, 22–26 Oct 2012
32. Hiatt, L.M., Harrison, A.M., Trafton, J.G.: Accommodating human variability in human-robot teams through theory of mind. In: IJCAI Proceedings-International Joint Conference on Artificial Intelligence, pp. 2066 (2011)
33. Scassellati, B.M.: Foundations for a Theory of Mind for a Humanoid Robot. Massachusetts Institute of Technology (2001)

34. Gratch, J., et al.: Exploring the implications of virtual human research for human-robot teams. In: *Virtual, Augmented and Mixed Reality*, pp. 186–196. Springer, Berlin (2015)
35. Schlinger, H.D.: Theory of mind: an overview and behavioral perspective. *Psychol. Record* **59**(3), 435 (2009)
36. Breazeal, C., Gray, J., Berin, M.: Mindreading as a foundational skill for socially intelligent robots. In: *Robotics Research*, pp. 383–394. Springer, Berlin (2010)
37. Das, D., et al.: Recognizing gaze pattern for human robot interaction. In: *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 142–143. ACM (2014)
38. Johansson, M., Skantze, G., Gustafson, J.: Head pose patterns in multiparty human-robot team-building interactions, pp. 351–360. In: *Social Robotics*. Springer, Berlin (2013)
39. Breazeal, C., et al.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005. IROS 2005, pp. 708–713. IEEE (2005)
40. Xu, T., Zhang, H., Yu, C.: Cooperative gazing behaviors in human multi-robot interaction. *Interact. Stud.* **14**(3), 390–418 (2013)
41. Cohen, P.R., Levesque, H.J.: Confirmations and joint action. In: *IJCAI*, pp. 951–959 (1991)
42. Scheutz, M., Schermerhorn, P., Kramer, J.: The utility of affect expression in natural language interactions in joint human-robot tasks. In: *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 226–233. ACM (2006)
43. Subramanian, R.A., Kumar, S., Cohen, P.R.: Integrating joint intention theory, belief reasoning, and communicative action for generating team-oriented dialogue. In: *Proceedings of the National Conference on Artificial Intelligence*, p. 1501. AAAI Press; MIT Press; Menlo Park, CA; Cambridge, MA; London (2006)
44. Hoffman, G., Breazeal, C.: Collaboration in human-robot teams. In: *Proceedings of the AIAA 1st Intelligent Systems Technical Conference*, Chicago, IL, USA (2004)
45. DeMarco, K.J., West, M.E., Howard, A.M.: Autonomous robot-diver assistance through joint intention theory. In: *Oceans-St. John's*, pp. 1–5. IEEE (2014)
46. Jennings, N.R.: Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. *Artif. Intell.* **75**(2), 195–240 (1995)
47. Schuster, D., et al.: A research approach to shared mental models and situation assessment in future robot teams. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 456–460. SAGE Publications (2011)
48. Nikolaidis, S., Shah, J.: Human-robot teaming using shared mental models. *ACM/IEEE HRI* (2012)
49. Jonker, C.M., et al.: Towards measuring sharedness of team mental models by compositional means. In: *AAAI Fall Symposium: Robot-Human Teamwork in Dynamic Adverse Environment* (2011)
50. Goodrich, M.A., Yi, D.: Toward task-based mental models of human-robot teaming: a bayesian approach. In: *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*, pp. 267–276. Springer, Berlin (2013)
51. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors J. Hum. Factors Ergon. Soc.* **46**(1), 50–80 (2004)
52. Lyons, J.B.: Being transparent about transparency: a model for human-robot interaction. In: *2013 AAAI Spring Symposium Series* (2013)
53. Hancock, P.A., et al.: A meta-analysis of factors influencing the development of human-robot trust. *DTIC Document* (2011)
54. Freedy, A., et al.: Mixed initiative team performance assessment system (MITPAS) for training and operation. In: *Interservice/Industry Training, Simulation and Education Conference (IITSEC)*, pp. 1–10 (2007)
55. Adams, J.A.: Multiple robot/single human interaction: effects on perceived workload. *Behav. Inf. Technol.* **28**(2), 183–198 (2009)

56. McBride, M., Morgan, S.: Trust calibration for automated decision aids. Institute for Homeland Security Solutions [Online]. Available: https://www.ihssnc.org/portals/0/Documents/VIMSDocuments/McBride_Research_Brief.pdf (2010)
57. Ogreten, S., Lackey, S., Nicholso, D.: Recommended roles for uninhabited team members within mixed-initiative combat teams. In: 2010 International Symposium on Collaborative Technologies and Systems (CTS), pp. 531–536. IEEE (2010)
58. Bainbridge, W.A., et al.: The effect of presence on human-robot interaction. In: The 17th IEEE International Symposium on Robot and Human Interactive Communication. RO-MAN 2008, pp. 701–706. IEEE (2008)
59. Parasuraman, R., Miller, C.A.: Trust and etiquette in high-criticality automated systems. *Commun. ACM* **47**(4), 51–55 (2004)
60. Li, D., Rau, P.L.P., Li, Y.: A cross-cultural study: effect of robot appearance and task. *Int. J. Soc. Robot.* **2**(2), 175–186 (2010)
61. Cring, E.A., Lenfestey, A.G.: Architecting human operator trust in automation to improve system effectiveness in multiple unmanned aerial vehicles (UAV) control. DTIC Document (2009)
62. Chen, J.Y., et al.: Situation Awareness-Based Agent Transparency. DTIC Document (2014)

Part IV
From Theory to Application:
UAV and Human-Robot Collaboration

Classification and Prediction of Human Behaviors by a Mobile Robot

D. Paul Benjamin, Hong Yue and Damian Lyons

Abstract Robots interacting and collaborating with people need to comprehend and predict their movements. We present an approach to perceiving and modeling behaviors using a 3D virtual world. The robot's visual data is registered with the virtual world to construct a model of the dynamics of the behavior and to predict future motions using a physics engine. This enables the robot to visualize alternative evolutions of the dynamics and to classify them. The goal of this work is to use this ability to interact more naturally with humans and to avoid potentially disastrous mistakes.

Keywords Virtual world · Soar cognitive Architecture

1 Introduction

The ADAPT project (Adaptive Dynamics and Active Perception for Thought) is a collaboration of three university research groups at Pace University, Brigham Young University, and Fordham University to produce a robot cognitive architecture that integrates the structures designed by linguists and cognitive scientists with those developed by robotics researchers for real-time perception and control. ADAPT is under development on Pioneer robots in the Pace University Robotics Lab and the Fordham University Robotics Lab. Publications describing ADAPT are [1–4].

D. Paul Benjamin (✉) · H. Yue
Pace University, 1 Pace Plaza, New York, NY 10038, USA
e-mail: dbenjamin@pace.edu

H. Yue
e-mail: yh19243n@pace.edu

D. Lyons
Fordham University, 340 JMH, 441 E. Fordham Road, Bronx, NY 10458, USA
e-mail: dlyons@fordham.edu

2 The ADAPT Architecture

Our approach is fundamentally different from other projects, which typically attempt to build a comprehensive system by connecting modules for each different capability: learning, vision, natural language, etc. Instead, we are building a complete cognitive robotic architecture by merging RS [5–7], which provides a model for building and reasoning about sensory-motor schemas, with Soar [8], a cognitive architecture that is under development at a number of universities. RS possesses a sophisticated formal language for reasoning about networks of port automata and has been successfully applied to robot planning. Soar is a unified cognitive architecture [9] that has been successfully applied to a wide range of tasks.

Soar’s model of problem solving utilizes a single mechanism of subgoaling and chunking to explain human problem solving performance; utilizing Soar as the basis of ADAPT permits us to unify the mechanisms underlying perception, language and planning. Furthermore, it permits us to explore possible interrelationships between learning in these areas, e.g. how learning language and learning perception may be related. Finally, it permits us to test our architecture on robotic versions of well-known cognitive tasks and explore how robot learning might be related to human learning.

RS provides a powerful representational language for the system’s dynamics, language and percepts; however, RS does not provide a mechanism for synthesizing the dynamics. Furthermore, RS lacks demonstrated cognitive plausibility, and in particular lacks a learning method.

We have implemented RS in Soar to take advantage of Soar’s cognitively plausible problem-solving and learning mechanisms. Soar uses universal subgoaling to organize its problem solving process into a hierarchy of subgoals, and uses chunking to speed and generalize that process. Universal subgoaling permits Soar to bring all its knowledge to bear on each subgoal. Chunking stores generalized preconditions for search control decisions, so that in future tasks similar search control decisions are made in a single step.

3 Visualization and Prediction of Human Behaviors

We believe that the comprehension of human movements and intentions requires the ability to visualize human movements and predict possible future movements. We view visualization as consisting of both a perceptual component and a reasoning component. The perceptual component is performed using the same perceptual mechanism that the robot uses to perceive its environment; the difference is that visualization perceives a simulation of the environment. Visual reasoning manipulates and superimposes representations that consist of a combination of symbolic knowledge and 3D animations. This approach to comprehension requires

the robot to be able to create different situations in which it can generate behaviors of robots, people and physical systems, and perceive the results of these behaviors. This requires implementing a virtual world that the robot can control.

ADAPT's virtual world is a multimedia simulation platform capable of realistic simulations of physical phenomena. It combines the various forms of map information found in most robots: topological, metric and conceptual information. ADAPT completely controls this virtual world, and can create arbitrary objects and behaviors in it, including nonexistent objects and behaviors that were not actually observed. Central to ADAPT's use of its virtual world is its ability to view these constructions from any point. This enables ADAPT to create visual representations with desired properties.

In the current implementation, ADAPT's world model is PhysX. PhysX gives the robot the ability to create a detailed and dynamic virtual model of its environment, by providing sophisticated graphics and rendering capabilities together with a physics engine based on the PhysX physics engine. PhysX models a wide variety of dynamic environments, including modeling other agents moving and acting in those environments.

This world model is used to represent the important entities and behaviors in the environment. The built-in physics capability of PhysX is then used to predict what is going to happen in the immediate future. Let's examine how ADAPT uses the virtual world together with its vision system to model and predict the environment. ADAPT's vision system consists of two main components, a bottom-up component that is always on, and a top-down goal-directed component controlled by Soar.

The bottom-up component is simple and fast. It does this by not producing much detail. The idea is for it to produce a basic stereo disparity map, a coarse-grained image flow, and color segmentation in real time. It runs on the robot's onboard computer using Intel's open vision library, and segments the visual data from the robot's two frame-grabbers. These "blobs" are transmitted together with stereo disparity data and optical flow to the off-board PC that is running Soar, where it is placed into working memory. This component is always on, and its output is task-independent.

The top-down component executes the more expensive image processing functions, such as object recognition, sophisticated image flow analysis, and application of particular filters to the data. These functions are called in a task-dependent and goal-dependent manner by Soar operators. This greatly reduces their frequency of application and speeds the operation of the vision system significantly. These two components are not connected to each other; instead, the output of the bottom-up component is used by Soar to determine when to call the top-down operations. Soar compares the bottom-up output to the visual data predicted by the virtual world.

The virtual world can display the view that the virtual copy of the robot "sees" in the virtual environment. The output of this graphics "camera" in PhysX is segmented and sent to the MMD (Match-Mediated Difference), together with distance information and motion information. The MMD tests for significant differences between the expected view and the actual view, e.g. the appearance of a large new

blob or a large change in optical flow. It aligns the real and virtual images with an affine map, then finds a set of matched key points and place a normalized Gaussian at each of them. The normalized match quality is the inverse of the distance between matched points divided by the sum of all match errors. We use this as a coefficient of the Gaussians to create the MMD measure. Any significant difference is placed in Soar's working memory, where it can cause an operator to be proposed to attend to this difference.

Soar controls all major aspects of perceptual processing, with the goal of constructing a virtual model of the environment that can be used in task planning. These aspects include focusing on regions of interest, choosing the depth of field, and deciding on the degree of detail for each part of the virtual model. For example, if a new blob appears, an abstract operator will be proposed to focus on this blob and try to recognize it. If this operator is selected (if there is no more important operator to do at the moment) then RS/Soar will instruct the robot to turn its

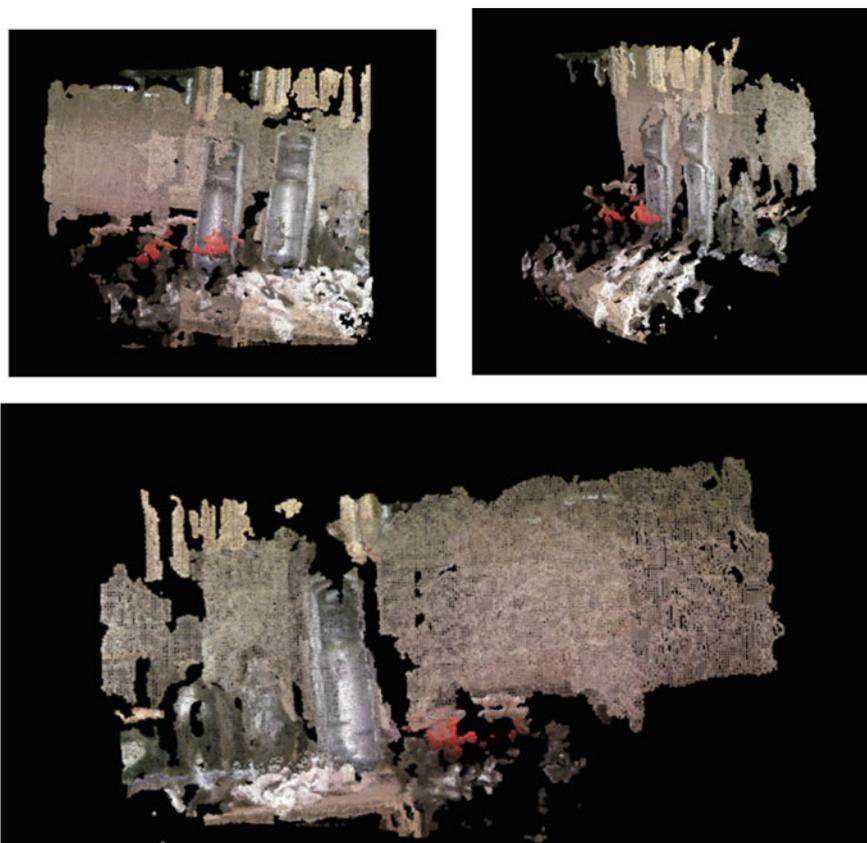


Fig. 1 The two point clouds at *top* are registered and joined to create the *bottom* cloud, which is transformed into a PhysX mesh

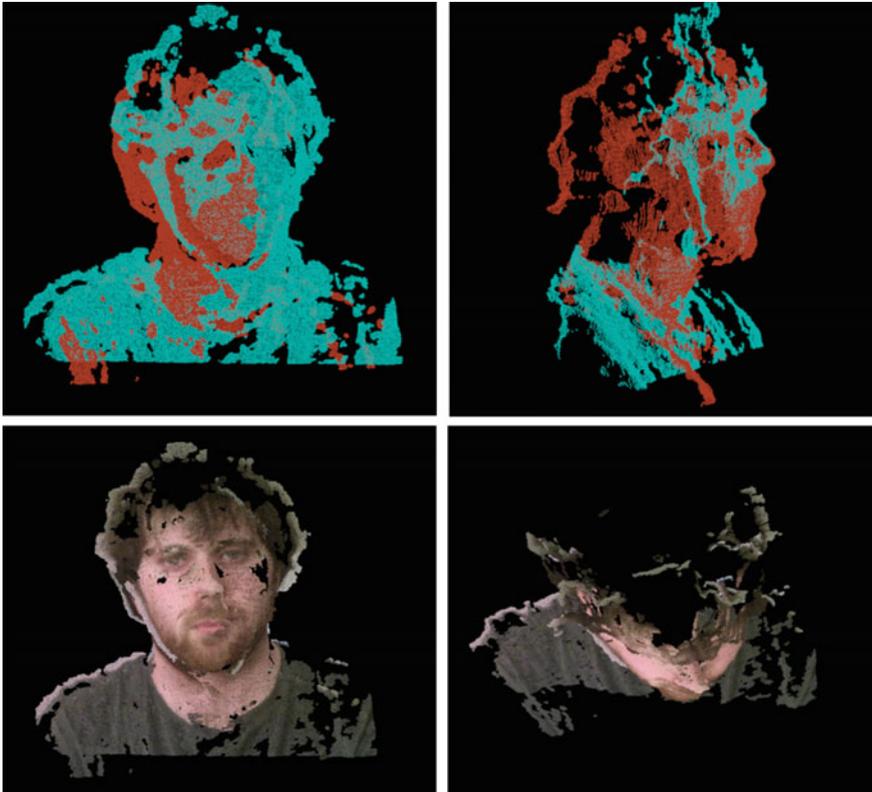


Fig. 2 Point clouds for a face, registered and joined to create a 3D mesh for a skeleton. The two point clouds at top are joined to create the cloud seen below, frontal view at *left* and from above at *right*

cameras towards this blob, focus at the appropriate depth, and obtain a point cloud from the visual input. Keypoints are extracted from the point cloud and used both in object recognition and to create a mesh that is registered with surrounding meshes.

Figure 1 shows two point clouds of a kitchen counter by this vision system. These are transformed into meshes. By joining small meshes together, the system can create larger meshes when necessary. Portions of the world that are not relevant to task goals are not rendered, greatly increasing efficiency.

This approach also works for people, as shown in Fig. 2. People are skeletonized in a manner similar to that used by the Kinect, and the skeleton is covered with mesh. Our initial work was with the Kinect, but now uses only stereo vision.

Once an object is recognized or rendered, a virtual copy is created in PhysX. The object does not need to be recognized again; as long as the blobs from the object approximately match the expected blobs from PhysX, ADAPT assumes it is the same object. Recognition becomes an explicitly goal-directed process that is much

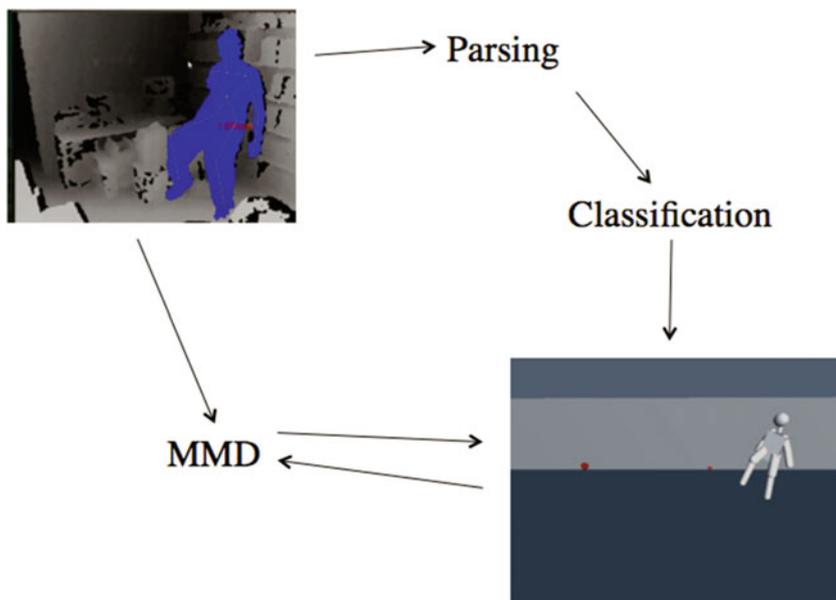


Fig. 3 The input from the physical world (*left above*) is parsed, classified and rendered into the virtual world (*below right*). The MMD compares both inputs, detects differences and updates the virtual world

cheaper than continually recognizing everything in the environment. The frequency with which these expensive operations are called is reduced, and they are called on small regions in the visual field rather than on the whole visual field.

Thus, ADAPT's vision system spends most of its time verifying hypotheses about its environment, instead of creating them. The percentage of its time that it must spend attending to environmental changes depends on the dynamic nature of the environment; in a relatively static environment (or one that the robot knows well from experience) there are very few unexpected visual events to be processed, so visual processing operators occupy very little of the robot's time.

Figure 3 shows the overall flow of control of our vision system [10, 11]. The real and synthetic images of the scene as viewed by the robot are compared. If the scenes are considered the same but from different viewpoints, then the viewpoint of the camera in the simulation is changed, and the simulation generates an image taken by the camera at the new location. If an unexpected object is seen in the real image, an object is introduced at the corresponding position in the simulated scene. The region of the real image responsible for the difference is used as video texture on the object and a new synthetic image generated. The information on whether there is no difference, an unexpected object, or an object missing between the image pairs is made available to action planning.

This loop of difference detection and simulation modification is used to keep the simulation synchronized to the observed environment. For prediction purposes, the

simulation can be allowed to ‘fast forward’ in time, so that the expected position, for example, of a target can be calculated and then compared to observations.

4 Summary

We have sketched the overall design of a new approach to the comprehension of behavior that is part of a robotic cognitive architecture. A powerful 3D multimedia world model is used to render the behaviors in the environment. This gives the robot the ability to visualize alternative evolutions of the dynamics and to classify them. The implementation of the basic components is complete.

Further information on this work, including video clips showing the robot moving under the control of schemas and the use of the world model, can be downloaded from the website for the Pace University Robotics Lab: <http://csis.pace.edu/robotlab>

References

1. Benjamin, D.P., Lyons, D., Lonsdale, D.: Embodying a Cognitive model in a mobile robot, In: Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision, Boston, (October 2006)
2. Benjamin D.P., Lyons, D., Achemichuk, T.: Obstacle avoidance using predictive vision based on a dynamic 3D world model, In: Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision, Boston, (October 2006)
3. Benjamin, D.P., Lyons, D., Lonsdale D.: Designing a robot cognitive architecture with concurrency and active perception, In: Proceedings of the AAAI Fall Symposium on the Intersection of Cognitive Science and Robotics, Washington, D.C., (October 2004)
4. Benjamin, D.P., Lyons, D., Lonsdale, D.: Cognitive robots: integrating perception, action and problem solving in behavior-based robots, In: Proceedings of (AAMAS-2004) pp. 1308–1309 (2004)
5. Lyons, D.M.: Representing and analysing action plans as networks of concurrent processes, IEEE Transactions on Robotics and Automation, (June 1993)
6. Laird, J.E., Newell, A., Rosenbloom, P.S.: Soar: an architecture for general intelligence. *Artif Intell* **33**, 1–64 (1987)
7. Newell, Allen: *Unified Theories of Cognition*. Harvard University Press, Cambridge, Massachusetts (1990)
8. Lyons, D.M., Hendriks, A.: Exploiting patterns of interaction to select reactions. *Spl Issue on Comput Theor Interact, Artif Intell* **73**, 117–148 (1995)
9. Lyons, D.M., Arbib, M.A.: a formal model of computation for sensory-based robotics, *IEEE Transactions on Robotics and Automation* **5**(3), (June 1989)
10. Lyons, D.M., Chaudhry, M., Benjamin D.P., Marius Monaco, A.J.V.: Integrating perception and problem solving to predict complex object behaviors, In: Conference on Multisensor, Multisource Information Fusion, SPIE, (April 2010)
11. Lyons D.M., Benjamin, D.P.: Robot video tracking by comparing real and simulated video scenes, In: Conference on Intelligent Robots and Computer Vision, SPIE, San Jose, Calif., (January 2009)

The Future of Human Robot Teams in the Army: Factors Affecting a Model of Human-System Dialogue Towards Greater Team Collaboration

A. William Evans, Matthew Marge, Ethan Stump, Garrett Warnell, Joseph Conroy, Douglas Summers-Stay and David Baran

Abstract Understanding of intent is one of the most complex traits of highly efficient teams. Combining elements of verbal and non-verbal communication along with shared mental models about mission goals and team member capabilities, intent requires knowledge about both task and teammate. Beginning with the traditional models of communication, accounting for teaming factors, such as situation awareness, and incorporating the sensing, reasoning, and tactical capabilities available via autonomous systems, a revised model of team communication is needed to accurately describe the unique interactions and understanding of intent which will occur in human-robot teams. This paper focuses on examining the issue from a system capability viewpoint, identifying which system capabilities can mirror the abilities of humans through the sensor and computing strengths of autonomous systems, thus creating a team environment which is robust and adaptable while maintaining focus on mission goals.

Keywords Human-robot teaming · Intent understanding · Shared mental model · Human-robot communication

1 Introduction

Much like ‘trust’ or ‘mind’, ‘intent’ is a term that is easy to understand but difficult to define. For operational purposes, Webster-Dictionary [1] defines intent as ‘the things you plan to do or achieve: an aim or purpose.’ Understanding intent is a critical component to creating an efficient team. As commands or instructions are given, all team members need to have an understanding of the team goals, as well as the processes being directed toward achieving those goals. This becomes even more important in task environments where team communications become degraded,

A. William Evans (✉) · M. Marge · E. Stump · G. Warnell · J. Conroy · D. Summers-Stay · D. Baran
US Army Research Laboratory, Adelphi, MD, USA
e-mail: arthur.w.evans20.civ@mail.mil

such as those in which military operations often occur. Vague or incomplete instructions related to command intent must still be executed in a timely and effective manner, and misunderstandings can have dire consequences. Understanding command intent is a difficult enough task in human-only teams, and incorporating autonomous systems into the team structure only serves to further complicate the issue.

In such human-robot teams, supervisory control solutions have become a prominent model. A key requirement for such solutions to enable greater human-robot integration is that each autonomous system in the team must be able to determine their own individual tactical behaviors based upon inferences made about the human supervisor's intent, rather than by direct response to specific command inputs. Previous research [2] supports this concept, finding that mixed initiative teams, with some decisions made by human controllers and some by autonomous systems, outperformed both fully manually controlled and fully autonomous teams while completing a search task. As autonomous systems become more capable and more integrated in teams with human teammates, their ability to communicate effectively becomes more important. Moving away from traditional joystick controls to speech control is a first step, but more may be required to allow autonomous systems to function as peers within a team structure. Autonomous systems will need to not only respond to structured language protocols, but will also need to understand natural language and participate in a dialogue with human teammates, incorporating language, gestures, eye gaze, mission and environmental factors.

For the last decade, human-robot teaming research has focused on the concept of moving the team structure from many humans controlling one robot, to one human controlling one robot, and for the future one human controlling many robots [3]. The motivation behind this concept was that a single Soldier can push their area of influence out further with each additional autonomous system they are able to control [4]. Realizing such a concept is not without its challenges. Current human-robot interactions typically require users to provide direct input of specific commands to the autonomous systems they control. Research and technology advances have helped to reduce the workload and cognitive demands imposed on the user of individual autonomous system [5, 6]. However, having to provide different commands to each individual system inherently places a limit on the number of entities that could be controlled by a single operator and represents a significant deviation from the processes seen in effective human-only teams. Previous research on RoboLeader, an agent for controlling multiple robots, has shown that distributed supervisory, rather than direct, control reduces overall user cognitive burden and may provide a unique avenue for multiple autonomous systems to observe and operate collaboratively within a human-robot team structure in more natural ways [7].

A new research program at the Army Research Laboratory is investigating how to use team dialogue as a way to enable autonomous systems to understand Soldier intent. Dialogue, in this case, is being defined as all of the communications (both verbal and non-verbal) that can reduce uncertainty about a command's intent. This can include spoken and text speech, eye gaze, position and heading, hand gestures,

environmental cues, etc., all of which can contain important tactical or contextual information. For the purposes of this program, a heterogeneous team of autonomous systems will be utilized to collectively observe human teammates, as well as the environment, to gather as much information as possible to aid in the understanding of the natural command language, eye gaze behavior, and gestures used in present day human squad activities. The information that the autonomous systems collect will then be utilized to infer intent and determine actions in the form of tactical behaviors, the specification of which will assist the squad in completing their objectives. This paper will focus on the factors associated with existing teamwork and communication models to determine which model(s) could support the inclusion of autonomous system capabilities. Additionally, a review of the factors that affect team performance, specifically shared mental models, verbal communication, and non-verbal communication will be discussed. Finally, a discussion is presented about how these factors can be translated to autonomous system capabilities, which can lead to the creation of a new model of human-robot team interaction.

2 Models of Communication

Here we consider classical models of communication and how they can inform the needs of a new Human-Robot Team Communication model. Communication has been a topic of research since the late 1940s and continues to be widely studied in this century, reaching out across domains like marketing, psychology, mathematics, business, and more. Osgood and Schramm’s [8] circular model of communication (Fig. 1), is still relevant and the basis of more advanced communications models today. This model expresses communication in simple ideals of a continuous loop, which involves encoding and decoding messages between two entities.

While Schramm’s model is used to explain the process of communication, Berlo’s model of communication [9] explains the components that fit that process in more detail (Fig. 2). It is through this model that we can see that communication is much more than simply a verbal passing of information. Every component from the ‘source’, through the ‘message’ and ‘channel’, to the ‘receiver,’ is affected by a number of various factors.

Fig. 1 Osgood and Schramm’s model of communication

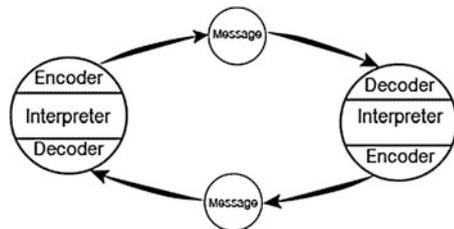
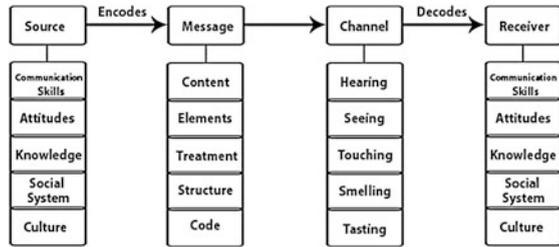


Fig. 2 Berlo's component model of communication



Berlo's model is useful because it not only accounts for explicit means of communication, such as communication skills, but also implicit factors, such as existing knowledge, social norms, and cultural bias, which can all affect the understanding of intent. Considering this model for human-robot interaction allows for the use of cues beyond simple direct commands and can help enhance the interpretation of the information found within vague command requests. Expanding on this model by also including factors that more easily incorporate non-verbal communications (such as eye gaze) could add valuable information to the intent understanding capabilities of autonomous systems.

A new Human-Robot Team Communication model will need to incorporate more implicit communication factors to allow autonomous systems to take advantage of more of the information that is available within the team environment. For more natural dialogue between human and autonomous system, autonomous systems will need to fully utilize components of communication, such as knowledge (i.e., mission goals, mission space, etc.) and culture (i.e., local norms, team norms, etc.).

Team communication and/or team performance models often build on the principles of the Osgood and Schramm and Berlo models, by incorporating a situation awareness component to help account for team processes toward a common goal. Situation awareness (SA) can be described in terms of overall team SA or individual team member SA, which is comprised of separate and overlapping elements [10]. Overlapping and correct SA for individual team members is a critical factor to creating high performing teams and leads to efficient and effective team communication patterns, as well as the development of shared mental models [11].

3 Shared Mental Models

Here we consider the importance of establishing a shared mental model to overall team performance, especially as it relates to communication. Team performance has been shown to be influenced by many factors [12, 13]. Of those factors, one of the most closely related to the understanding of intent and enabling high performing teams is the presence of shared mental models [14–17]. Shared mental models allow team members to understand the context within which actions or

communications are being executed and which actions or communications significantly affect decision making, planning, and coordination [18]. Shared mental models can allow teammates to efficiently divide tasks toward the achievement of a common goal [19], and can also allow teammates to more correctly predict behaviors of other teammates based on fewer cues providing less or incomplete information, which in turn creates a more productive, more efficient team. These cues can be team member based (e.g., a head nod, a point, or eye movements), task based (e.g., specific team behaviors based on situation), or environmentally based (e.g., changes in environmental condition) and are often non-verbal in nature. Shared mental models, along with the previously mentioned cues, could also provide context and clarity to ambiguous statements or commands. Based on mission goals, Soldiers in well-functioning teams will understand their Sergeant's order to "take that building" to mean 'breach and clear the third building on the left side of the road', if accompanied by correct cues about gaze direction, semantic understanding, existing mission knowledge, and evaluation of the current environment.

The establishment of shared mental models in human-only teams is often created via education, training, and familiarity. The introduction of autonomous systems into the team structure yields some obstacles to the creation of shared team mental models. While goal-based concepts might be easily understood by autonomous systems, more abstract or vague communications, which are widely used in high performance team processes, produce a barrier in the human-autonomous system team communication process. In addition, assessing shared mental models of human-robot teams could be difficult. Some recent success has been found by Nikolaidis and Shah [20] in exploiting task procedure assessments to evaluate autonomous system mental models. While this research focused on a more structured manufacturing setting, this approach might prove viable for more complex military tasks. The application of procedural knowledge could be leveraged to enable autonomous systems to engage in the use of set tactical behaviors in a variety of configurations. Utilizing tactical behaviors in this way could help to increase not only the capabilities of the system but potential areas for overlap with human mental models about the associated tasks.

4 Verbal Communication

Humans accomplish tasks by communicating, where they often use dialogue to achieve a mutual understanding of information [21]. Verbal communication is governed by the communicative acts that dialogue partners select to establish and maintain a shared mental model. These actions consist of anything that a dialogue partner does to make others understand his or her intentions, like declaring information to or requesting information from a dialogue partner, reporting problems, or correcting a misunderstood belief. These actions can be in the form of verbal behaviors, like speech or vocal gestures, or in the form of non-verbal behaviors like pointing and gaze. In a conversation between Alice and Bob, Alice is either

presenting a communicative act to Bob, or accepting one from Bob. Generally, when Alice accepts a communicative act from Bob, new information has been grounded (i.e., added to Alice’s perceived common ground of information with Bob). Dialogue is the vehicle by which two or more dialogue partners establish this mutual understanding of information.

A long-standing challenge in artificial intelligence research has been to develop conversational interfaces that are user-friendly and useful for people. Following existing psycholinguistic principles, like grounding in communication [22], helps us move towards these goals. At the same time, advances in technologies like automatic speech recognition and speech synthesis have allowed conversational interfaces called *spoken dialogue systems* to run on mobile devices like smartphones and tablet computers (e.g., Siri on iOS). The traditional focus for dialogue system developers has been on tasks that link to a structured database or computer application. Generally, these systems are best equipped to handle structurally-defined tasks like travel booking.

There are many recent advances in robotics, computer vision, and navigation. Their emergence as potential team members in a variety of tasks has also entailed the need to communicate with them effectively. Natural language is one potential form of communication. Natural language dialogue provides an intuitive and flexible way for people to communicate with robots. Dialogue between people and robots is *physically situated* [23]—it refers to the environment: a streaming source of information that is often crucial for processing plans and accomplishing tasks. Dialogue systems can allow people to communicate with robots at a more abstract task level than typical command-and-control devices. However, such dialogue cannot be treated like the dialogue systems used to book travel; interactions will often require dynamic reference to the robot’s surroundings. To this end, there have been several human-to-robot dialogue systems developed for unmanned ground vehicles, unmanned aerial vehicles, and virtual agents [24, 25].

The nature of physically situated interaction yields a relatively unexplored set of problems that sets it apart from traditional human-computer dialogue. Interactions with robots may fail due to a robot’s dependency on specific environment information, or if its path to a goal is obstructed [26]. People engaged in dialogue with such a robot can help; in existing human-human navigation dialogues, Skantze [27] showed that people ask proactive, task-relevant questions instead of simply indicating task failure. Dialogue enables people to supplement a robot’s representation of the world and allow it to complete tasks. Robots could issue clarification questions to human partners, or those same partners could interrupt a robot’s current task with time-critical information. Verbal communication using natural language provides a hands-free method to engage in a dialogue with robots.

5 Non-verbal Communication

Non-verbal communication is an important component in many types of human collaboration and teaming. The ‘messages’ in non-verbal communication, like gestures or glances, can be thought of as symbols, which help to pair meaning and form [28]. The meanings gained from this type of communication are largely dependent on context. One of the most-studied ways in which it is done is through eye gaze behavior. Baron-Cohen et al. [29] showed that humans utilize knowledge of eye-gaze direction when attempting to understand the mental states of others. That is, when one human can determine where another is looking, implicit information, such as whom another person is addressing, what another person is referring to, or even some level of what another is thinking about, can be more accurately inferred by the observer. Clearly, then, the ability to know where other humans are looking is an important factor in generating an accurate shared mental model among teammates. For example, for the specific collaborative task of one-on-one learning from demonstration, it has been argued that both the instructor and the learner must each be able to follow the eye gaze of their partner in order to form a *shared* or *joint attention* model that affords them critical spatial and feedback information [30, 31]. Strabala et al. [32] showed that even for the seemingly simple collaborative task of one human handing an object to another, eye gaze is a critical non-verbal cue that allows the team to successfully complete the handoff. Further, Huang et al. [33] have shown that analyzing the eye gaze of humans can be useful in predicting human intent, such as desired objects. Therefore, based on this evidence from the study of human-human teams, we argue here that endowing an autonomous system with the ability to extract human eye gaze information will give it a greater chance of being an effective teammate in a human-robot team.

6 Consideration of Autonomous Teammates

Here we consider the specific capabilities of modern unmanned vehicles and what technologies we can use to address the issues of communication discussed above and incorporate these vehicles into Human-Robot Teams.

6.1 Unmanned Vehicle Capabilities

For Army field operations, autonomous vehicles useful at a tactical level fall into two classes: aerial and ground. Aerial robotic vehicles are well-suited for applications requiring speed and a top-down point of view. Common applications now include examples such as crop analysis, 3D digital terrain mapping, search and rescue, and wildlife monitoring [34]. Ground robotic vehicles are well-suited for

applications that require long-duration operation and more extensive payload options. They have routinely been deployed in situations that require standoff, such as explosive ordnance disposal.

Aerial robotic platforms are highly desired for military applications as agents for intelligence, surveillance, and reconnaissance. They suffer, however, from a poverty of payload, which limits power, processing capabilities, and sensor options. These constraints lead to lower average data bandwidth, possible communications drop-outs or delays, and limited on-board intelligence. On the other hand, because the payload capacity of ground robotic platforms can be as high as several tens of kilograms, they can be outfitted with cameras, ranging sensors, manipulators, and significant computing power, as required for their mission.

Because robotic platforms can operate at long distances from their operators, it becomes possible to perceive the world from a completely different viewpoint, complicating the correspondence problem between the human's point-of-view and that of the robot. However, real-time estimates of pose (3D position and orientation in a global reference frame) are now ubiquitous in robot systems, making it possible to present information to operators within a third-person view.

While current systems are typically flown and driven either manually or via global positioning system (GPS) waypoint navigation, support for a true dialogue with a human requires some degree of autonomous operation. Primarily, the vehicle must exhibit a capability for reliable obstacle avoidance. Several current research implementations and a few commercial implementations support some level of generic obstacle avoidance, thus enabling the possibility of exploration algorithms and point-to-point travel in cluttered environments. Furthermore, current research systems can support many common on-board video/image processing functions, such as object recognition and target tracking, thus enabling the vehicle to implement some degree of autonomous perception, 3D vision-based mapping, target recognition and tracking. Memory and data storage has become miniaturized to the point that research-grade systems can support hundreds of gigabytes (for aerial platforms), or several terabytes (for ground platforms), of data storage to support large databases representing their world model. This world model tracks all of the spatial information that each system has collected, forming a basis for the shared mental model that could exist within the autonomous system.

6.2 Verbal Communications

The dialogue-based view of verbal communication forms the basis for our general semantic reasoning engine that considers higher-level mission goals communicated by the Soldier and decides upon the commanded actions that will best support them. Goals may be permanent, mission-specific (delivered as an explicit "statement of intent"), or implicit in the ongoing dialogue. Commanded actions can often be fulfilled in a variety of ways, some of which will be faster, less risky, or less costly than others.

Ultimately, robots must make decisions when given a command, and the process of decoding incoming data into information, comparing information with desired goals, and selecting an action is called *planning*. Dialogue enables robots to have collaborative communication about their plans. Planning has been considered a form of abductive reasoning, whereby an autonomous system attempts to generalize a new observation according to a hypothesis, and determines what to do next [35]. Autonomous systems could observe the Soldier performing a series of actions, and then follow the observed actions by engaging in a particular plan of their own. This would result in the robots performing actions complementary to the Soldier's. Robots could ask the Soldier a clarification question, which relies on reasoning. Conducting a dialogue with the Soldier involves selecting a communicative action that could resolve an inconsistency in the robot's plan. This type of interaction indicates to both the Soldier and robot teammates that the higher level mission is being followed. For real-world situations, this capability requires an in-depth knowledge of the possible plans and how they can vary, and the ability to robustly determine what actions are being performed.

Toward achieving this goal, we will make use of existing common sense reasoning technology. Common sense reasoners attempt to define a "core" set of knowledge that humans possess, that could potentially be transferred to artificial agents like robots. More specifically, we will use Cyc, software which encapsulates a common sense knowledge base to relate goals, events, and causes. Existing predicates built-into Cyc (e.g., the action Preconditions predicate, which defines what must happen before executing an associated action) allow us to express how any particular type of event can have effects that either further or harm the goals of an agent. We also plan to make use of analogical and approximate reasoning in Cyc by incorporating representations of assertions and predicates.

Each action that an autonomous agent can deliberately perform will be characterized in terms of some direct effects that it is likely bring about. These events, in turn, will be characterized in terms of what effects they might bring about and under what conditions. If the robot is able to deductively reason that one of its actions will set in motion a chain of events (i.e., a plan) that will further the goals of its team and another will harm those goals, it will choose the beneficial option.

6.3 *Non-verbal Communications*

We focus on the use of human eye gaze as our primary channel for collecting non-verbal information to augment the dialogue and help develop our shared mental model. Human eye gaze has been studied by the autonomous systems community in several different capacities. For a learning from demonstration task, Lockerd and Breazeal [31] gave a robot learner the ability to exhibit human gaze behavior as a way to provide nonverbal feedback to a human instructor. Conversely, Hoffman et al. [30] autonomously estimated human gaze direction from video, and used this information to better accomplish a similar learning task. Sakita et al. [36] also used

video to estimate human eye gaze during a joint human-robot construction task, and proposed several interesting collaborative behaviors that relied on this information, including “taking over” or “settlement of hesitation.” In the computer vision community, Park et al. [37] have argued that estimating *social saliency*, or what regions of space draw human attention, is a critical component of autonomous scene understanding and they present a technique to compute this quantity based on the head pose information of humans in the scene.

The success of the above techniques belies the difficulty associated with accurate, automatic estimation of both head-pose and eye-gaze information. Park and Shi [38] discuss the shortcomings of these techniques when estimating gaze from both third- and first-person cameras, and propose a technique for estimating social saliency using the spatial distribution of humans in the scene instead of eye-gaze information. In fact, the topic of accurately estimating eye gaze from third-person video is still quite active [37]. In contrast with these techniques, we extract gaze information from a different modality: eye-tracking glasses (e.g., the Tobii Glasses 2). These devices provide both egocentric (i.e., first-person) video and very robust estimates of the wearer’s eye gaze location as overlaid onto this video. We believe that acquiring gaze information in this fundamentally different way will allow us to overcome many third-person gaze estimation issues, and thereby pave the way for us to extend the previously-discussed human-robot teaming successes from the laboratory to more natural environments.

6.4 *Mapping Plans into Vehicle Actions*

In our envisioned system, we have two systems that support the human: (1) the reasoning system that interacts with the human through both verbal and non-verbal communication channels to infer their intent and render a general plan to accomplish this intent; and (2) the heterogeneous system of autonomous vehicles that can be tasked to carry out primitive maneuvers and collect raw sensor data. We imagine that we may not want the reasoning system to be endowed with knowledge of the specific capabilities of the vehicles that are being tasked; the job of the reasoning system is to decide *what* should be done to respect the user’s intent, not exactly *how*. In this case, how can we be assured that the reasoning system will produce plans that can actually be realized by the physical systems?

Our answer lies in the use of formal language as way of constraining the domain of plans to those that are realizable. State machines, or automata, have long been used in the engineering of systems as a way to model the relationships between input signals, output signals, and sequenced operations that need to be performed. There is an intimate connection between automata that can respond to a sequence of input signals and the formal grammars used to produce strings in a formal language [39]. One way of exploiting this connection is to take the plan language that is output from the reasoning system and learning the mapping from the plan semantics to the actions that the physical robots can perform, as has been done for a controlled

natural language [40]. Another way would be to begin from the capabilities of the system and design the language that matches these capabilities, as was done with Motion Grammars [41]. The first represents engineering the physical systems to match the language, and the second represents engineering the language to match the physical systems. A middle ground would be to use an intermediate representation, such as statements in a process logic, that incorporates the capabilities of the platforms as primitives and that we can expand the plan language into [42]. In any case, the reasoning system can issue plans without specifically referring to the platform capabilities, because either the plan language can be mapped on to them later or the plan language can only express capabilities that are available.

7 Toward a Human-Robot Team for Intent Understanding

Conceptually, a number of factors have been identified, which could be used to create a new model of human-robot team communication mirroring long established models from the human communication literature. Incorporating autonomous system sensor and reasoning capabilities, any future model need to be descriptive in the means to which autonomous systems can acquire and utilize data. Ultimately the goal of this model will be to provide guidelines about how autonomous systems can utilize data gathered from verbal and non-verbal means to disambiguate information about the operational space and execute plans of action in complex and dynamic environments. Moving forward, matching human models of communication with engineering framework models of autonomous processing will provide the basis to begin testing new models of human-robot team communication.

References

1. Intent. (n.d.); merriam-webster.com. Retrieved February 21, 2016, from <http://www.merriam-webster.com/dictionary/intent>
2. Wang, J., Lewis, M.: Human control for cooperating robot teams. In: 2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 9–16. IEEE, Mar 2007
3. Chen, J.Y., Barnes, M.J.: Human-agent teaming for multirobot control: a review of human factors issues. *IEEE Trans. Hum. Mach. Syst.* **44**(1), 13–29 (2014)
4. Doare, R., Danet, D., Hanon, J.P.: Robots on the battlefield: contemporary issues and implications for the future. Maroon Ebooks (2014)
5. Chen, T., Campbell, D., Gonzalez, F., Coppin, G.: The effect of autonomy transparency in human-robot interactions: a preliminary study on operator cognitive workload and situation awareness in multiple heterogeneous UAV management. In: Proceedings of Australasian Conference on Robotics and Automation 2014, pp. 1–10. Australian Robotics & Automation Association ARAA, Dec 2014

6. Zhang, Y., Narayanan, V., Chakraborti, T., Kambhampati, S.: A human factors analysis of proactive support in human-robot teaming. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2015)
7. Chen, J.Y., Barnes, M.J., Qu, Z.: RoboLeader: an agent for supervisory control of multiple robots. In: *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 81–82. IEEE Press, Mar 2010
8. Schramm, W.: How communication works. In: Schramm, W. (ed.) *The Process and Effects of Communication*, pp. 3–26. University of Illinois Press, Champaign (1954)
9. Berlo, D.K.: *The Process of Communication: An Introduction To Theory And Practice*. Holt, Rinehart and Winston, New York (1960)
10. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. *Human Factors J. Hum. Factors Ergon. Soc.* **37**(1), 32–64 (1995)
11. Salas, E., Prince, C., Baker, D.P., Shrestha, L.: Situation awareness in team performance: implications for measurement and training. *Hum. Factors J. Hum. Factors Ergon. Soc.* **37**(1), 123–136 (1995)
12. Salas, E., Cooke, N.J., Rosen, M.A.: On teams, teamwork, and team performance: discoveries and developments. *Hum. Factors J. Hum. Factors Ergon. Soc.* **50**(3), 540–547 (2008)
13. Salas, E., Dickinson, T.L., Converse, S.A., Tannenbaum, S.I.: *Toward an understanding of team performance and training* (1992)
14. Converse, S.: Shared mental models in expert team decision making. *Individ. Group Decis. Making Current* **1993**, 221 (1993)
15. Lim, B.C., Klein, K.J.: Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy. *J. Organ. Behav.* **27**(4), 403–418 (2006)
16. Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., Cannon-Bowers, J.A.: The influence of shared mental models on team process and performance. *J. Appl. Psychol.* **85**(2), 273 (2000)
17. Rouse, W.B., Cannon-Bowers, J.A., Salas, E.: The role of mental models in team performance in complex systems. *IEEE Trans. Syst. Man Cybern.* **22**(6), 1296–1308 (1992)
18. Stout, R.J., Cannon-Bowers, J.A., Salas, E., Milanovich, D.M.: Planning, shared mental models, and coordinated performance: An empirical link is established. *Hum. Factors J. Hum. Factors Ergon. Soc.* **41**(1), 61–71 (1999)
19. Gurtner, A., Tschan, F., Semmer, N.K., Nägele, C.: Getting groups to develop good strategies: effects of reflexivity interventions on team process, team performance, and shared mental models. *Organ. Behav. Hum. Decis. Process.* **102**(2), 127–142 (2007)
20. Nikolaidis, S., Shah, J.: *Human-robot teaming using shared mental models*. ACM/IEEE HRI (2012)
21. Clark, H.H.: *Using Language*. Cambridge University Press, Cambridge (1996)
22. Clark, H.H., Brennan, S.E.: Grounding in communication. *Perspect. Soc. Shared Cogn.* **13** (1991), 127–149 (1991)
23. Bohus, D., Horvitz, E.: On the challenges and opportunities of physically situated dialog. In: *AAAI Fall Symposium: Dialog with Robots*, Nov 2010
24. Lemon, O., Bracy, A., Gruenstein, A., Peters, S.: The WITAS multi-modal dialogue system I. In: *INTERSPEECH*, pp. 1559–1562, Sept 2001
25. Marge, M., Pappu, A., Frisch, B., Harris, T.K., Rudnicki, A.I.: Exploring spoken dialog interaction in human-robot teams. In: *Proceedings of Robots, Games, and Research: Success Stories in USARSim IROS Workshop* (2009)
26. Marge, M., Rudnicki, A.I.: Miscommunication recovery in physically situated dialogue. In: *Proceedings of SIGdial* (2015)
27. Skantze, G.: Exploring human error recovery strategies: implications for spoken dialogue systems. *Speech Commun.* **45**(3), 325–341 (2005)
28. Streeck, J., Knapp, M.L.: The interaction of visual and verbal features in human communication. *Adv. Nonverbal Commun.* 3–23 (1992)
29. Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., Walker, J.: Are children with autism blind to the mentalistic significance of the eyes? *Br. J. Dev. Psychol.* **13**(4), 379–398 (1995)

30. Hoffman, M.W., Grimes, D.B., Shon, A.P., Rao, R.P.: A probabilistic model of gaze imitation and shared attention. *Neural Netw.* **19**(3), 299–310 (2006)
31. Lockerd, A., Breazeal, C.: Tutelage and socially guided robot learning. In: 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 2004. Proceedings, vol. 4, pp. 3475–3480. IEEE, Sept 2004
32. Strabala, K.W., Lee, M.K., Dragan, A.D., Forlizzi, J.L., Srinivasa, S., Cakmak, M., Micelli, V.: Towards seamless human-robot handovers. *J. Hum. Robot Interact.* **2**(1), 112–132 (2013)
33. Huang, C.M., Andrist, S., Sauppé, A., Mutlu, B.: Using gaze patterns to predict task intent in collaboration. *Front. Psychol.* **6** (2015)
34. Handwerk, B.: 5 Surprising drone uses (Besides Amazon Delivery), National Geographic. Retrieved March 2, 2016 from <http://news.nationalgeographic.com/news/2013/12/131202-drone-uav-uas-amazon-octocopter-bezos-science-aircraft-unmanned-robot/>
35. Charniak, E., Goldman, R.P.: Probabilistic Abduction for Plan Recognition. Brown University, Department of Computer Science (1991)
36. Sakita, K., Ogawara, K., Murakami, S., Kawamura, K., Ikeuchi, K.: Flexible cooperation between human and robot by interpreting human intention from gaze information. In: 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 2004. Proceedings, vol. 1, pp. 846–851. IEEE, Sept 2004
37. Park, H.S., Jain, E., Sheikh, Y.: 3d social saliency from head-mounted cameras. In: Advances in Neural Information Processing Systems, pp. 431–439 (2012)
38. Soo Park, H., Shi, J.: Social saliency prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4777–4785 (2015)
39. Hopcroft, J., Motwani, R., Ullman, J.: Introduction to Automata Theory, Languages, and Computation. Pearson, London (2013)
40. Matuszek, C., Herbst, E., Zettlemoyer, L., Fox, D.: Learning to parse natural language commands to a robot control system. In: Proceedings of the International Symposium on Experimental Robotics, pp. 403–415 (2013)
41. Dantam, N., Stilman, M.: The motion grammar: analysis of a linguistic method for robot control. *IEEE Trans. Robot.* **29**(3), 704–718 (2013)
42. Kress-Gazit, H., Fainekos, G.E., Pappas, G.J.: Temporal-logic-based reactive mission and motion planning. *IEEE Trans. Robot.* **25**(6), 1370–1381 (2009)
43. Parks, D., Borji, A., Itti, L.: Augmented saliency model using automatic 3d head pose detection and learned gaze following in natural scenes. *Vision. Res.* **116**, 113–126 (2015)

Human-Autonomy Teaming Using Flexible Human Performance Models: An Initial Pilot Study

Christopher J. Shannon, David C. Horney, Kimberly F. Jackson
and Jonathan P. How

Abstract Recent advances in autonomy have highlighted opportunities for tight coordination between humans and autonomous agents in many current and future applications. For operations involving cooperating humans and robots, autonomous teammates must have flexibility to respond to the inherently unpredictable behavior of their human counterparts. To investigate this issue in detail, this paper uses an unmanned aerial vehicle (UAV) simulation to evaluate flexible human performance models over traditional static modeling approaches for multi-agent task allocation and scheduling. Additional comparisons are drawn between adaptive human models, which are adjusted in real-time by the autonomous planner according to realized human performance, and adaptable human models, where the human operator is given sole authority over model adjustments. Results indicate that adaptive human performance models significantly increase total mission reward over both the baseline static modeling framework ($p = 0.0012$) as well as the adaptable modeling technique ($p = 0.0028$) for this system.

Keywords Human-Autonomy teaming · Multi-Agent planning · Human performance modeling · Adaptive and adaptable autonomy

C.J. Shannon (✉) · D.C. Horney · J.P. How
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: cshannon@mit.edu

D.C. Horney
e-mail: dchorney@mit.edu

J.P. How
e-mail: jhow@mit.edu

K.F. Jackson
Draper, Cambridge, MA, USA
e-mail: kjackson@draper.com

1 Introduction

Traditional approaches to multi-agent planning typically view agents as independent, autonomous vehicles. However, a variety of modern and future applications require highly coordinated efforts between humans and autonomous systems [1]. Planning for humans in the loop presents several key challenges, including the difficulties associated with adequately modeling actual human capabilities. In particular, reliable prediction of human task performance can be hard, and even well-established models from the human factors community are subject to inter-personal differences—such as experience and skill—as well as dynamic intrapersonal factors like distraction and fatigue [2]. These challenges motivate human performance models that can be updated in real-time to ensure better consistency between predictive planning and actual execution.

A promising strategy for human modeling is to use an adaptive approach in which actual human performance is leveraged as feedback for autonomous adjustments to models throughout mission execution. Adaptive autonomy has been steadily gaining traction in human-robot interaction (HRI) applications over recent years [3]. State-of-the-art autonomous systems that are able to reason on, and respond to, the needs and intentions of a person present new opportunities for close synergy between people and autonomy [4]. Recent studies have produced planners that enable a single robot to dynamically execute a shared plan with a human [5], while other efforts have developed intelligent task management strategies for human operators that regulate and optimize over human workload parameters [6, 7]. Additionally, adaptive levels of automation have shown promise for human supervisory control of multiple robots [8].

Apart from feedback based on realized human performance, an adaptable autonomy construct can provide information to the planner directly from the user. Examples include actively changing mission objective functions in real-time [9] as well as adjusting levels of automation [8]. This can enhance the user’s control of the environment, which has been shown to help in terms of change detection, system transparency, and overall user satisfaction within a variety of domains [10]. However, since adaptable autonomy requires deliberate user feedback, it can increase workload and draw attention away from the main task at hand. In addition, relying on human interaction may be less optimal than a well-designed adaptive framework [3].

Drawing from these strategies, this pilot study focuses on an evaluation of adaptive and adaptable human performance models for the allocation and scheduling of joint human-robot tasks amongst a heterogeneous, dynamic human-robot team. Models for the human operators’ predicted task durations and workload thresholds are used to investigate the advantages and drawbacks of adaptive and adaptable human modeling in the context of these highly coordinated multi-agent systems. Results from the human subject experiment ($n = 12$) provide insights into the use of flexible human performance models for human-autonomy teaming.

2 Background

For this pilot study, the human participant’s role within the multi-agent mission is to classify a series of discrete pieces of imagery. These imagery classifications are time-critical: the values of accurate imagery information acquisition are functions of time (e.g. time-windows and exponential time-decays). Therefore, the amount of time an operator needs for correct classification and the rate at which he can be presented with new demands are important for the allocation and scheduling of tasks.

Figure 1 shows the two human performance models used in this pilot study. Pew’s model (Fig. 1a) predicts that the correctness of a decision in a binary choice increases as a sigmoidal function of time spent on the task [11]. In order to predict or allocate task durations, probability of success may be optimized [7] or chosen (as in this case) depending on requirements for speed versus accuracy. The workload threshold model (Fig. 1b) moderates the rate at which tasks appear by incurring a cost on proposed schedules that exceed the prescribed threshold. This model, derived from the Yerkes-Dodson Law [13], uses the *utilization* metric defined as operator busy time over total mission time in order to keep the human at a desirable level of workload [12].

In our recent work, we have created a fast algorithm called ASSIST that integrates humans into the traditional task allocation and scheduling problem to produce plans for tightly coupled (i.e. synchronized joint-task) heterogeneous human-robot teams [14]. ASSIST can be embedded into a closed-loop replanning framework—in which humans are treated as dynamic agents—in order to adapt human and environment models in real-time. This pilot study evaluates the ASSIST planning framework and its various feedback mechanisms through actual human-autonomy teaming experiments.

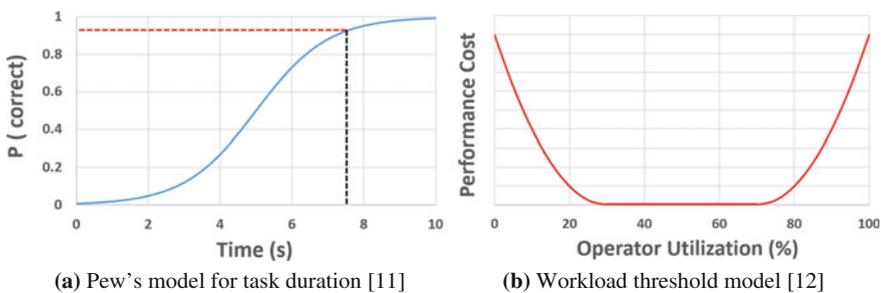


Fig. 1 Human performance models for operator’s imagery classification tasks. **a** Pew’s model for task duration [11]. **b** Workload threshold model [12]

3 Aim of the Pilot Study

The quantitative hypotheses tested in this experiment compare static (baseline open-loop approach), adaptive, and adaptable modeling strategies using an *overall mission reward* metric. This reward is calculated as the summation of all successfully completed tasks' values at their respective execution times, and incorrect imagery classifications result in zero reward for a task. Thus overall reward represents a measure of timely, accurate task completion as well as efficiency in the allocation and scheduling of the tasks among the multi-agent team to optimize execution in terms of specified mission goals. Subjective trends from post-experiment participant questionnaires—plus other recorded measures such as classification accuracy, mission makespan (total duration), and total vehicle distance traveled—supplement the evaluation.

We hypothesize that adaptive human performance modeling will result in the most overall mission reward, with adaptable modeling resulting in the second most reward, and static modeling resulting in the least. Autonomous adjustments of the models based on realized performance from actual mission execution in real-time should alleviate mismatches between the planning phase's models and the humans' actual ability. For the adaptable mode, the user's adjustment of these models should also mitigate model mismatches, albeit less efficiently than the autonomous adaptive mode. Additionally, the baseline static modeling case will likely illuminate the issues of rigid, one-size-fits-all open-loop approaches to modeling human agents.

We also hypothesize that the adaptable human performance modeling strategy will be the most preferred by users, with adaptive modeling being the second most preferred, and static modeling being the least preferred. We predict that giving the human operator authority over the models that predict his own performance (and thus drive the allocation and scheduling of his tasks) will result in higher user satisfaction than if the adjustments were made by the autonomous system instead. We further predict that the adaptive mode will be preferable to the static mode due to multi-agent plans that have flexibility and are more conformed to each participant's actual ability.

4 Experimental Methods

The human-autonomy teaming mission consists of discrete surveillance tasks in distinct locations throughout a map. Points of interest may have differing priorities or time-criticalities which correspond to their respective task values. The ASSIST planner allocates and schedules coupled human-robot pairs to the tasks, optimizing plans for maximum task value while adhering to human operator heuristics and constraints. The architecture uses the Robot Operating System [15] for a high-fidelity multi-vehicle simulation. Participants interact with a standard laptop

(2 GHz Intel Core i7 processor, 8 GB RAM) using a keyboard and mouse to accomplish missions.

4.1 User Interface

The full interface is illustrated in Fig. 2. Multi-agent plans from the ASSIST algorithm are executed by the human-robot team with the UAVs flying to task locations and providing imagery to the human operator. The top right portion of the interface provides an overhead mission view that allows the operator to anticipate vehicles arriving to their specified survey waypoints. The top-left command line prompts the user to alert him to incoming imagery. The bottom of the interface shows the live, top-down video feeds from simulated on-board UAV cameras that present the imagery to be classified.

The participant performs classification by inputting into the command line “c” for circle imagery or “s” for square imagery as well as the count of moving images. In all modes, this imagery flashes yellow to give the human operator an indication as to when he is exceeding the duration allotted for the task by the planner. Overall mission reward accumulates with timely, accurate completion of the joint human-robot tasks. The user continues the process of classifying images throughout the duration of the mission, cross-checking the overhead view for increased situational awareness and looking to the command line for incoming tasks and feedback on his performance.

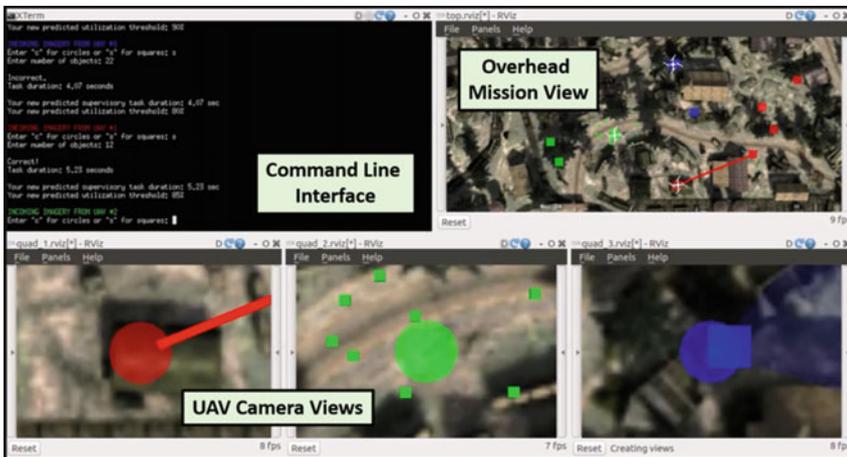


Fig. 2 Full user interface for human-robot teaming mission

4.2 *Experiment Conditions*

There are three conditions in this experiment which relate to the modeling of human performance. Static modeling is a baseline control condition that uses open-loop modeling of humans with planning parameters set a priori. Adaptive modeling leverages actual human performance as feedback throughout each mission to autonomously adjust models. Finally, adaptable modeling uses inputs from the human operator to adapt his performance models during the mission. Replanning is conducted at 15 s intervals in all modes with their respective models.

Static Modeling. The static modeling mode provides a baseline open-loop approach to the representation of human performance. In this mode, humans are considered to be homogeneous agents whose predicted task performance is equivalent among different individuals. Moreover, they are assumed to be static, modeled by rigid functions throughout the length of the mission. As illustrated in Fig. 1, Pew's model for predicted task duration and the utilization model for projected workload threshold are specified a priori. For this pilot study, predicted task duration was set to 6.4 s, derived from the average length of time required for novice operators to complete a task in preliminary testing. Projected workload threshold was set to 70 %, consistent with previous studies involving human operators and multiple UAVs [12] and aligning with subjective responses from novice operators interacting with the current system.

Adaptive Modeling. The adaptive modeling mode leverages realized human performance throughout mission execution to adjust the planner's underlying models in real-time. Human operators are treated as heterogeneous, dynamic agents in that the specified models are adapted to fit the skills of each individual and remain flexible to track actual human performance throughout the mission. Pew's model is initially set to 6.4 s for predicted task duration and the utilization model originally specifies 70 % as the operator workload threshold. These values are then updated in real-time based on actual performance by the operator.

The two human performance models are autonomously adapted by the closed-loop system after each task. The actual time from the imagery's appearance to the operator's task completion becomes the new predicted task duration for future human-robot surveillance tasks. If the operator's inputs are correct, the utilization workload threshold is increased by 5 %, functioning as a heuristic to allow more demand on a successful operator. If the classification or count is incorrect, the utilization threshold is decreased by 15 %, decreasing overall workload on the operator in the near future. In this mode, replanning not only alleviates discrepancies between the current system state and past projections, but also leverages up-to-date realized agent performance to generate plans that more closely mirror the actual behavior of the multi-agent system. This closed-loop modeling strategy alleviates model mismatches arising from heterogeneous human operators (e.g. a model that is appropriate for one person may be ill-fitting for another) as well as dynamic, stochastic human performance.

Adaptable Modeling. As an alternative to the autonomous system providing feedback of realized performance to the planner, an adaptable mode was developed in which the human operator had sole authority over the adjustment of his performance models. This strategy shifts responsibility of model updates from the autonomy to the human operator in attempts at providing insights into both the objective effects on multi-agent teaming as well as the subjective perceptions of the user.

Adaptable adjustments are accomplished through the use of additional operator inputs into the user interface. First, each human-robot surveillance task requires the usual “c” or “s” classification input along with specifying the number of images. Following these two inputs, the command line prompts the subject to “Enter ‘z’ if too busy, ‘r’ if too bored (else, hit enter).” Inputting “z” decreases the operator utilization workload threshold by 10 %, while “r” increases it by 10 %. The command line then prompts the operator to “Enter ‘l’ for less allotted task time, ‘m’ for more allotted task time (else, hit enter).” Entering “l” shifts Pew’s model to decrease predicted task duration by three seconds, whereas “m” moves it in the opposite direction to increase predicted task duration by three seconds. These discrete adjustment values were chosen from a range of possibilities upon demonstrating satisfactory objective and subjective performance in preliminary testing with novice users.

4.3 Procedure

Twelve individuals were selected to participate in the pilot study. Each experiment began with the subject reading an instruction document that explained the study’s purpose, the responsibilities of the participant, and the interaction procedures for the user interface. Each subject was then guided through two full mission trials (with three autonomous vehicle teammates and 15 surveillance tasks) for training purposes. The first practice trial was conducted in the adaptive human performance modeling mode to allow the subject to become familiar with his primary task. The second practice trial was performed in the adaptable human performance modeling mode in order for the subject to become familiar with providing feedback to the planner. Upon completion of the two practice missions, the participant completed a human-robot teaming mission in each of the static, adaptive, and adaptable human performance modeling modes. The ordering of modes was counterbalanced among the twelve participants (using the six possible combinations of mode ordering for two participants each) in order to mitigate potential confounding variables of learning and fatigue. The process required less than an hour of each participant’s time, and all testing took place over a two week period.

Objective mission metrics were recorded throughout each trial. Total makespan (or mission duration) and total vehicle distance traveled were noted. The start time, duration, and accuracy of each task was recorded and used to compute overall mission reward, average operator accuracy, and average task duration. All model

shifting inputs in the adaptable mode (“too busy,” “too bored,” “less time,” “more time”) and their associated input times were also logged. This level of detail into all relevant aspects of the mission allowed analysis of underlying reasons for human-robot team performance trends, such as temporally overloaded operators, model discrepancies between the planning phase and mission execution, and dynamic human performance.

Upon completing all trials, each participant filled out a questionnaire. The survey was used to evaluate situational awareness, team fluency, user interface satisfaction, and human performance modeling mode preference. Specific questions relating to the feedback modes (e.g. which was “most efficient in helping you complete the task” and which mode “you most enjoyed using”) were included in the survey. In order to garner as much information as possible from the pilot study, all questions and choices were open-ended and prompted the subject to specify why they felt a certain way. Participants’ subjective inputs provided insight into trends in the objective metrics and aided in future system development after pilot study completion.

5 Results

After all testing was completed, data were investigated both qualitatively and quantitatively to evaluate the effects of the various human modeling strategies on team performance and operator satisfaction. Primarily, statistical analysis of overall mission rewards and evaluations of user mode preferences were used to compare the three modes.

5.1 Objective Analysis

A one-way balanced ANOVA F-test was used to test whether there exists any statistically significant differences among the three modes’ associated mission rewards (Fig. 3). Based on the data and using a 95 % confidence interval, there is reason to believe that there exists significant variation between the three modes in terms of overall mission reward ($F_{2,33} = 8.645$, $p < 0.001$). Therefore, direct comparisons between each pair of modes using post hoc paired t-tests with Bonferroni corrections [16] are appropriate.

Comparing adaptive human performance modeling against the baseline static approach shows that average mission reward for the adaptive mode ($M = 8116.25$, $SD = 933.80$) is significantly different than the static mode’s ($M = 6649.21$, $SD = 1372.39$) average reward ($t(11) = 4.353$, $p = 0.0012$). With mean overall reward being much higher for the adaptive case, results indicate that adaptive human performance models significantly increase total mission reward over the static modeling framework in the context of this experiment. A second paired t-test

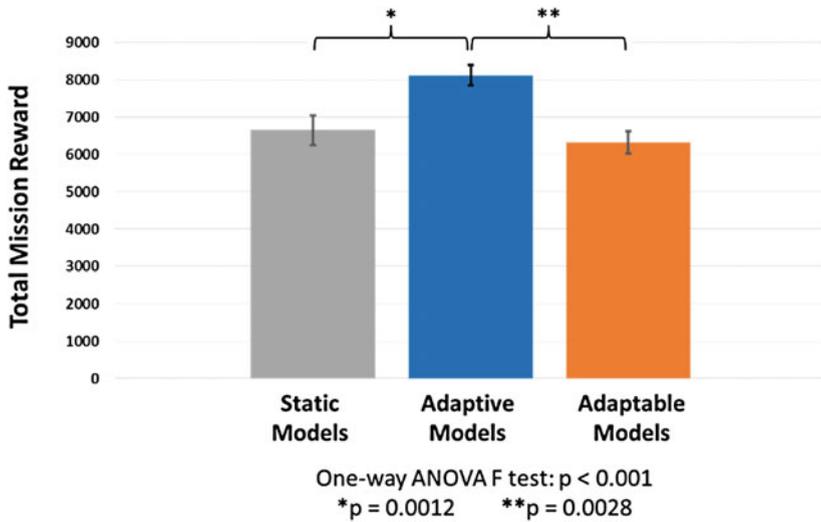


Fig. 3 Overall mission rewards between human modeling modes with standard error bars

comparing adaptable human operator modeling against the baseline static modeling strategy gives no indication that mission reward from adaptable modeling ($M = 6311.36$, $SD = 1036.45$) is significantly distinct from reward with static modeling ($M = 6649.21$, $SD = 1372.39$) in this pilot study ($t(11) = 0.631$, $p > 0.05$). A final paired t-test examines adaptive and adaptable human performance modeling, demonstrating that adaptive human performance models ($M = 8116.25$, $SD = 933.80$) result in statistically significant improvement of overall mission reward in comparison to adaptable models ($M = 6311.36$, $SD = 1036.45$) for this experiment ($t(11) = 3.826$, $p = 0.0028$).

The data suggest that adaptive human performance models can promote more effective coordination between humans and autonomous agents than traditional open-loop modeling approaches. Leveraging feedback from actual human performance can alleviate model mismatches at both the system level (e.g. workload thresholds for difficult vs. routine missions) and the individual level (e.g. different experience and skill between operators). Assuming humans to be homogeneous agents with one-size-fits-all models can be detrimental to overall team performance. In fact, average task duration varied by as much as 87.5 % between participants in this pilot study. A histogram of participants' average task durations in Fig. 4 shows a Gaussian-like distribution with substantial variance between subjects.

The advantages of adaptive modeling for heterogeneous human agents are further illustrated in Fig. 5, which shows mission timelines for a single participant in each of the three modes. This subject generally accomplishes tasks more quickly than the predicted task duration of 6.4 s set a priori, which causes a model mismatch for the static open-loop modeling approach. The adaptive human performance modeling mode, on the other hand, is able to adjust to the operator's realized

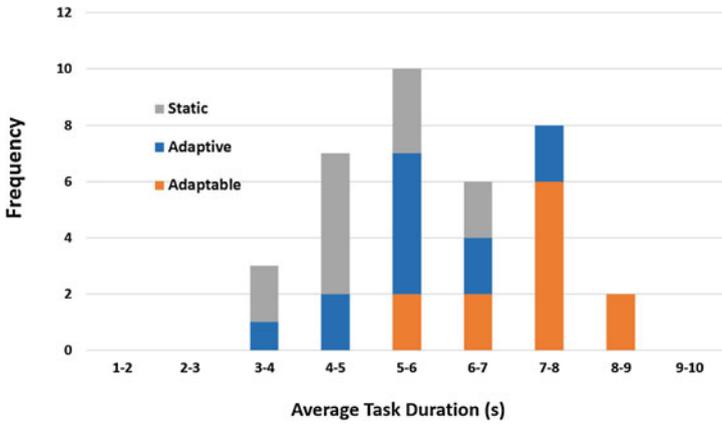


Fig. 4 Histogram of average task duration for each mission trial

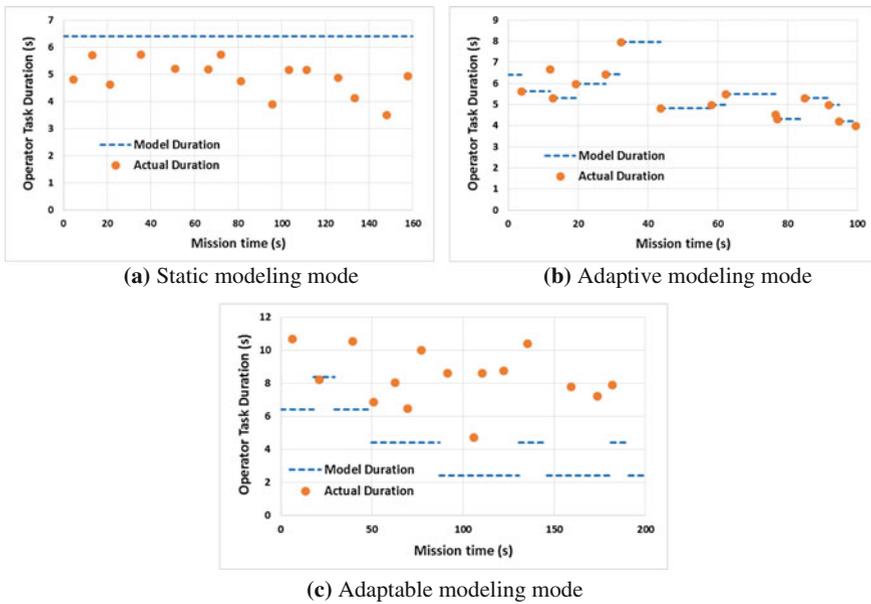


Fig. 5 Planning phase’s predicted human operator task duration from underlying models versus actual task duration over the course of one participant’s trials in each mode. **a** Static modeling mode. **b** Adaptive modeling mode. **c** Adaptable modeling mode

performance once the mission begins. As an additional advantage for adaptive modeling, the human’s dynamic behavior during this trial—which takes the form of decreasing durations as the mission progresses—is tracked by the closed-loop system.

Results also indicate (within the context of this system) that incorporating real-time autonomous adjustments of human performance models provides quantitative advantages over relying on human input for adjustments. This finding may be explained as the result of two main causes. First, the adaptable mode required slightly more workload than the static and adaptive modes. Figure 4 shows the adaptable trials favor the right side of the distribution, and a paired t-test confirms that the adaptable mode's additional inputs resulted in significantly longer average task durations ($M = 7.246$, $SD = 0.997$) over the adaptive approach ($M = 5.617$, $SD = 0.986$) in this pilot study ($t(11) = 3.745$, $p = 0.00324$). Second, the human operator may be worse than the autonomous system at adjusting his model appropriately to minimize model mismatches between the planning phase and actual mission execution. Figure 5c illustrates that the subject fails to track actual performance accurately. Even more so, at times throughout the trial his adjustments increase model discrepancies rather than mitigate them.

Other aspects of human-robot team performance among the feedback modes were analyzed to supplement comparisons of overall mission reward. Mission makespan, or total mission duration, indicates that adaptive human performance modeling significantly reduces total mission time ($F_{2,33} = 10.729$, $p < 0.001$) over both baseline static ($t(11) = 6.710$, $p < 0.001$) and adaptable modeling approaches ($t(11) = 4.563$, $p < 0.001$). This further signifies adaptive modeling's ability to mitigate model discrepancies between the planning and execution phases, resulting in more efficient task allocations and schedules amongst the human-robot team. Additionally, overall vehicle distance was longer in the adaptable modeling mode than both the static ($t(11) = 1.976$, $p = 0.0738$) and adaptive ($t(11) = 2.012$, $p = 0.0594$) strategies ($F_{2,33} = 2.656$, $p < 0.085$), pointing again to suboptimal multi-agent plans for the adaptable case.

5.2 Subjective Measures

Post-experiment questionnaires allowed subjects to provide qualitative feedback on their perceptions of the autonomous system, the user interface, and the various human performance modeling modes. All participants provided positive comments on the general system, specifically citing the efficiency of the interface under temporal pressure. In addition, real-time feedback was said to be especially useful in terms of helping operators stay motivated throughout the mission and allowing them to self-correct after errors. Finally, subjects enjoyed having the overhead view available for team-wide situational awareness and projections of incoming surveillance tasks.

Participant responses to questions on mode preference align with objective findings, as 92 % of subjects chose adaptive human performance modeling as the "most efficient" mode of the three tested. Additionally, 83 % of users stated that they "most enjoyed using" the adaptive scheme, describing it as "stimulating," "well-tuned," and "most engaging." This autonomous closed-loop modeling

approach was generally perceived as being more effective than the baseline static case, and its lower workload requirements relative to the adaptable strategy allowed users to focus on the primary classification task without distraction.

6 Discussion

The increase in overall mission reward for adaptive human performance modeling over both the baseline static case and the adaptable approach may be attributed to the adaptive strategy's ability to minimize model mismatches between the planning phase and actual human ability. Proper prediction of human operator performance within the human-robot teaming missions is important for effective multi-agent planning towards achieving specified mission goals. The ASSIST algorithm is allocating and scheduling tasks amongst the vehicles with the aim of maximizing mission reward. Again, this reward metric is calculated by totaling all successfully completed tasks' values at their execution times, and these time-varying task score functions may take various functional forms.

When the planning phase over-predicts operator task duration (Fig. 5a), ASSIST imposes a longer scheduling constraint on future tasks. This may lead to allocations in which other vehicles travel longer distances in order to reach higher-value tasks around the time the human is expected to become available. However, if the operator finishes the task more quickly than predicted, he may then be forced to wait idly for an undesirable amount of time before the next task arrival. If, on the other hand, the planning phase under-predicts operator task duration (Fig. 5c), ASSIST may allocate tasks in an overly-optimistic fashion. A vehicle may be required to stay committed to a task longer than expected due to the human operator, thus resulting in the vehicle arriving later-than-expected to subsequent tasks. This can lead to both inefficient scheduling throughout the multi-agent team (e.g. two vehicles each arriving to their tasks at the same time, requiring one to wait an extended period of time for the human) as well as "missed rendezvous" (e.g. missing a task time window).

In addition to task durations, predicting the human operator's workload threshold is important for the heuristics of the algorithm. Ideally, an operator would be able to achieve 100 % accuracy on imagery classification tasks while remaining as busy as possible in order to reduce scheduling constraints on task execution times and minimize overall mission duration. With this in mind, adaptive workload thresholds were increased following accurate imagery classifications to reflect the fact that the operator was able to succeed at the prescribed workload. Upon task failures, adaptive workload thresholds were decreased to allow the user to gather himself for subsequent demands. Instead of this heuristic approach, future experiments will incorporate secondary tasks to analyze and adjust the operator's workload threshold in a more direct fashion.

In terms of subjective data, the majority of participants preferring adaptive modeling over the adaptable approach is contrary to original predictions which

assumed that providing the human operator with more control over the system would result in greater user satisfaction. However, these results align with other findings in the literature. Specifically, Gombolay et al. [17] showed that humans generally prefer to work within an efficient team rather than have a heightened role in the planning process if that increased role is detrimental to overall team performance. In the current pilot study, allowing human operators to adjust their own performance models imposes additional workload requirements and generally produces larger model mismatches and thus less efficient multi-agent plans.

7 Conclusions

Closed-loop planning approaches in which the autonomous system adapts flexible human models throughout mission execution can improve overall mission performance for tightly coordinated human-robot teams. The challenges of unpredictable, heterogeneous human agents with dynamic, stochastic behavior can be addressed through responsive replanning that leverages realized execution information. Also, while adaptable human modeling failed to demonstrate improved mission performance over the baseline open-loop approach, future work concerning efficient input methodologies and comprehensive interface feedback—particularly as it relates to model mismatches—may reveal added benefits.

This human-robot teaming pilot study highlights many potential avenues for future work, such as testing longer missions to capture more dynamic human performance and utilizing more comprehensive subjective questionnaires and workload assessments. Additional work could be done to test more complicated feedback modes, such as improving adaptable adjustment mechanisms and incorporating stochastic averaging [18], filtering [19], and learning techniques [20] into the adaptive approach. A hybrid “adaptablive” mode could be created to investigate whether building on both adaptive and adaptable strengths can be synthesized. This could also distinguish effects of additional workload versus model discrepancies to help explain the adaptive mode’s advantageous performance over the adaptable strategy in this work.

Acknowledgments The authors would like to thank Draper (Cambridge, MA, USA) for funding this research as well as Professor Julie Shah for her insight and guidance throughout this effort.

References

1. Unmanned Systems Integrated Roadmap: FY2013-2038. Washington, DC, USA (2013)
2. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multirobot control: a review of human factors issues. *IEEE Trans. Human-Mach. Syst.* **44**, 13–29 (2014)
3. Kaber, D.B., Prinzl, L.J.: Adaptive and Adaptable Automation Design: A Critical Review of the Literature and Recommendations for Future Research (2006)

4. Alami, R., Clodic, A., Chatila, R., Lemaignan, S.: Reasoning about humans and its use in a cognitive control architecture for a collaborative robot. In: *Human-Robot Interaction Workshop* (2014)
5. Shah, J.A., Wiken, J., Williams, B.C., Breazeal, C.: Improved human-robot team performance using Chaski, a human-inspired plan execution system. In: *6th International Conference on Human-Robot Interaction*, pp. 29–36 (2011)
6. Savla, K., Frazzoli, E.: A dynamical queue approach to intelligent task management for human operators. *Proc. IEEE* **100**, 672–686 (2012)
7. Srivastava, V., Surana, A., Bullo, F.: Adaptive attention allocation in human-robot systems. In: *American Control Conference* (2012)
8. Kidwell, B., Calhoun, G.L., Ruff, H.A., Parasuraman, R.: Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **56**, 428–432 (2012)
9. Clare, A.S., Cummings, M.L., How, J.P., Whitten, A.K., Toupet, O.: Operator object function guidance for a real-time unmanned vehicle scheduling algorithm. *J. Aerosp. Comput. Inf. Commun.* **9**, 161–173 (2012)
10. Sauer, J., Kao, C.-S., Wastell, D.: A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics* **55**, 840–853 (2012)
11. Pew, R.W.: The speed-accuracy operating characteristic. *Acta Psychol. (Amst)* **30**, 16–26 (1969)
12. Nehme, C.E.: *Modeling Human Supervisory Control in Heterogeneous Unmanned Vehicle Systems* (2009)
13. Yerkes, R.M., Dodson, J.D.: The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* **18**, 459–482 (1908)
14. Shannon, C.J., Johnson, L.B., Jackson, K.F., How, J.P.: Adaptive mission planning for coupled human-robot teams. In: *American Control Conference* (2016)
15. Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., Berger, E., Wheeler, R., Ng, A.: ROS: an open-source robot operating system. *ICRA Work.* **3** (2009)
16. Weisstein, E.W.: Bonferroni Correction (2004)
17. Gombolay, M.C., Gutierrez, R.A., Sturla, G.F., Shah, J.A.: Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams. *Robot. Sci. Syst. X.* (2014)
18. Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T.: Bayesian model averaging: a tutorial. *Stat. Sci.* **14**, 382–417 (1999)
19. Bertuccelli, L.F., Choi, H., Cho, P., How, J.P.: Real-time multi-UAV task assignment in dynamic and uncertain environments. In: *AIAA Conference on Guidance, Navigation, and Control* (2009)
20. Bishop, C.M.: *Pattern Recognition. Machine Learning* (2006)

Self-scaling Human-Agent Cooperation Concept for Joint Fighter-UCAV Operations

Florian Reich, Felix Heilemann, Dennis Mund and Axel Schulte

Abstract In this article, we describe human automation integration concepts that allow the guidance and the mission management of multiple UCAVs (Unmanned Combat Aerial Vehicles) from aboard a manned single-seat fighter aircraft. The conceptual basis of our approach is dual-mode cognitive automation. This concept uses two distinct modes of human-agent cooperation, a hierarchical relationship with agents working in delegation mode, and a heterarchical relationship with an agent working in assistance mode. For the hierarchical relationship we suggest three delegation modes (team-, intent-, and task-based). The agent in heterarchical relationship, i.e. the assistant system, adapts the operator-assistant system cooperation and the guidance of UCAVs according to the named delegation modes. The adaptation is shaped by the assessment of the operator's mental state and external situation features. Thereby, we aim at balancing the operator's activity and work demands. Future research at our institute will concentrate on developing a software prototype for human-in-the-loop experiments.

Keywords Self-scaling automation · Human-agent cooperation · Dual-mode cognitive automation · Assistant system · Multiple UCAV guidance · Delegation modes · Operator-centered automation adaption

F. Reich (✉) · F. Heilemann · D. Mund · A. Schulte
Institute of Flight Systems (IFS), Universität der Bundeswehr Munich (UBM),
Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
e-mail: florian.reich@unibw.de

F. Heilemann
e-mail: felix.heilemann@unibw.de

D. Mund
e-mail: dennis.mund@unibw.de

A. Schulte
e-mail: axel.schulte@unibw.de

1 Introduction

In this article, we present our concept of self-scaling human-agent cooperation, which is supposed to enable a single-seat fighter pilot to perform joint fighter-UCAV missions with multiple UCAVs. In such operations, the pilot's range of responsibility includes the operation of the own fighter aircraft as well as the guidance and the mission management of multiple cooperating UCAVs in a dynamic air warfare environment. As a consequence, the task of flying the own fighter aircraft is in conflict with the operation of the UCAVs. State of the art armed medium-altitude long-endurance unmanned aerial vehicles such as the MQ9-Reaper (General Atomics) are operated by a two-man team, consisting of a pilot and a sensor and weapon operator [1]. The operation of multiple UCAVs from aboard a single-seat fighter aircraft therefore requires an appropriate automation concept for effective joint operations, since the UAV to operator ratio increases considerably. Our automation integration concept is supposed to enable such operations and, beyond this, to balance the operator's activity-related work demands, and to address automation-induced shortfalls in human-system interaction.

In the following section, we first summarize the findings from former own studies on multiple UCAV operations with highly-automated agents and joint fighter-UCAV operations. Then, we identify key issues of human-machine system design. Chapter 4 gives a detailed overview of our concept. In Chap. 5 we analyze our approach with regard to similar international work. Finally, in Chap. 6, we conclude and provide an outlook on open conceptual issues.

2 Background

The Institute of Flight Systems (IFS) investigates joint fighter-UCAV operations in future air warfare scenarios, in which we expect a mix of manned and unmanned combat aircraft to be deployed (compare [2]). In the first phase of our research, we mainly examined the capabilities of highly automated cognitive and cooperative agents operating multiple UCAVs in joint air-to-ground attack missions. For this purpose, a desktop simulation was developed in that artificial cognitive agents autonomously operated the associated UCAVs on the basis of high-level abstract goals, without the need of further human involvement [3]. A key feature of this solution was the ability of the agents to negotiate the allocation of the tasks to be performed during the mission between each other based on explicit goals for cooperation and rules of coordination and communication.

In the second phase of our research, we focused on the realization of a mixed manned-unmanned fighter-UCAV team. Therefore, we replaced the on-board agent of one UCAV by a human pilot, who used a simple desktop human-machine interface to control the aircraft and to communicate with the other agent-guided UCAVs [4]. In human in-the-loop experiments, we could show the applicability of

human-agent cooperation to joint fighter UCAV operations. For a more realistic evaluation of the human-machine cooperation, in a next step the agents of the desktop simulation were modified and integrated in a generic single-seat fighter cockpit simulator [5–7]. For this setup and the following human-in-the-loop experiments, a high level of automation was chosen for the UCAV agents. This team-based cooperative guidance only required an absolute minimum of interventions by the pilot. The cooperative allocation of tasks between the UCAVs (i.e., suppression of enemy air defense, target reconnaissance, target designation, battle damage assessment, and fighter escort) was performed by the agents fully automated, based upon negotiation mechanisms. This high level of automation produced the need for an associative assistant system, which consisted of a Team Coordination Module (TCM) [5], and a Self-Explanation Capability Module (SECM) [7]. The assistant system informed the operator about the behavior of the unmanned team members, and thereby, ensured a required minimum of Situation Awareness (SA) and trust in the automation. The TCM operated in an associative assistance mode, which provided the pilot with spatial and temporal coordination information of the UCAVs, and in a few use cases also in an alerting assistance mode that directed the pilot's attention in the case of coordination conflicts. Additionally, the agents routinely explained the UCAVs' behavior to the pilot, and provided further information upon request. To evaluate the joint fighter-UCAV system, an experimental study with German Air Force pilots was conducted. These experiments included air-to-ground missions, where the manned fighter was supported by three UCAVs. The main tasks of these missions were the reconnaissance, the designation and the engagement of a high priority target, the suppression of ground-based enemy air defense, and the protection of the manned fighter in general. To accomplish these tasks, the UCAVs and the manned aircraft had different capabilities and payloads such as High-Speed Anti-Radiation Missiles (HARM) for Suppression of Enemy Air Defense (SEAD), Laser-Guided Bombs (LGB) for the target destruction, cameras for reconnaissance, and laser designators. The manned fighter aircraft could only perform the high-level mission goal, the engagement of high-priority targets, after a visual verification of reconnaissance pictures, provided by one of the UCAVs. All other mission-relevant tasks were autonomously anticipated by the UCAVs' agents according to their capabilities and the dynamic environment. Throughout the mission, the UCAV agents proactively pursued the overall mission goal. The experiments showed, that the cognitive agent-controlled UCAVs could adapt to the changing environment and to unforeseen situations (e.g., pop-up threats) effectively. The assistant system, consisting of the TCM and SECM, was considered as helpful and could slightly increase the SA. However, a significant increase of trust in the unmanned members could not be shown [7]. The study also indicated that the constant high level of automation temporarily led to mental under-load of the pilots and was lacking in adaptability to balance the operator's activity and work demands over the course of the mission. The experimental subjects further expressed the desire to be able to assign specific tasks to the UCAVs during mission execution, especially in less demanding situations.

Generally speaking, the experiments approved that high-level automation in the form of cooperating cognitive agents is realizable and suitable for multi-UCAV guidance as well as for joint fighter-UCAV operations in demanding future air warfare scenarios. However, the study also showed the need to balance the operator's workload in higher as well as in less demanding situations.

3 Key Aspects of Automation in Human-Machine Systems

Due to severe human-induced system errors in highly-automated civil aviation aircraft, researchers invested considerable effort to identify negative effects correlated to automation in complex human-machine systems. In this chapter, we want to summarize important aspects that may also negatively affect joint fighter-UCAV operations. Afterwards we derive some basic guidelines for the design of our system.

The first issue, called human-out-of-the-loop performance, is attributed to several factors. These factors include vigilance decrements, complacency, over-reliance, under-reliance, loss of SA, automation surprise, inappropriate feedback, and skill degradation, which are closely coupled to each other. Inappropriate automation in human-machine systems may initiate vigilance decrements or reduced operator alertness and as a consequence undetected system failures, especially for systems where the human has a supervising role. Out-of-the-loop issues due to operator vigilance are in many cases closely linked to complacency and over-reliance on automation (compare [8, p. 438, 9, p. 544]). According to [9, p. 544] detection and SA problems, as well as skill degradation can be negative consequences of automation-induced complacency. In [10, p. 192, 11] the authors also identify over-reliance as a negative effect of automation. At the same time under-reliance on automation as a result of system unreliability (or automation surprises) may also be a shortfall of automation [9, p. 543, 12]. Another important source of errors in human-machine systems is the loss of SA. According to [13] SA is mainly affected by complacency and vigilance, active-passive role switching, and system feedback. Out-of-the-loop issues related to automation surprise base on automation complexity, where the human operator may not be able to understand the system behavior [9, p. 542, 14]. Another automation-induced out-of-the-loop-performance issue is skill degradation. In [13] the authors link skill degradation to changed operator vigilance and complacency. In [8, p. 438, 10, p. 195, 11, 15] the authors also name automation-induced skill degradation as a general drawback of automation.

A second aspect, which is associated to automation problems in human-machine systems is unbalanced mental workload. According to [8, p. 438] automation may increase as well as decrease mental workload, which can cause mental over-load as well as mental under-load. In over-load situations the human operator is not able to cope with the situation, under-load situations can be linked to the named out-of-the-loop issues. Another problem with regard to workload is "clumsy

automation” [10, p. 193], which affects the operator’s mental workload counter-productively. In low-workload situations the workload is reduced, in high-workload situations the operator’s workload may even be increased.

Based on these theoretical underpinnings, we identified approaches such as *human-centered automation* (compare [10]), *adaptable and adaptive automation* (e.g., [16]), and *cognitive and cooperative automation* (compare [17]) as key design elements for our concept.

4 The Concept of Self-scaling Human-Agent Cooperation

To achieve the aforementioned capabilities, we suggest a human automation integration concept that features adaptable and adaptive human-agent cooperation. The concept bases on dual-mode cognitive automation [17], which has already been applied successfully in ground-based UAV guidance [18] and multi-UAV guidance from aboard a helicopter cockpit (compare [19, 20]).

4.1 Framework for Automation Design in Complex Human-Machine Systems

According to [17, 21] the starting point of sophisticated automation in human-machine systems is the definition of a work process (WProc), the associated work objective (WObj), the associated work process output (WPOut), and the affected work object (WO). Furthermore, other work processes, the environment (Env), supplies (Sup), and information (Inf) are considered. After defining the WProc, the physical system running the WProc is considered. This system is called a work system (WSys). Within a WSys, in principle, two roles are distinguished—the *Worker* and the *Tools* role [21] (see Fig. 1).

- Worker: The Worker knows, understands, and pursues the WObj by own initiative. By definition a WSys cannot exist without a human Worker.
- Tools: The Tools receive tasks from the Worker and only perform them when told to do so. Hence, the Worker has a hierarchical relationship to the Tools.

Fig. 1 Work system with worker and tools roles according to [21]

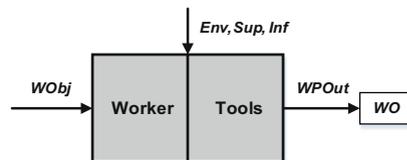
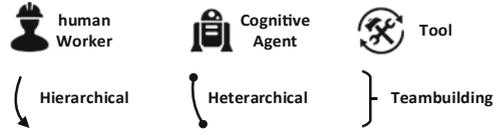


Fig. 2 Symbolic building blocks following [21]



In order to describe highly automated work systems involving human-cognitive agent teaming the authors in [21] suggest a symbolic language that allows describing a small number of building blocks. These are the human Worker, the Tool, and the Cognitive Agent blocks. Cognitive Agents are entities providing higher cognitive capabilities and can exist in the role of a Worker or a Tool. In addition, to describe the relation between these entities, hierarchical relationship, heterarchical relationship, and teambuilding symbols are used (see Fig. 2).

One approach to facilitate automation in complex work systems is dual-mode cognitive automation as suggested by Onken and Schulte [17]. This concept is supposed to overcome the drawbacks of conventional automation by incorporating Cognitive Agents into the WSys. The second key aspect of this approach concerns the design of human-agent relationships. Dual-mode cognitive automation applies two modes of relationship in the form of (1) a hierarchical and (2) a heterarchical link between the human Worker and the Cognitive Agents.

4.2 Work System Design for Joint Fighter-UCAV Operations

To integrate self-scaling human-agent cooperation in the fighter-UCAV system we define the WProc “Perform Fighter-UCAV Mission”, the WO “Target”, and the highly abstract WObj “Perform Fighter-UCAV Mission” as presented in Fig. 3. This WProc is initiated by the WProc “Command & Control Operation”, which represents the WProc of a superior command and control authority (e.g., Air Operation Center), as well as the variables Env (e.g., weather conditions in target

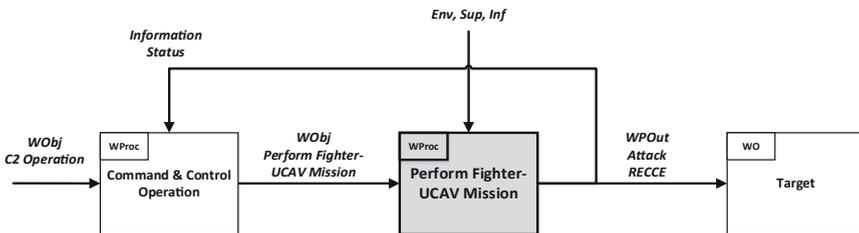


Fig. 3 WProc “Perform Fighter-UCAV Mission” with WO “Target”, WPOut “Attack/RECCE”, and WObj “Perform Fighter-UCAV Mission”

area), Sup (e.g., fuel), and Inf (e.g., status of reconnaissance of target area). By introducing the human Worker (fighter pilot), the Tools (manned aircraft and UCAVs), and Cognitive Agents, the WProc is afterwards materialized as a work system.

As a conceptual basis of our automation integration concept, we choose the aforementioned dual-mode cognitive automation. For the heterarchical relationship we incorporate a Cognitive Agent in the form of an *assistant system*, which cooperates with the human Worker. Furthermore, we incorporate Cognitive Agents controlling the unmanned aircraft. The human Worker and the assistant system have a hierarchical relationship towards these agents, although these UCAV agents are supposed to hold a Worker role. The hierarchical relationship is realized in the form of three *delegation modes*, featuring *team-based*, *intent-based*, and *task-based* guidance of the UCAVs. These modes are used to scale the hierarchical relations between the pilot and the assistant system agent towards the UCAVs as well as the cooperation among the UCAV agents.

1. Team-based guidance (Fig. 4a): This delegation level comprises multiple UCAVs pursuing a highly abstract mission goal as a team, e.g., a coordinated target attack. Therefore, each agent needs to have the capabilities to anticipate tasks in consultation with other team members. Furthermore, the team members

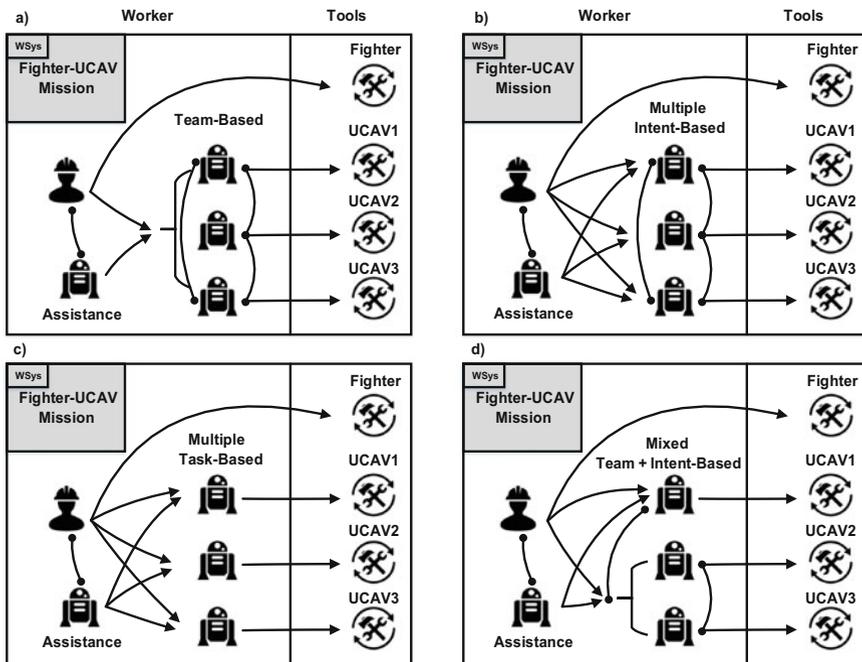


Fig. 4 Work system configurations. a Team-based guidance, b Multiple intent-based guidance, c Multiple task-based guidance, and d Mixed team + intent-based guidance

- need to cooperate effectively within the team to keep the team plan and schedule, but also to cooperate with other UCAVs/UCAV teams. This delegation mode may be applied in high workload situations, where the pilot is not able to guide and manage individual UCAVs on an intent- or task-based level.
2. (Multiple) intent-based guidance (Fig. 4b): Here a single UCAV is supposed to pursue a mid-level goal, e.g., the observation of a target area. Therefore, the associated agent needs to be able to autonomously plan and schedule towards the mid-level goal. Although the UCAV pursues an individual task, the associated agent cooperates with other UCAVs/UCAV teams to avoid conflicts with regard to the high-level mission goal. This delegation mode is supposed to be appropriate for balanced workload.
 3. (Multiple) task-based guidance (Fig. 4c): The unmanned team member receives a low-level task, e.g., taking a recce picture. For this delegation mode the associated UCAV agent must provide low-level task assignment and execution. This mode is supposed to balance low workload phases by increasing the operator's delegation tasks. It might as well be used for tasks, where immediate action or ethically responsible decision (e.g., weapon deployment) implementation is required that shall not be left to the automation.

Besides these levels, the delegation of the UCAVs could also be applied in a mixed manner (Fig. 4d), where e.g., a sub-team of UCAVs shall suppress the enemy air defense in the target area, while one UCAV shall reconnoiter the egress route.

The delegation modes correlate to a two-dimensional definition of automation. As suggested in [22], we want to take into account the ability of an entity to take care of itself (*self-sufficiency*) and the freedom from outside control (*self-directedness*). In our application the Cognitive Agents operating the UCAVs are supposed to have a high self-sufficiency (high-level skills), but are restricted in their self-directedness according to the delegation modes team-based, intent-based, and task-based. In contrast, the assistant system (heterarchical component) has a high self-sufficiency and a high self-directedness. For the definition of the operator-assistant system relationship, we introduce a third dimension of automation, which we call *assistance*. This dimension provides alarming, supporting, and cooperating functionalities, which enable the assistant system to vary between attention guidance, proposal of task modifications, and task adaptations (due to ethical issues, task adaptations may require human approval).

4.3 Self-scaling Capabilities

To enable our system to work effectively in joint fighter-UCAV operations, our automation integration concept is supposed to actively balance the operator's mental state, and to address automation-induced negative effects. Therefore, the heterarchical and the hierarchical relationship between the human Worker and the

Cognitive Agents feature scalability. The hierarchical relationship (the guidance of the UCAVs) is scaled according to the proposed delegation modes, whereas the heterarchical relationship (operator-assistant system cooperation) is scaled by adapting the mode of assistance and the Human-Machine Interface (HMI).

As the assistant system is able to initiate system adaptations autonomously, it represents the key component of our automation concept. To trigger system adaptations, we intend to pursue an operationalization of the pilot's mental workload similar to the approach in [23]. The operationalization of mental workload considers the operator's tasks, activities, and related mental resource demands, as well as observable variations in behavior patterns.

Although the assistant system primarily initiates system adaptations (*system-initiated adaption mode*), the pilot is able to intervene and to directly assign tasks to the UCAVs in case of faulty system behavior or when such interaction is desired (*operator-initiated adaption mode*). That implies that our assistant system allows *two modes of cooperation* with the human Worker. In case of direct operator task assignment not only the delegation of the UCAVs is adapted, but also the HMI and the mode of assistance. Nevertheless, the assistant system is still able to propose system adaptations when working in the operator-initiated adaption mode. If the human operator accepts such a proposal, the assistance, the HMI, and the delegation of UCAVs is adapted according to the pilot's mental state and the environmental situation (as in the system-initiated adaption mode). By monitoring operator inputs and mixed-initiative capabilities we want to enable the assistant system to solve evolving conflicts in the operator-initiated adaption mode interactively with the human operator.

4.4 Relations Between Human Worker and Cognitive Agents

In this section we want to particularize the relations between the Cognitive Agents and the human Worker. We intend to design effective human-agent cooperation by taking account of the four requirements basic compact (to work together), maintenance of a common ground, directability, and predictability as suggested in the joint-activity concept for effective team work in [24].

To achieve a basic compact for common-grounding activities, all team members need to understand and accept their role in the work system. In the context of our military application the Cognitive Agents shall not be able to refuse operator commands, unless the situation does not allow the execution of the mission or any subtasks due to ethical reasons (e.g., the bombing of a target or an air defense position, where civilians may be affected).

To maintain a common ground, we intend to use shared knowledge in the form of a blackboard, and Cognitive Agents, which are enabled to model the human Worker and other Cognitive Agents. The use of a blackboard is supposed to

permanently provide appropriate information on the status, actions, and intentions of all team members. By linking all agents to the blackboard they are able to observe the other team members. Understanding of this information shall be enabled by incorporating appropriate agent knowledge. The assistant system agent is supposed to pertinently interpret information on the blackboard and thereby shall be enabled to manage attention, e.g., by informing the human Worker about conflicts and problems which may negatively influence the mission. In addition, we want to incorporate reasoning capabilities in our agents, which take individual goals and mutual impact on actions into account. Our assistant system is supposed to act de-conflicting and enable goal negotiation by mixed-initiative strategies. To reduce coordination costs we want to build up our system on a Multi-Agent System (MAS) which offers sufficient coordination.

Mutual directability shall be assured by applying the aforementioned hierarchical delegation modes, mutual predictability shall be generated with the blackboard, which is accessible and understandable by each team member.

5 Related Work

In the following we present selected works which are closely linked to our concept. In the context of multiple UAV guidance we identified resemblances with the works of [25, 26]. In [25] an assistant system with mixed-initiative planning and execution capabilities was incorporated in a future mountain search and rescue system. Their approach made use of hierarchical delegation modes to enable a co-located operator to guide and manage multiple UAVs on adjustable levels of automation. In [26] a manned fighter aircraft was accompanied by three simulated and one real UCAV in mixed-reality flight tests. In their scenarios the pilot largely took on a supervisory control role after defining mission goals taken from a “pool” of goals. In their previous work the authors had recognized the need for team-based task allocation to UCAVs due to operator over-load issues. The self-organized task allocation among participating UCAVs was realized with a MAS, in which four different types of agents cooperated (user, group, specialist planning, and UAV agent). Although, the named concepts both bear a resemblance with our approach, they lack of self-scaling capabilities to autonomously adapt to the operator’s mental state and the environment. Related works with respect to adaptability and/or adjustability are numerous (e.g. [16, 27]). However, such works often apply system- and user-initiated dynamic function allocation whereas the scaling in our approach aims at adapting human-agent and agent-agent cooperation. Besides the abovementioned workload operationalization method [23], there exist further approaches, which are used to trigger adaptive automation. The authors of [28] correlated the mental workload to eye fixation time in combat management scenarios with different cognitive task load. Further works base on EEG (electroencephalography) measurements. In [29] a subject-specific discrimination of two workload levels was

achieved with artificial neural networks. The cross-subject training and testing of a hierarchical Bayes model in [30] enabled the separation of three workload levels.

6 Conclusion

In this article, we propose a human-automation integration concept, which is supposed to enable a single-seat fighter pilot to guide and manage multiple UCAVs from aboard the cockpit in joint fighter-UCAV operations. In a first step, we define the associated work process, in which we incorporate Cognitive Agents, Tools, and the pilot as a human Worker to instantiate a work system. The allocation of Cognitive Agents in the work system bases on the concept of dual-mode cognitive automation. We make use of an assistant system and multiple UCAV guidance on the basis of three different delegation modes (team-based, intent-based, and task-based). The conceptual core of our approach is the self-scaling human-agent cooperation, which features human-centricity, adaptivity, and cognitive and cooperative automation. Thereby, we intend to balance the operator's mental state, and to address automation-induced negative effects. In general, system adaptations are initiated by the assistant system on the basis of the operator's mental state and the dynamic battlefield environment. To intervene in faulty system behavior or when desired, our approach also allows operator-initiated system adaptations.

Although the named capabilities are supposed to allow effective and efficient joint fighter-UCAV operations, we identified some open issues which show the need for further investigations. The first essential point we identified is the choice of an appropriate cognitive (multi-)agent software. Another challenge will be the design of the Human-Machine Interface in the cockpit. It must enable effective interaction between the human operator and the automation for changing system configurations.

To investigate the proposed concept, we intend to implement a software prototype in our single-seat fighter simulator for experimental studies in the mid-term future.

References

1. Chappelle, W.L., McDonald, K., McMillan, K.: Important and Critical Psychological Attributes of USAF MQ-1 Predator and MQ-9 Reaper Pilots According to Subject Matter Experts, pp. 1–35 (2011)
2. de Freitas, M.M., Cunha, F.S., Ribeiro, A.M.R., Azinheira, J.R., Carvalho, R.J.S., Cabrita Freitas, J., Avalle, M.: UCAV : A Technology Assessment Project as a Complex Problem Case Study (2009)
3. Meitinger, C., Schulte, A.: Cognitive machine co-operation as basis for guidance of multiple UAVs. In: NATO RTO HFM Symposium on Human Factors of Uninhabited Military Vehicles as Force Multipliers. Biarritz, France (2006)

4. Meitinger, C., Schulte, A.: Human-UAV co-operation based on artificial cognition. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 91–100 (2009)
5. Gangl, S., Lettl, B., Schulte, A.: Management of multiple unmanned combat aerial vehicles from a single-seat fighter cockpit in manned-unmanned fighter missions. In: *AIAA Infotech@Aerospace (I@A) Conference*, pp. 1–18. American Institute of Aeronautics and Astronautics, Reston, Virginia (2013)
6. Gangl, S.: *Kooperative Führung mehrerer unbemannter Luftfahrzeuge aus einem einsitzigen Kampfflugzeug* (2015)
7. Lettl, B., Schulte, A.: Self-explanation capability for cognitive agents on-board of UCAVs to improve cooperation in a manned-unmanned fighter team. In: *AIAA Infotech@Aerospace (I@A) Conference*, pp. 1–11. American Institute of Aeronautics and Astronautics, Reston, Virginia (2013)
8. Wiener, E.L.: Cockpit automation. In: Wiener, E.L., Nagel, D.C. (eds.) *Human Factors in Aviation*, pp. 433–461. Academic Press, London (1988)
9. Wickens, C.D., Hollands, J.G.: *Engineering Psychology and Human Performance*. Prentice Hall, Upper Saddle River (2000)
10. Billings, C.E.: *Aviation Automation: The Search for a Human-Centered Approach*. Lawrence Erlbaum Associates, Mahwah (1997)
11. Kaber, D.B., Endsley, M.R.: Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Saf. Prog.* **16**, 126–131 (1997)
12. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. *Hum. Factors J. Hum. Factors Ergon. Soc.* **39**, 230–253 (1997)
13. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. *Hum. Factors J. Hum. Factors Ergon. Soc.* **37**, 381–394 (1995)
14. Sarter, N.B., Woods, D.D.: Team play with a powerful and independent agent: operational experiences and automation surprises on the Airbus A-320. *Hum. Factors J. Hum. Factors Ergon. Soc.* **39**, 553–569 (1997)
15. Wiener, E.L., Curry, R.E.: Flight-deck automation: promises and problems. *Ergonomics* **23**, 995–1011 (1980)
16. Kidwell, B., Calhoun, G.L., Ruff, H.A., Parasuraman, R.: Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **56**, 428–432 (2012)
17. Onken, R., Schulte, A.: *System-Ergonomic Design of Cognitive Automation: Dual-Mode Cognitive Design of Vehicle Guidance and Control Work Systems*. Springer, Berlin (2010)
18. Theißing, N.: Designing a support system to mitigate pilot error while minimizing out-of-the-loop-effects. In: *13th Conference on Engineering Psychology and Cognitive Ergonomics. Lecture Notes Computer Science* (2016)
19. Uhrmann, J., Schulte, A.: Task-based guidance of multiple uav using cognitive automation. In: *COGNITIVE 2011, The Third International Conference on Advanced Cognitive Technologies and Applications*, pp. 47–52 (2011)
20. Rauschert, A., Schulte, A.: Cognitive and cooperative assistant system for aerial manned-unmanned teaming missions. *NATO Res. Technol. Agency, Hum. Factors Med. Panel, Task Gr. HFM-170 Superv. Control Mult. Uninhabited Syst. Methodol. Enabling Oper. Interface Technol. RTO-TR-HFM* **170**, 1–16 (2012)
21. Schulte, A., Donath, D., Lange, D.S.: Design patterns for human-cognitive agent teaming. In: *13th Conference on Engineering Psychology and Cognitive Ergonomics. Lecture Notes Computer Science* (2016)
22. Bradshaw, J.M., Feltovich, P.J., Jung, H., Kulkarni, S., Taysom, W., Uszok, A.: Dimensions of adjustable autonomy and mixed-initiative interaction. In: *Agents and Computational Autonomy*, pp. 17–39. Springer, Berlin (2003)

23. Schulte, A., Donath, D., Honecker, F.: Human-system interaction analysis for military pilot activity and mental workload determination. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1375–1380. IEEE (2015)
24. Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J.: Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intell. Syst.* **19**, 91–95 (2004)
25. Bevacqua, G., Cacace, J., Finzi, A., Lippiello, V.: Mixed-initiative planning and execution for multiple drones in search and rescue missions. In: ICAPS, pp. 315–323 (2015)
26. Baxter, J.W., Horn, G.S., Leivers, D.P.: Fly-by-agent: controlling a pool of UAVs via a multi-agent system. *Knowledge-Based Syst.* **21**, 232–237 (2008)
27. Parasuraman, R., Barnes, M., Cosenzo, K., Mulgund, S.: Adaptive Automation for Human-Robot Teaming in Future Command and Control Systems. DTIC Document (2007)
28. Greef, T., Lafeber, H., Oostendorp, H., Lindenberg, J.: Eye movement as indicators of mental workload to trigger adaptive automation. In: Schmorow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience: 5th International Conference*, pp. 219–228. Springer, Berlin (2009)
29. Wilson, G.F., Russell, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiologicaly determined adaptive aiding. *Hum. Factors J. Hum. Factors Ergon. Soc.* **49**, 1005–1018 (2007)
30. Wang, Z., Hope, R.M., Wang, Z., Ji, Q., Gray, W.D.: Cross-subject workload classification with a hierarchical Bayes model. *Neuroimage* **59**, 64–69 (2012)

Experimental Analysis of Behavioral Workload Indicators to Facilitate Adaptive Automation for Fighter-UCAV Interoperability

Dennis Mund, Felix Heilemann, Florian Reich, Elisabeth Denk, Diana Donath and Axel Schulte

Abstract In this article, we present an experimental study investigating the operationalization of behavioral indicators of pilots' mental workload in a military manned-unmanned teaming scenario. For the identification of such behavioral workload indicators, we conducted an explorative experimental campaign. We chose an air-to-ground low-level flight mission with multiple target engagements. To further increase the task load of the pilots, we introduced an embedded secondary task, i.e. the classification of target pictures delivered by remote UCAVs. This is a typical task, which we expect in future manned-unmanned teaming setups. The examination of the subjective ratings shows that high individual workload states were achieved. In these high workload situations, the subjects used various behavioral adaptations to keep a high performance level while regulating their subjective workload. As these behavioral adaptations occur prior to grave performance decrements, we consider to use behavioral changes as indicator for high workload and as trigger for adaptive support.

Keywords Behavioral workload · Ucavs · Adaptive automation · Manned-unmanned teaming

D. Mund (✉) · F. Heilemann · F. Reich · E. Denk · D. Donath · A. Schulte
Institute of Flight Systems, University Bundeswehr Munich,
Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
e-mail: dennis.mund@unibw.de

F. Heilemann
e-mail: felix.heilemann@unibw.de

F. Reich
e-mail: florian.reich@unibw.de

E. Denk
e-mail: elisabeth.denk@unibw.de

D. Donath
e-mail: diana.donath@unibw.de

A. Schulte
e-mail: axel.schulte@unibw.de

1 Introduction

With manned-unmanned teaming being realized in airborne helicopter cockpits like the Apache E, in which unmanned vehicles are controlled from aboard the helicopter [1], we anticipate similar applications in the fighter jet domain in the near future, as for example [2]. This will be especially challenging for single-seated fighter aircraft. Although it is expected that these aircraft will be equipped with high-level onboard automation, their pilots will be additionally burdened with certain tasks concerning the supervision and handling of unmanned assets, which may lead to unbalanced workload conditions.

Depending on the given situation, this supervisory control task results in high variations of task-load for the human pilot. In combination with individual parameters, such as experience and training, these variations in task-load will entail a massive fluctuation of the mental workload of the pilot. In order to support pilots adequately, we suggest using adaptive automation techniques as part of an assistant system to alleviate peaks of workload (as in e.g. [3]). Therefore, a continuous determination of the mental workload of the pilot is required. Considering the available workload measurement methods, i.e. subjective ratings, psychophysiological-, performance-, situation based-, and behavior measurements, only a subset of these methods may be suited for continuous individual workload state gathering [4]. Subjective ratings rely on questionnaires and thus the workload is not accessible continuously. In contrast, situation based measurements can be accessed, but use generalized models, which makes them insensitive to individual factors of the given subject. The remaining methods are appropriate, but each of them has its benefits and drawbacks. While physiological measurements allow conclusions about the cognitive workload by using parameters like heart rate or the electrical resistance of the skin, these methods do not consider information about the task responsible for the high workload. Performance measures indicate high workload situations and the associated task context, but this information is only available after performance already declined. Similar to human operator performance, human behavior is also task specific and indicates high workload situations before performance decrements occur. The reason for this is that humans adapt their behavior according to their perceived workload situation to cope with high workload situations and to maintain their performance on a high level as long as possible (see [5]). By taking advantage of these objectively observable behavioral adaptations as indicator for high workload situations within future adaptive assistant systems, these systems might be able to offer task specific support prior to grave performance decrements.

In order to identify such regulating behaviors we conducted experiments in our fighter aircraft research simulator to identify behavioral adaptations. We use fighter-UCAV manned-unmanned teaming missions as application context. The remainder of this section describes the behavioral adaptations in more detail, followed by a description of our experimental setup in section two. The results of our

experimental campaign are presented in section three. The final section summarizes our findings and presents open issues which need for further investigation.

1.1 Behavior in High Workload Situations

Many publications indicate that human performance depends on the subjective mental workload a person is exposed to (e.g. [6]). In low workload situations, the performance may be also low, due to boredom and complacency effects. In contrast high workload situations require additional effort of the human to retain a certain level of performance compared to a normal workload situation, in which the human reaches a high performance without difficulties. Figure 1 illustrates this relationship between performance and workload. The limits between low, normal, and high workload depend on individual factors that include, but are not limited to training and experience, and also are influenced by behavioral precursors like health state or fatigue (see [6]). One expression of a person spending effort is the self-adaptation of behavioral patterns in high workload situations, where the human deviates from regular task execution, in order to execute tasks more efficiently at the expense of performance (see [7]). Suchlike adaptations in behavioral patterns are, in principle, accessible to technical measurement, and can be exploited as indicator for high workload situations [8]. According to [9] they are called self-adaptive strategies and can be separated into two classes:

- *Load-sharing*: A task or subtask is assigned to another person/crew-member or to automation in order to distribute the task load over multiple operators.
- *Load-shedding*: A task is executed more efficiently at the expense of the performance. By load-shedding the performance is slightly decreased, but the task can be performed more efficiently.

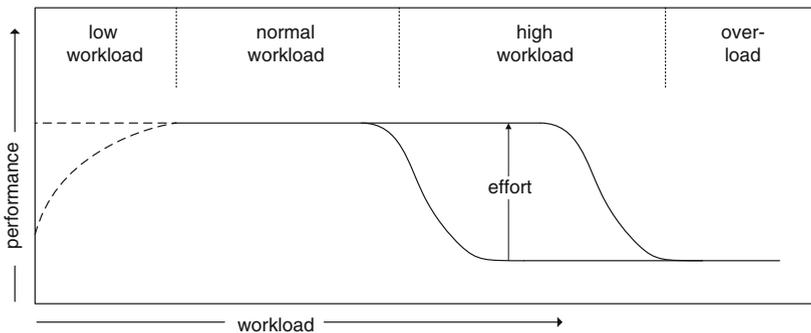


Fig. 1 Schematic relation between workload and performance

2 Experimental Setup

To identify such self-adaptive strategies we conducted experiments with German Air Force pilots in our generic jet fighter research simulator cockpit. We expected behavioral changes between situations of low-subjective workload and those of high workload. The experiments contained four military low-level flight missions for behavioral data acquisition. Each of these missions consisted of five segments with different levels of complexity, mainly determined by the frequency of the embedded secondary task. In order to distribute known and unknown individual disturbances uniformly over the group of subjects, we chose a duplex randomized, partial, within subject experimental design. The randomization varied the sequence of missions to be accomplished by the subjects and within each mission the sequence of levels of difficulty.

The primary task of each mission was the engagement of four targets along a predefined route in an air-to-ground combat scenario. In order to increase the workload of the subjects they had to meet given mission-specific constraints with respect to the primary task. Therefore, we used low-flight level constraints defining the minimum and maximum flight altitude (minimum flight altitude: 100 ft AGL, maximum flight altitude: 300 ft AGL). This task was impeded by a high threat level evoked by multiple Surface to Air Missile (SAM)-Sites in the mission area. Furthermore the subjects had to meet Time-over-Target (TOT) intervals for each target engagement ($TOT \pm 5$ s).

We additionally charged the subjects with an embedded secondary task, the classification of objects regarding type (i.e. vehicle, person, building, SAM-site) and hostility (i.e. friend/foe) using sensor images provided to the test subjects during their primary task execution (see Fig. 2 middle and right). This secondary task is expected to occur in future manned-unmanned teaming setups, where the unmanned assets are operated on Level-of-Interoperability (LoI) II/III, which includes amongst others the receipt, monitoring and evaluation of UAV (sensor) payload data. In order to increase the workload, the density of the object classification tasks were varied between each mission segment and hence over the course of mission.



Fig. 2 *Left:* Setup of our simulation environment; *Middle:* Exemplary image that had to be classified as a friendly Building; *Right:* Exemplary image that had to be classified as an enemy vehicle

To capture human behavior during mission accomplishment, the simulator (see Fig. 2 left) was equipped with Smart Eye Pro[®], a contact free, video-based eye movement measurement system. Additionally, the manual interactions of the pilot (throttle-, stick-, and display/button interactions) were recorded. In order to relate the operators' behavioral adaptations to his subjective workload, we additionally gathered subjective workload ratings and performance measurements. To assess the subjective workload ratings we used the NASA-Task Load Index [10], because it is common in the flight domain. The questionnaires were presented directly after each target attack. The measurement of the operator performance referred to the primary and the secondary task comprising the following performance and behavior parameters:

- Primary task related parameters: number of successful target attacks, violation of the given low-level flight constraints, violation of the specified TOT tolerances.
- Secondary task related parameters: number of accomplished object classification tasks, number of errors in the object classification task, required time for the task accomplishment (i.e., interaction time), and the time between the appearing of the task and the beginning of the execution of the task (i.e., wait time).

The experiments were distributed over a span of two days per subject. The first day began with an introduction to the simulation environment, the methods for subjective and behavioral data acquisition (i.e., subjective ratings and eye-tracking measurement). Afterwards, the primary and secondary tasks were introduced in four training missions, and the day ended with the first mission for data acquisition. On the second day, a refresher training was offered to the subjects before conducting three further missions for behavioral data acquisition. The experiments were run with six German Air Force fighter pilots with a minimum age of 29 and a maximum age of 56 years. This results in an average age of 42.3 years with a standard deviation of 23.86 years.

3 Experimental Findings

This section describes the findings from our experiments with German Air Force pilots. First the results of the subjective workload assessment are presented. Afterwards the performance and the observed behavioral adaptations (self-adaptive strategies) are investigated.

The performance measures for the primary as well as the secondary task do not show any clear correlation with the density of images that had to be classified by the pilot. This indicates that the image density is insufficient as a sole driver for the difficulty of a mission segment. We identified the structure of the encountered terrain, approximated by the standard deviation of the terrain altitude, as a second important driver for the difficulty of a mission segment. Based on these findings, the mission segments were grouped into three clusters of similar terrain difficulty and

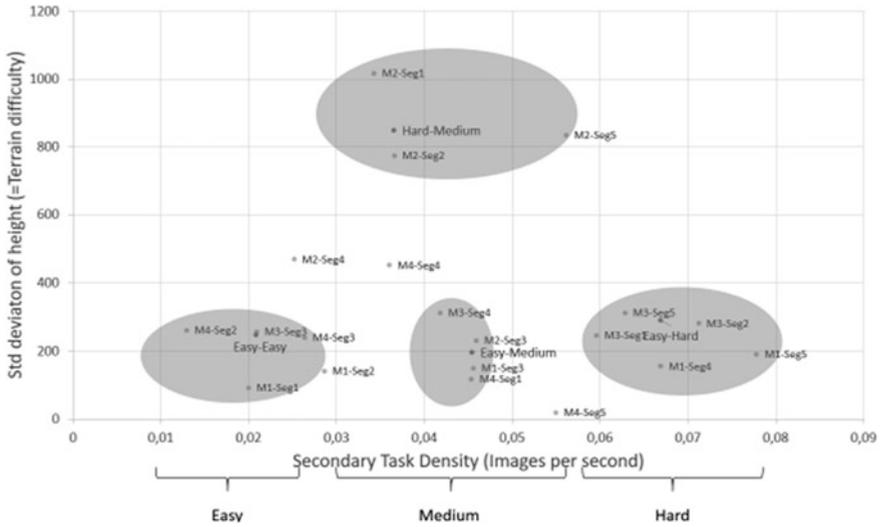


Fig. 3 Clusters of varying difficulties. MX-SegY is a short hand notation for mission X segment Y

increasing image density (low, medium, high) and one cluster with medium image density and high terrain difficulty (see Fig. 3). Using these clusters, it is possible to evaluate the effects of increased image density, as well as increased terrain difficulty on gathered workload and performance independently of each other.

3.1 Subjective Workload Ratings

The combined workload ratings from the NASA TLX questionnaires in Fig. 4 show medium to high workload demands across all experiments. This is a good indication that the experimental setup stimulated the workload range as intended. Additionally, we observed an insignificant increase in workload ratings with increased image density (Easy-Easy \geq Easy-Medium \geq Easy-Hard). On the other hand, there is no change in the subjective workload ratings between the clusters with different terrain difficulty (Easy-Medium vs. Hard-Medium). The mean, as well as the median, of these ratings are similar for both clusters. Figure 4 shows boxplots of the workload ratings attributed to the four different clusters. The whiskers depict the extreme values (Minimum, Maximum) of the given ratings.

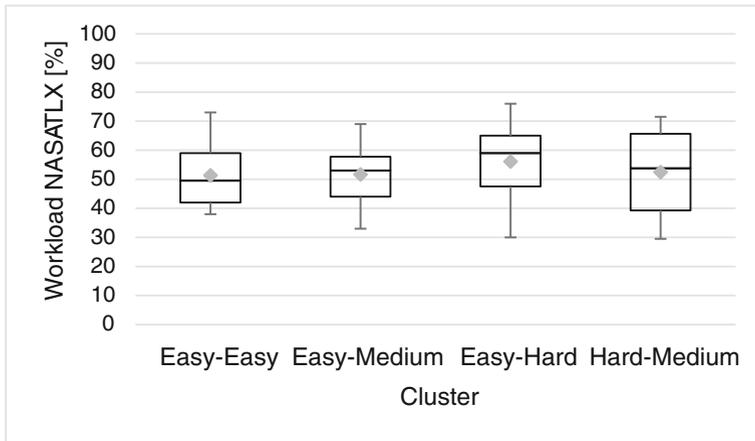


Fig. 4 NASA TLX subjective workload evaluation for the four clusters (*whiskers*: extreme values, *diamonds*: mean values)

3.2 Performance Evaluation

All subjects succeeded in engaging the mission targets, with the exception of one subject missing its first target in its first mission. We assume the one failed target attack to be connected to a lack of training. Concerning the given time constraints, in general, there were only very few deviations from the demanded TOT intervals. An exception was mission two segment three, in which all but one subject missed the timing interval. The third performance criterion for the primary task, i.e. the adherence of the altitude range, shows a correlation with the density of the secondary task. A higher density of this classification task results in an increased

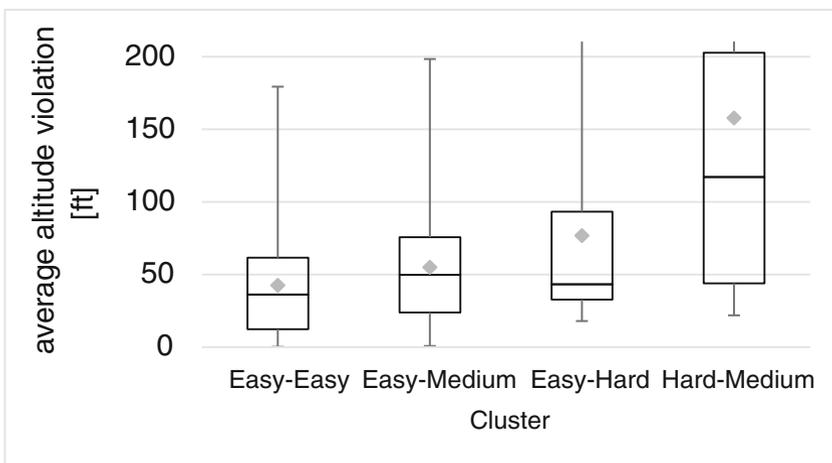


Fig. 5 Average altitude violation for the four identified clusters

Table 1 Mean and standard deviation of the average secondary task execution per cluster

(Sec)	Easy-easy	Easy-medium	Easy-hard	Hard-medium
Mean	4.53	4.23	3.91	4.19
Std. deviation	2.05	1.34	0.94	1.07

violation of the altitude range, as shown in the mean values for the clusters Easy-Easy, Easy-Medium, and Easy-Hard in Fig. 5. Similarly, the comparison between the Easy-Medium and the Hard-Medium cluster shows a strong negative influence of the terrain difficulty on the adherence of the altitude range.

Regarding the secondary task, all subjects were able to process the 56 images per mission that were provided by the UAVs. The average processing time for one identification task was affected neither by the terrain difficulty nor by the image density as shown in Table 1. In contrast, the time between the appearance of a classification task and the start of its processing increased significantly with increasing task frequency from a mean of 7.5 s (sample size = 180; std. deviation = 6.3 s) in the Easy-Easy to 9.41 s (sample size = 450; std. deviation = 8.4 s) in the Easy-Hard cluster. As observed for the processing time, this performance measure is not affected by the terrain difficulty.

3.3 Behavioral Adaptations

In a second step, we performed a more in-depth analysis of the experimental data in order to detect behavioral changes induced by subjective workload effects. In the following, we will describe the identified self-adaptations.

Deliberate Relaxation of Flight Constraints. The first strategy refers to the relaxation of the given flight altitude constraints (lower flight limit: 100 ft AGL, upper flight limit: 300 ft AGL). We observed that some of the subjects relaxed this constraint right from the beginning in order to prevent critical workload conditions by freeing mental resources, otherwise required for the low-level flight within the limits. Other subjects tried to stay within the given flight constraints. The intentional use of this strategy was confirmed by the subjects in a debriefing at the end of the experiments.

Delay in Object Classification Task. In this strategy, subjects delayed singular image processing tasks (object classification) in favor of other tasks. The use of this strategy was frequently observed, especially when the subject encountered situations of increased difficulty in the primary task, e.g. during navigational turns at route waypoints, or during attack runs. Figure 6 shows the execution of the object classification task during a navigational turn (grey area) compared to regular task execution. Each task performance is depicted by using the three task-relevant times: the time when the task appears first (circle), the time when the pilot starts working on the task (diamond), and the time of task completion (square). The time of the

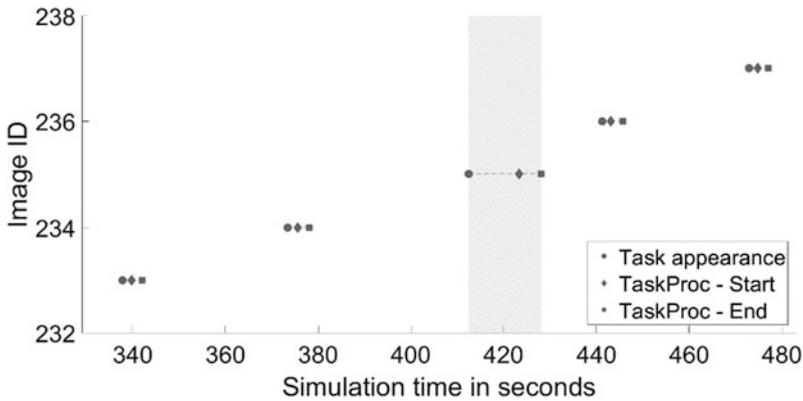


Fig. 6 Timing information for secondary task execution. The *grey area* shows task execution with the delay strategy due to a navigational turn opposed to regular execution in the *non-gray areas*

task appearance (*Task appearance*) is defined by the point in time, when the sensor image is received. Task performance starts (*TaskProc—Start*), when the pilot opens the image browser. Finally the task is completed (*TaskProc—End*), when the pilot presses the submit button to enter the classification result to the system. Images 233, 234, 236, and 237 were processed routinely, while image 235 was delayed, because of a navigational turn.

Reprioritization. Within this behavioral adaptation, the subjects interrupted their working on the secondary task (i.e., object identification task) and focused on the primary task for a short duration, to the primary task in order to avoid violations of the aforementioned constraints (flight altitude, flight path, TOT-Times). After resolving any violations, the subjects switched directly back to the processing of the object identification task. As this strategy comes along with attentional switching it can be identified and quantified using eye-movement analysis. Figure 7 visualizes the regular behavior for the object identification task on the left side, and the reprioritization strategy on the right side. The upper graphs visualize the test subjects’ attention allocation to the cockpit displays and the out-the-window view. For the object identification task, the subjects had to focus their attention on *MHDD1* (Multifunctional Head-Down Display), as this was the display where they received the sensor images. The lower graphs in Fig. 7 visualize the time line of the simultaneous secondary task performance of the object classification task (*Task appearance* (circle), *TaskProc—Start* (diamond), and *TaskProc—End* (square), as described earlier). During a regular object classification task (left side), the subject’s visual attention remained on *MHDD1* during the whole task performance. The task duration was about 4.22 s in average. In cases where the test subjects applied the reprioritization strategy (right side), they started with an attention switch to *MHDD1* and then started the task. Afterwards the subjects interrupted this secondary task to check constraint violations in the primary task in the out-the-window

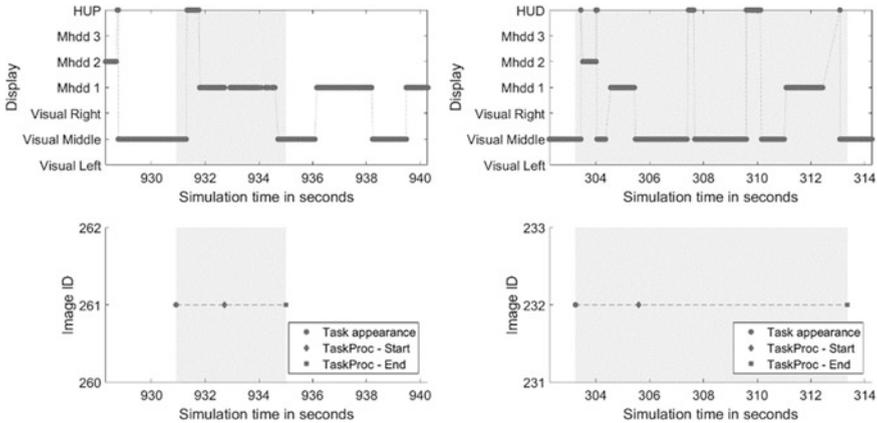


Fig. 7 Example of regular task execution (*left*) compared to the reprioritization strategy (*right*) for test subject 3. The *upper images* show the attention allocation of the subject towards the displays (cockpit displays and visuals), the *lower images* show the execution timeline of the object classification task

view (attention on *Visual Middle*), and finally switched their attention back to *MHDD1* to finalize the object classification task. Due to the reprioritization the object classification task performance took about twice as long as regular.

Batch Processing. This strategy, we interpret as a combination of the strategies *deliberate relaxation of flight constraints* and *delay in object classification task*. The test subjects postpone the secondary task performance until there is a batch of them. Then they heavily abandon from the primary task to temporarily reallocate attentional resources to the secondary task and process the whole batch at once. This reduces the amount of attention switching costs [11] and consequently reduces the overall cognitive work demands. The upper half of Fig. 8 plots the flight altitude over time as well as the terrain elevation, whereas the bottom half shows the

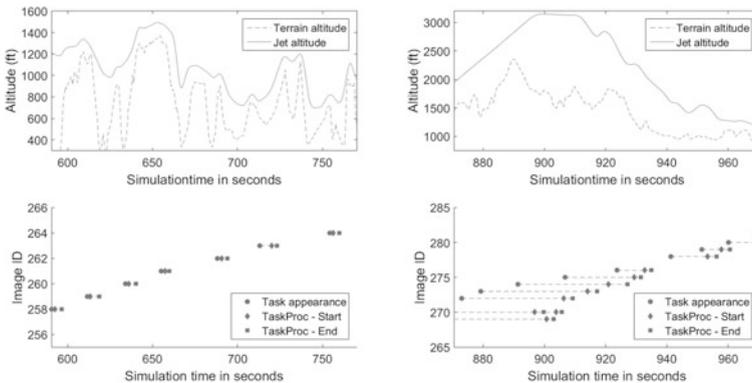


Fig. 8 Example of regular task execution (*left*) versus batch processing (*right*) for test subject 6. The *upper plots* show the flight altitude and the terrain elevation, the *lower plots* show the events in the object classification task

execution of the image classification task over the same time span. On the left hand side, the regular task execution is depicted, while the right side shows the task execution with the batch processing strategy.

3.4 Findings on Usage of Strategies

In general, the utilization of self-adaptive strategies was increased with a rising density of object classification tasks. This was particularly evident for the *batch processing* strategy, which was not used in any mission segment of the easy-easy cluster, but frequently in the hard-easy cluster. In contrast, a comparison between the easy-medium and the hard-medium cluster shows a less frequent usage of the strategies *batch processing*, *reprioritization*, and *delay in object classification task* with a more challenging terrain. This may be linked to an excessive usage of the *deliberate relaxation of flight constraints* strategy, as the altitude violation increased substantially in the hard-medium cluster, and the according subjective workload ratings were not influenced by the terrain. This means that the subjects decreased their performance in the primary task to a level, so that the object classification task could be performed without reaching excessive work demands.

4 Conclusions

Investigating data from our workload experiments with German Air Force pilots, we identified several behavioral adaptations. These adaptations were used by the test subjects to sustain high performance levels in high workload situations. In the mid-term future we want to examine the identified strategies in more detail to integrate the recognition of self-adaptive strategies into automated human machine systems. Therefore, we will need to generate machine runnable models of measured behavior patterns, e.g. by using machine-learning methods. This will allow for objective, non-intrusive, and context-sensitive assessment of workload for the use in adaptive assistant systems.

References

1. Hawkins, K.: "www.army.mil," U.S.Army, October, 6th 2014 (Online). Available: http://www.army.mil/article/135412/Unmanned_aircraft_soar_with_new_capabilities_for_Apache_teaming/. Accessed 25 Feb 2016
2. Baxter, J., Horn, G., Leivers, D.: Fly-by-agent: controlling a pool of UAVs via a multi-Agent System. In: Knowledge-Based Systems (2008)

3. Parasuraman, R.: Adaptive automation for human-robot teaming in future command and control systems. In: Army Research Lab Aberdeen Proving Ground Md Human Research and Engineering Directorate (2007)
4. Honecker, F.: Human-System Interaction Analysis for Military Pilot Activity and Mental Workload Determination, Hong Kong (2015)
5. Sperandio, J.-C.: Variation of operator's strategies and regulating effects on workload. *Ergonomics* **14**(5), 571 (1971)
6. Veltman, J., Jansen, C.: The role of operator state assessment in adaptive automation. In: TNO-DV2 2005 A245 (2006)
7. Sperandio, J.-C.: The regulation of working methods as a function of workload among air traffic controllers. *Ergonomics* **21**(3), 195 (1978)
8. Schulte, A., Donath, D.: Measuring self-adaptive UAV operators' load-shedding strategies under high workload. In: *Engineering Psychology and Cognitive Ergonomics*, Heidelberg (2011)
9. Canham, L.: Operability testing of command, control & communications in computers and intelligence (C4I) systems. In: *Handbook of Human Factors Testing and Evaluation*, p. 433 (2001)
10. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* **52**, 139 (1988)
11. Jones, R.M., et al.: Using cognitive workload analysis to predict and mitigate workload for training simulation. *Procedia Manuf.* **3**, 5777 (2015)

Part V
Supporting Sensor and UAV Users

Model-Driven Sensor Operation Assistance for a Transport Helicopter Crew in Manned-Unmanned Teaming Missions: Selecting the Automation Level by Machine Decision-Making

Christian Ruf and Peter Stütz

Abstract One of the research fields at the Institute of Flight Systems (IFS) of the University of the Armed Forces (UniBwM) focuses on the integration of reconnaissance sensor operation support in manned-unmanned teaming (MUM-T) helicopter missions. The purposive deployment of mission sensors carried by a team of unmanned aerial vehicles (multi-UAV) in such missions is expected to bring in new and impactful aspects, especially in workload-intensive situations. Paradigms of variable automation in the sensor domain and cognitive assistant systems are intended to achieve an operationally manageable solution. This paper provides an overview of the sensor assistant system to be deployed in a MUM-T setup. To manage sensor deployment automation functions, a machine decision making process represented by an agent system will be described. Depending on a workload state input, a suitable level of automation will be chosen from a predefined set. A prototype system of such agent with its capability to react on varied stimuli will be demonstrated in a reduced toy problem setup.

Keywords MUM-T · Multi-UAV · Mission sensors · Human operator · Assistant system · Adaptive automation · Machine decision making · Rational agent

1 Introduction

The R&D project CASIMUS (Cognitive Automated Sensor Integrated Unmanned Mission System) deals with the challenge to support the crew of a two-seated military transport helicopter with sensor equipped unmanned aerial vehicles (UAVs).

C. Ruf (✉) · P. Stütz
Institute of Flight Systems, University of the Bundeswehr Munich (UniBwM),
Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
e-mail: christian.ruf@unibw.de

P. Stütz
e-mail: peter.stuetz@unibw.de

The MUM-T—approach ensures that no further chains of command are involved as these UAVs are directly deployed during missions by the helicopter crew. They shall enable the crew to reconnoiter the intended routes of flight and to survey certain regions of interest such as operation areas or potential landing zones. To achieve such nearby and time-critical reconnaissance, the UAVs and their payload sensor systems have to be scheduled, controlled and sensor output needs to be monitored by the helicopter crew. This is expected to influence the crews' workload through the extension of the crews' task spectrum and a higher overall mission complexity. To counter this potential workload increase, an adaptive, cognitive assistant system [1, 2] provides situation-adapted support by continuous crew supervision and aims to balance crew workload. For UAV platform guidance, a high-level task-based-guidance paradigm has been proposed [3]. Similarly, a corresponding automation layer will be proposed in this paper for the domain of mission sensor deployment, where human centered adaptive automation is expected to reduce and simplify reconnaissance tasks, regarding technical limits of the automated system functions. To select a suitable level of automation, an agent system suggests the automation level by reweighting the current user needs and the available technical capabilities. The sensor assistance system developed in this context is prototyped and tested in simulated MUM-T transport helicopter missions at the Institute of Flight Systems (UniBwM).

2 Problem Scope

2.1 *Automation as a Solution*

The nature of the UAV reconnaissance system used in CASIMUS, containing multiple distributed sensor carrying platforms directly guided from an airborne manned helicopter, differs in many ways from conventional UAV systems. In conventional systems, the platforms are controlled and monitored from a ground control station (GCS), where a dedicated payload operator monitors and controls the sensor system, while the UAV pilot controls and operates the vehicle itself. In the MUM-Teams considered, one of the helicopters crew members has to take on these tasks, most likely the pilot-not-flying (PNF). A task spectrum similar to the GCS' payload operator ones would be added to the PNF's regular tasks. The need to control multiple UAVs and monitor their sensor output by only one operator further aggravates the situation, which may eventually lead to an excessive amount of workload. Moreover, the resulting operative workstation environment differs substantially from those of conventional UAV systems.

A higher degree of process automation would be the classical way to deal with these problems resulting from intensive user involvement, continuous monotonous activity or high human workload. To efficiently incorporate the new sensor operation activities in the regular task range of a PNF, machine perception capabilities

integrated into the sensor systems should to be on his hand as supporting, automated high-level functionality.

2.2 Automation as a Problem

In today's cockpit environments, human crews are used to work with highly automated systems, characterized by reliable and deterministic behavior. Still, complex automation can lead to out-of-the-loop—effects of the crew [4], misinterpretation of automation states (“Opacity-effect” [5, 6]) or relying without question in automation systems (“Over-Reliance” [7]). Wrong belief about the automation systems state and capability is known to be a reason for human failure. Distrust in automation will occur when automation systems don't work in a reliable way. The operators' degree of “trust in automation” is a criterion whether to use automation support at all or perform the task in person [8].

When it now comes to offer automated functions in the domain of mission sensor deployment and data assessment, the problem may be further aggravated due to domain specific technical circumstances. Software based exploitation of data acquired by mission sensors bears the vulnerability to fail or deliver non-reliable results under changing circumstances (e.g. false positive or true negative detections). Mainly, environmental changes (daytime, weather, nature of ground, illumination conditions and signal-to-noise-ratio) are affecting the correctness of delivered output. Specifically computer vision algorithms have system inherent uncertainties that yield to technical result imperfection, especially under a wide operational range. Knowledge about critical conditions for successful usage is often hidden or reserved to expert users.

Furthermore, mission sensor systems carried on UAVs have performance limits, based on physical measurement principle and sensor resolution. Again, expert user knowledge is necessary to cope with these features. To take this into account, the probability of failure [9] during operation must be considered when automatic sensor data processing is employed.

2.3 General Approach

From the above one can summarize that sensor data evaluation can be performed alternatively by machine based automation systems or by the human operator. Machine-based evaluation can process data with high bandwidth, works fatigue-proof, but can't ensure perfect detection performance. Man-based evaluation is exhausting, causes high human workload, only on limited bandwidth, but brings in very high cognitive human capabilities.

Main requirement for a technical system which tackles this antagonism is to achieve a balance between tolerable human workload and necessary reconnaissance

performance. Thus, in our context, the proposed assistant system has to deal with the occurrence of human workload as well as reliability of complex technical systems' automation. To achieve this goal, variable, situation-dependent levels of automation are proposed to be used in this domain to represent adaptive machine capabilities combined with different stages of human involvement.

3 Related Work

This chapter aims to relate the proposed approach to background paradigms and previous work.

In [10], three basic requirements for assistant system behavior were proposed to be applied by assistant systems for crew support. Here the assistant system acts on a stepped de-escalation scheme. It ranges from guiding the crews' attention to the most urgent task over transferring a complex task situation in a manageable one up to allocating specific automated means for the execution of tasks that are not accomplishable by the crew [1]. Applying these design requirements to a sensor assistant system fosters the idea of situation-dependent crew support and machine-process execution. In this context the paradigm of a variable degree in automation is referred to as *adaptive automation* or *variable automation*. Different level arrangements result in stepped involvement of the human operator [11, 12]. Adaptive automation systems select a suitable level of automation or choose a situation-adequate mode depending on the current context [13]. Changes in automation degree can be initiated by both, the human operator and the technical system. The system is specified as adaptable (if the user invokes changes) or adaptive (change initiative by automation itself) [14]. As pointed out above, offering variable levels of automation shall be used in the proposed sensor assistant system to counteract work-load intensive situations.

The guidance of multiple UAVs out of a helicopter cockpit with support of an cognitive assistant system was successfully demonstrated in simulated missions [2] in a past R&D project. However there, idealized sensor deployment was assumed, which did not require the crew to interact with the UAVs sensor system. The assistant system proposed in this paper now should also support the crew when operating more state-of-the-art payload systems.

Also in [2] an error-free, nearly fully automated target recognition (ATR) system approximated sensor data processing. In CASIMUS now data processing will be aspired which includes typical artefacts and imperfections, in order to investigate on the reliability issue discussed above. Therefore, a "Sensor- and Perception-Management System" framework [15, 16] (SPMS) shall be integrated which hosts mission sensor data processing algorithms. This framework administers environment—and context-adapted machine perceptive capabilities by selection and application of appropriate image processing algorithms. High level capabilities realized this way can be applied for typical airborne surveillance tasks such as aerial

mapping as well as object detection and tracking. In addition, each algorithm is accompanied by a “trustworthiness” figure which assesses its reliability in the given circumstances.

4 Sensor Assistant System

Figure 1 gives an overview of the system blueprint for offering sensor deployment support, implemented along the requirements set out above. The figure shows three divisions:

- *Automation Functions*: a repository of installed machine-performable automation functionality in the domain of Computer Vision and Gimbal Control
- *Management*: A decision mechanism to determine the appropriate automation degree depending on different input conditions. Results are forwarded to the operator or, under specific conditions executed automatically.
- *Knowledge Models*: models of human working processes and their costs of human resources allocated occurring during performance for different automation levels (“Sensor Taskmodel”) like illustrated in [17]; model of automation levels (“LOA”)

The sensor assistant system receives several inputs. Two are of main interest, the workload measure of the human operator (provided by an external system [18]) and the computer vision algorithms’ trustworthiness (provided by the “Sensor- and Perception-Management” System [15, 16]). These inputs are model-based values; the workload is a taskmodel-based construct, and the trustworthiness is delivered by

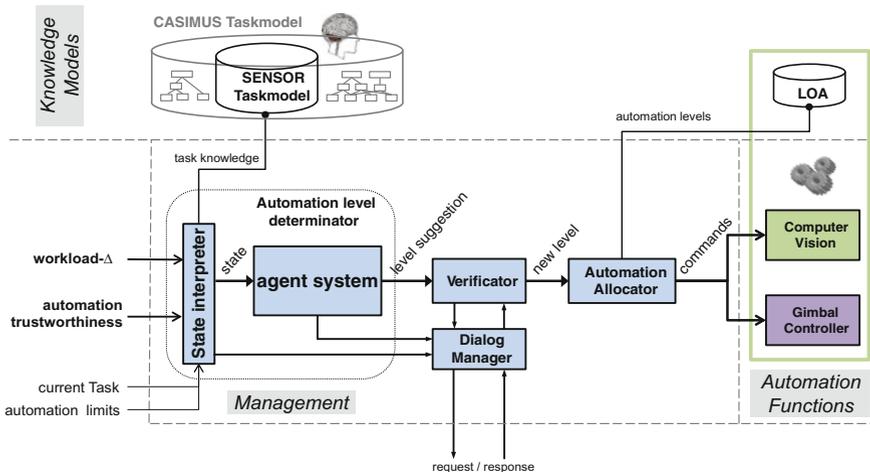


Fig. 1 Sensor assistant system overview

a reliability estimation process. The sensor assistant systems' goal is to allocate the best fitting automation function setup for varying task, workload and automation trustworthiness parameters. To react on this input signals, an agent system was designed, which is able to propose changes in the automation level. These level suggestions can be verified by superordinate authorities, before the automation allocation for those levels takes place.

4.1 *Sensor Automation Functions*

In this chapter the available automation functions are explained in more detail.

“Computer Vision.” Automated functions in this group are intended to relieve the human operator from perceptive-cognitive activity (e.g. interpreting acquired sensor data), which demands high user involvement, specifically when it comes to monitoring parallel data sources from teaming UAVs. Therefore, a spectrum of different task-related algorithms from the computer vision domain has been implemented and integrated. The implementation of the algorithms follows the modular scheme as laid out in [16]. Further, variations of sensor data access and data visualization functions are available (e.g. mosaicking of aerial images).

“Gimbal Controller.” The mission sensors onboard the UAVs considered are typically attached to a cardanic mount (*gimbal*) to enable dynamic sensor gaze control during flight. The degrees of freedom of the two axis demand for continuous user steering when special ground tracks or ground patterns are required to be scanned. An automated gimbal control application was developed to decrease the humans' sensorimotor activity when operating the gimbal. The automation functions offer several control modes that are useful for the existing task spectrum.

4.2 *Levels of Automation*

As described in Sect. 2.3, the two aspects human workload and technical system reliability should be balanced. For this purpose, different levels of automation from the repertoire of automated functions described above had to be laid out. According to the scheme of “Levels of Automation” [11, 12], several levels with different depth of user involvement are suggested. This ensures, that the distribution of task load can be varied between man and machine purposefully.

Motivated by the aspect of human workload reduction during task performance, the level design proposed foresees in one direction an increasing automated support which potentially results in workload reduction. However high-levels of automation through computer vision bear the potential of high machine unreliability. This aspect indicates to prefer lower level of automation and therefore more robust data processing. This leads to the assumption that moving towards lower level of automation increases trustworthiness but also increases workload and vice versa.

This interrelation leads to the effect, that human workload can be reduced as long as systems' reliability does not undercut a certain value. In the opposite direction, the fallback in lower levels, induced primarily by low automation trustworthiness, involves the human tighter, but ensures that the automation delivers more trustable results.

Automation Levels for Sensor Data Evaluation. Figure 2 shows the arrangements and composition of image processing and information presentation functionalities as levels. These levels can be mapped on the left side scale indicating the abstraction level and presentation properties of resulting surveillance data.

With upwards level steps, data will be transformed from signal (image) data to abstract information, similar as pointed out in [19]. This efficiently can reduce the bandwidth of data to be analyzed by the operator and supports their spatial correlation to deduce mission relevant results. Further, the representation form of data can be varied from an "image stack" (chronology of images/Full Motion Video) to a spatial map representation by either image stitching (level 2) or spatial arranged object information on a tactical map (level 3). The red/blue schematic in above figure thereby symbolizes the expected "workload" division between human and machine system.

In terms of trustworthiness, moving on levels upwards introduces machine imperfections step by step. Applying computer vision methods to detect objects on level 1 may be prone to high false-positive and true negative detections. Additionally when moving to level 3, false classifications can contribute to a further reduction in trustworthiness.

In contrast, the selective level adaption in downwards direction is aimed to involve the operator when automated functions appear to fail, thus introducing a more intense cognitive fusion performance on human side [9], where preprocessed data and information provided by maximally exploited automation capabilities are delivered to the crew for confirmation and/or further information extraction.

Automation Levels for Gimbal Controlling. The domain of payload steering also offers variable automation, Fig. 3 shows the automation functions' usage graded in four levels.

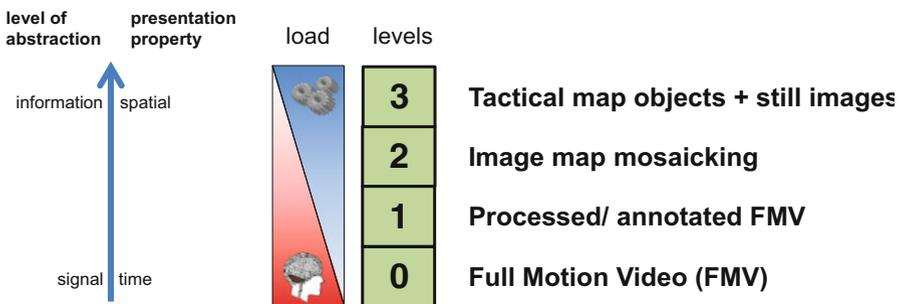


Fig. 2 Levels of automation in data evaluation

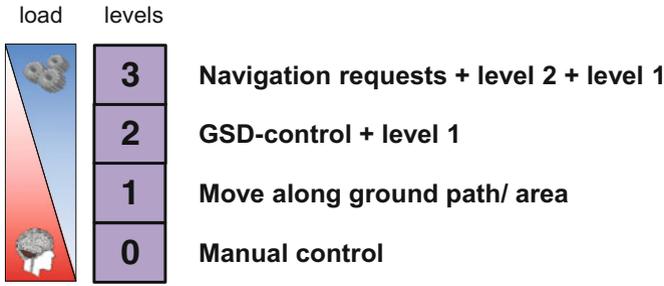


Fig. 3 Levels of automation in payload gimbal automation

The automated sensor control functions offer the capability to scan ground patterns of routes and areas automatically (level 1). In higher levels, system initiated zoom adaptations update the sensors' ground sample distance (GSD) automatically (level 2). The highest automation level (level 3) issues navigational requests to the platform's FMS (flight management system), requesting platform repositioning. These levels should reduce the human's sensorimotor activity, control process complexity, system process supervision times and intervention demands.

5 Machine Decision Making for Selecting the Automation Level

5.1 Working Method

When using the suggested automation levels in a (simulated) surveillance UAV scenario, the critical parameters that initiate the automation level adaption need to be observed as base for the decision process. Figure 4 shows the functional core that is aimed to determine the necessary change of the automation degree based on the two main input variables. First of all, the humans' workload is considered as the main indicator for the decision process to act. In addition, also the automation functions trustworthiness has to be taken into regard, as motivated above, to find the best available solution (automation level) from the predefined set to satisfy the workload reduction request.

The core of the automation management consists of an artificial agent system, which utilizes of a *Markov Decision Process* (MDP). With regard to modelling, this method was chosen because the problem comprises full control ability about the state transitions (in contrast to *Markov Chain* or *Hidden Markov* models) and, in this first prototype, completely observable states were assumed (in contrast to *Partially Observable MDP*). Such agent system, according to [20], can be classified as: an artificial, controlled, reactive, individualistic, conservative, nonreasoning no perception and no memory agent.

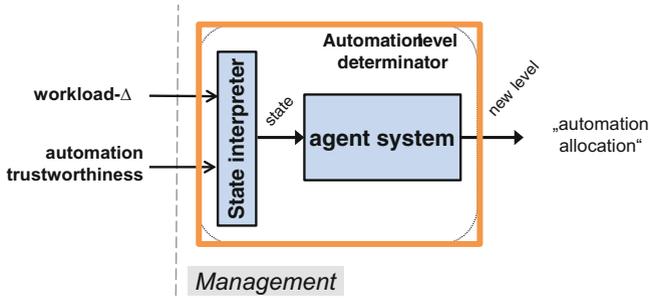


Fig. 4 “Automation level determinator” based on an agent system

First, a manageable state scenario had to be found, on which the MDP operates. The MDP consists of set of such internal states S , a set of actions A to transfer between the states when they are applied, a transition model T (which defines the transition probabilities between the states) and a reward model R (defines rewards for applying actions in the states).

By this a state space model was build (Fig. 5). The five states S represent the impact of automation usage on the HC-Crew and the consequence of imperfect automation. The available agent actions A describe five different modifications of the automation degree. The agent states and actions were defined as:

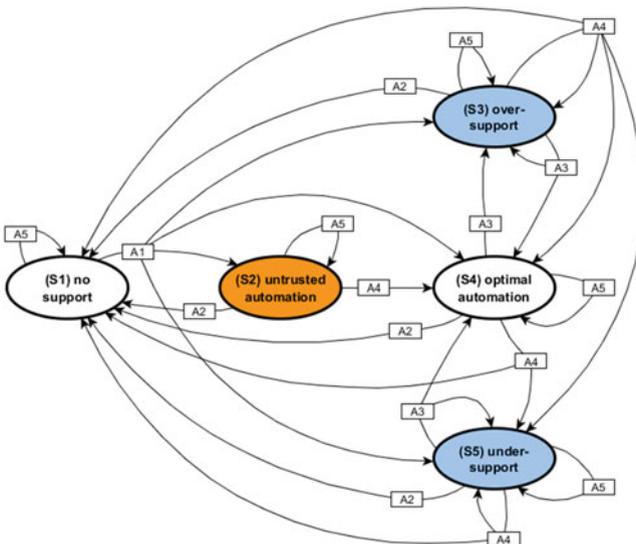


Fig. 5 State space transition model

- $\mathbf{S} = \{S1, S2, S3, S4, S5\} = \{\text{"no Support"}, \text{"untrusted automation"}, \text{"over-support"}, \text{"optimal automation"}, \text{"under-support"}\}$
- $\mathbf{A} = \{A1, A2, A3, A4, A5\} = \{\text{"start Automation"}, \text{"stop Automation"}, \text{"increase level"}, \text{"decrease level"}, \text{"keep current state"}\}$

Figure 5 shows the state space transition model with the feasible transitions (emitting specified agent actions) between the states. The states regard human workload (blue marked in Fig. 5) and automation trustworthiness (orange marked in Fig. 5) aspects. The actions are affiliated with specified rewards, directing the desired agent decision behavior.

The solution of this MDP—problem produces a policy that yields to the highest reward for action application. This policy is a list of optimal agent actions that are applied in the different states.

On stimulation, the agent’s inner system state is updated by an observation. The state interpreter in Fig. 4 determines a current state. On this state S , a specified action A is applied, and a subsequent state S' will be reached. The agents’ objective is now to reach the inner target state $S4$ “optimal support” by applying selected actions, which is aimed to transform the environment observed by the agent system (human workload) to an optimized range (minimized workload). The following chapter shows the automation management system described above working in a reduced toy problem setup.

5.2 Proof of Concept in a Toy Problem Setup

During mission performance, the human workload is assumed to vary over time, mainly depending on the performance of tasks for mission fulfilment. When using automated sensor systems in mission environments, the productivity and efficiency of machine recce capabilities may change (expressed by the automation trustworthiness), caused by varying environment conditions, viewing distances, viewing angles. Hence, the best fitting automation level has to be determined online whenever such changes occur.

To demonstrate the sensor assistant systems’ capability to handle such situations, a set of input variables was created, comprising a trend of measures of human workload and automation trustworthiness (Fig. 6). In this trend, a fixed threshold (“minAutomationTrustworthiness” in Fig. 6) of 75 % is assumed to classify automation as suitable to support.

In the toy problem setup, the received input triggered automation level adaption activities. From the trend in Fig. 6, four typical use cases (1–4) were taken that were to be handled and solved by the agent at discrete time steps. The agents’ actions were applied on these use cases and the automation level adaption took place.

The assistant systems’ outputs are presented in Fig. 7. Depicted in magenta are the agents’ decisions and the consequential changes between the automation levels.

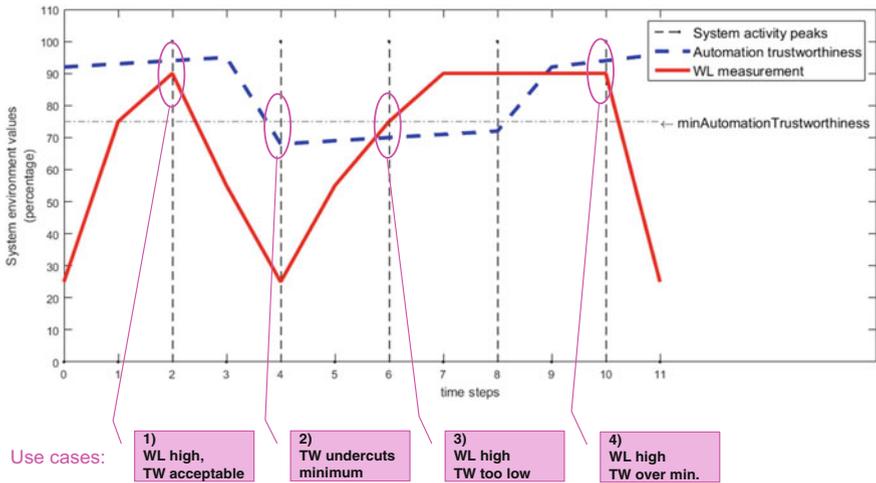


Fig. 6 Timeline of schematic *input signals*, *use cases*, a *value threshold* and *activity peaks*

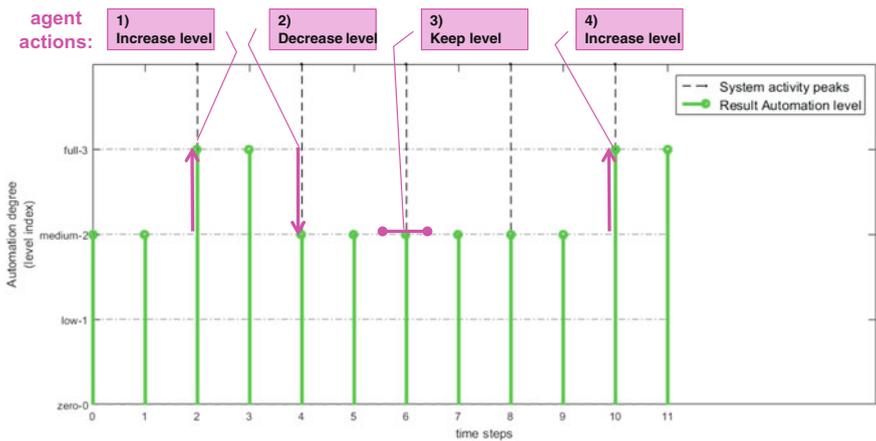


Fig. 7 Timeline representation of *agent actions* and resulting *automation levels*

In this setup, a high base workload of the HC-Crew was assumed, so the sensor automation runs already in level 3 of 4 from the beginning.

As shown in Figs. 6 and 7, the agent automatically reacts to the varying input parameters mentioned above. High human workload initiates automation degree increase, as long as minimum trustworthiness margin is not undercut (use case 1). Too low automation reliability decreases the automation level (use case 2) or prevents from changing to an automation level that potentially induces higher automation failure rates (use case 3). At last, the assistant system notices the return of sufficient automation trustworthiness and initiates a level increase if necessary (use case 4).

With the application of the agent decisions on the use cases illustrated above it can be demonstrated that the agent systems' models and actions are suitable to handle the different situations occurring in this problem field.

6 Realization Process and Future Work

After this first functionality proof of the decision making core, the transfer of the concept in an executable implementation will follow. Therefore, the system will be interfaced to make use of the combined task-, resource—and interaction model (“Sensor Taskmodel” in Fig. 1) to regard task-specific differences.

For functional demonstration of the holistic HC-assistant system applying sensor automation with usage of the proposed sensor assistant system, a closed-loop-operation will be realized. Finally, the evaluation of the concept and the effects on operator performance will follow in human-in-the-loop experiments with transport helicopter crews.

References

1. Onken, R., Schulte, A.: System-ergonomic design of cognitive automation: dual-mode cognitive design of vehicle guidance and control work systems. Springer Publishing Company, Incorporated (2012). http://www.springer.com/gp/book/9783642031342?wt_mc=ThirdParty.SpringerLink.3.EPR653>About_eBook
2. Strenzke, R., Uhrmann, J., Benzler, A., Maiwald, F., Rauschert, A., Schulte, A.: Managing cockpit crew excess task load in military manned-unmanned teaming missions by dual-mode cognitive automation approaches. AIAA Guid. Navig. Control Conf. (2011)
3. Uhrmann, J., Schulte, A.: Concept, design and evaluation of cognitive task-based UAV guidance. *J. Adv. Intell. Syst.* **5**(1) (2012)
4. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. *Hum. Factors J. Hum. Factors Ergon. Soc.* **37**(2), 381–394 (1995)
5. Billings, C.E.: *Aviation Automation: The Search for a Human-Centered Approach* (1997) Lawrence Erlbaum Associates, Incorporated, NJ, USA
6. Wiener, E.L.: Human factors of advanced technology (glass cockpit) transport aircraft. (Nasa-Cr-177528), 222 (1989)
7. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. *Hum. Factors J. Hum. Factors Ergon. Soc.* **39**, 230–253 (1997)
8. Muir, B.M.: Trust in automation: Part I. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* **37** (1994)
9. Llinas, J., Liggins, M.E., Hall, D.L.: *Handbook of Multisensor Data Fusion: Theory and Practice*. CRC press (2008)
10. Onken, R.: Funktionsverteilung Pilot-Maschine: Umsetzung von Grundlagenforderungen im Cockpitassistenzsystem CASSY. In: Gärtner, K.-P. (ed.) *DGLR-Bericht 94–01*, Berlin: Deutsche Gesellschaft für Luft- und Raumfahrt (1994)
11. Sheridan, T.B.: adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: distinctions and modes of adaptation. *Syst. Man Cybern. Part ASyst. Hum. IEEE Trans.* **41**(4), 662–667 (2011)

12. Endsley, M.: Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 462–492 (1999)
13. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **30**(3), 286–297 (2000)
14. Scerbo, M.W.: Adaptive automation. In: Karwowsky, W. (ed.) *International Encyclopedia of Human Factors*, pp. 1077–1079. Taylor & Francis, London, U.K. (2001)
15. Russ, M., Schmitt, M., Hellert, C., Stütz, P.: Airborne sensor and perception management: a conceptual approach for surveillance UAS. In: *Information Fusion (15th FUSION)* (2012)
16. Hellert, C., Smirnov, D., Russ, M., Stütz, P.: A High Level Active Perception Concept For UAV Mission Scenarios. *Dtsch Luft- und Raumfahrtkongress* (2012)
17. Ruf, C., Stütz, P.: Ergonomische Einbindung des Sensor-Operators in eine MUM-T/Multi-UAV Umgebung: Problemanalyse, Konzeptdarstellung und erste Modellbildung. *Kooperation und kooperative Systeme in der Fahrzeug- und Prozessführung* **2015–01**, 79–96 (2015)
18. Honecker, F., Schulte, A.: Konzept für eine automatische evidenzbasierte Online-Pilotenbeobachtung in bemannt-unbemannten Hubschraubermissionen. In *4 Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten*, Bielefeld (2015)
19. Rohr, K.: *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Kluwer Academic Publishers, Norwell, MA, USA (2001)
20. Burgin, M., Dodig-Crnkovic, G.: *A Systematic Approach to Artificial Agents* (2009). <http://arxiv.org/abs/0902.3513>

Using Natural Language to Enable Mission Managers to Control Multiple Heterogeneous UAVs

Anna C. Trujillo, Javier Puig-Navarro, S. Bilal Mehdi
and A. Kyle McQuarry

Abstract The availability of highly capable, yet relatively cheap, unmanned aerial vehicles (UAVs) is opening up new areas of use for hobbyists and for commercial activities. This research is developing methods beyond classical control-stick pilot inputs, to allow operators to manage complex missions without in-depth vehicle expertise. These missions may entail several heterogeneous UAVs flying coordinated patterns or flying multiple trajectories deconflicted in time or space to pre-defined locations. This paper describes the functionality and preliminary usability measures of an interface that allows an operator to define a mission using speech inputs. With a defined and simple vocabulary, operators can input the vast majority of mission parameters using simple, intuitive voice commands. Although the operator interface is simple, it is based upon autonomous algorithms that allow the mission to proceed with minimal input from the operator. This paper also describes these underlying algorithms that allow an operator to manage several UAVs.

Keywords Unmanned aerial vehicles · Voice commands · Autonomy · Coordinated flight · Mission operator

A.C. Trujillo (✉)

NASA Langley Research Center, MS 492, Hampton, VA 23681, USA

e-mail: anna.c.trujillo@nasa.gov

J. Puig-Navarro · S.B. Mehdi

University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

e-mail: puignav2@illinois.edu

S.B. Mehdi

e-mail: mehdi1@illinois.edu

A.K. McQuarry

Analytical Mechanics Associates Inc, Hampton, VA 23681, USA

e-mail: andrew.k.mcquarry@nasa.gov

© Springer International Publishing Switzerland 2017

P. Savage-Knepshield and J. Chen (eds.), *Advances in Human Factors in Robots and Unmanned Systems*, Advances in Intelligent Systems and Computing 499,
DOI 10.1007/978-3-319-41959-6_22

1 Introduction

Small unmanned aerial vehicles (sUAV) are starting to become ubiquitous because they are relatively cheap and are fairly easy to fly while the potential immediate productivity gain is large for applications such as photography, inspection, and package delivery. As more people find innovative ways to employ sUAVs [1]—such as crop monitoring [2], photography [3], filming [4], package delivery [5], pipeline inspection [6], search and rescue (SAR) [7], and fire monitoring [7] just to name a few—the way humans interact with them will become critical. Current interaction methods typically include manual controllers [8, 9], smartphones [10] and tablets [11], or graphical ground control stations (GCS) [12, 13]. Interacting with all these types of controllers requires the operator to learn and understand the specifics of the controller and also the sUAV's dynamic behavior rather than having a more natural and higher level teaming relationship with each vehicle. A lack of teaming typically results in increased workload, decreased situation awareness, and trust issues among all active agents. However, with the possibility of communicating with various types of unmanned vehicles (UxVs) by more natural language (NL) methods, such as speech and gestures, the teaming aspect may come to full fruition. NL may also decrease workload [14] and increase situation awareness [15].

Using speech recognition to give commands is becoming pervasive, especially as speech commands move from controlled to natural language. Many people are now comfortable with speech commands beyond the primitive phone tree systems, such as Apple's Siri, Microsoft's speech recognition system, and Google's Now. Great progress has been made in these systems to understand human speech without training [16–19]. However, the word error rate is still rather high for typical conversational speech recognition especially in noisy environments [20]. Various methods to improve on this have been implemented [21] but have only been partially successful.

An area that may benefit from speech recognition is a dispatch scenario with a dispatcher scheduling multiple package deliveries to a defined neighborhood. This application highlights several aspects regarding UAV control by untrained UAV operators. Allowing UAV operators or mission managers with no UAV pilot expertise to control multiple vehicles is critical to fully realize the new missions that UAVs enable; however, more importantly, the inherent ease to control and the stability of many off-the-shelf small UAVs enable casual users to command and control these vehicles without the need for knowledge of aerodynamics, stability and control, weight and balance, etc. Thus, the goal of this research is allowing an inexperienced UAV pilot, an operator, to define and manage a mission. This mission may entail several heterogeneous UAVs flying coordinated patterns or flying multiple trajectories deconflicted in time or space to locations defined by a dispatcher. This mission may be accomplished with a relatively simple interface. For the package dispatcher, this interface allows a dispatcher to easily define locations for packages to be delivered to and then to easily generate and inspect the deconflicted flight paths to ensure on-time delivery.

This paper describes the functionality and preliminary usability measures of an interface that allows an operator, in this case a dispatcher, to define a mission of delivering packages with multiple coordinated UAVs and then to start the mission. This interface includes using natural language, in this case speech, to make inputs beyond the traditional input methods of keyboard, mouse, and touchscreen. With a relatively well-defined, simple vocabulary and using open-source speech-recognition software, the mission manager can input the vast majority of the mission parameters using simple, intuitive voice commands. Furthermore, although the interface is simple, underneath are autonomous algorithms that allow the mission to proceed with minimal operator input. On-going work at NASA Langley Research Center's Autonomy Incubator and the University of Illinois at Urbana-Champaign regarding the underlying algorithms that allow an operator to manage several UAVs is also described.

2 Initial Voice Usability Experiment

An experiment was conducted to begin to measure the efficacy and user acceptance of using voice commands to define a multi-UAV mission and to provide high-level vehicle control commands such as "takeoff" and "land."

2.1 Independent Variables

The primary independent variables were input type (voice or mouse) and order used. Half of the subjects used the mouse input method first and then used voice, while the other half of the subjects first used voice input and then the mouse.

Subject was included as an independent variable because so few were run because this was an exploratory effort. None of the subjects had used a speech recognition system for input before besides traditional phone trees.

2.2 Dependent Variables

The primary dependent variables consisted of the correctness of the mission parameter inputs and the time needed to make all inputs. Other dependent variables were NASA-TLX workload ratings [22] and subjective ratings on a final questionnaire. The NASA-TLX and subjective questionnaire ratings were all normalized to a 100-point scale and all measures were continuous within that scale.

Fig. 1 Package delivery setup screen. The pull down menus and the length input were options. The commands were the pushbuttons such as “Done” and “Takeoff.”

2.3 Procedure

The experiment required each subject to fill in an online form that contained comparable required information that would be needed for a package dispatcher to deliver packages (Fig. 1). The input screen was programmed in Matlab 2015b[®].¹ For each run, subjects typed in a simple numeric code for the package code. Then, they defined the initial starting position (“From”), the delivery location (“To”), and the return location (“Return”) using either pull-down menus or voice input. Voice input was accomplished using CMU Sphinx4-5prealpha [23] for speech recognition. Next, they inputted the length of the package, which are detailed in Table 1. If another package was to be added, they indicated that by the “Add Another Package” or if done entering packages, the subject indicated “Done.” The subject had the system “Calculate Trajectory” and then “Takeoff” once the trajectory was calculated. Later, the subject used “Land” to finish the run. These last commands (i.e., “Add Another Package”, etc.) were considered command fields.

After the voice and mouse input experiment runs, all subjects completed a NASA-TLX. At the conclusion of all runs, subjects completed a questionnaire asking them about their experience in inputting the mission parameters, and starting and stopping the mission using both mouse and voice input.

¹The use of trademarks or names of manufacturers in the report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

3 Results

Because this was a preliminary experiment to test the methodology, only four subjects were run. Therefore, only averages and standard errors of the means will be reported with comments on trends.

3.1 Parameter Input Accuracy

All subjects inputted the parameters for the package delivery specification (i.e., “To”, “From”, “Return”, and “Length”) with minimal errors. Out of the 288 inputs, there were only five errors (<2 % error rate) split between mouse and voice inputs. Therefore, both methods of input appear to be accurate.

However, this initial test was done in a quiet environment with a limited/restricted/controlled vocabulary (see Table 1) and no homophones. Thus, the accuracy of the speech recognition was high [24]. In a noisier environment, possibly with multiple persons talking simultaneously, using push-to-talk headsets or a key word to wake up the system [25] may limit inadvertent voice activation.

3.2 Input Time

Input Specifications. In general, subjects used slightly more time to input package information using voice input (Figs. 2a, b and 3). The required times for voice input

Table 1 Option fields accepted inputs

Option field	Possible inputs
Package code	Integer 1–9
From	Current location Forest Net Red house Yellow house Green house
To	Red house Green house Yellow house
Return	Round trip Forest Net South lake Red house Green house Yellow house
Length	Integer 1–9

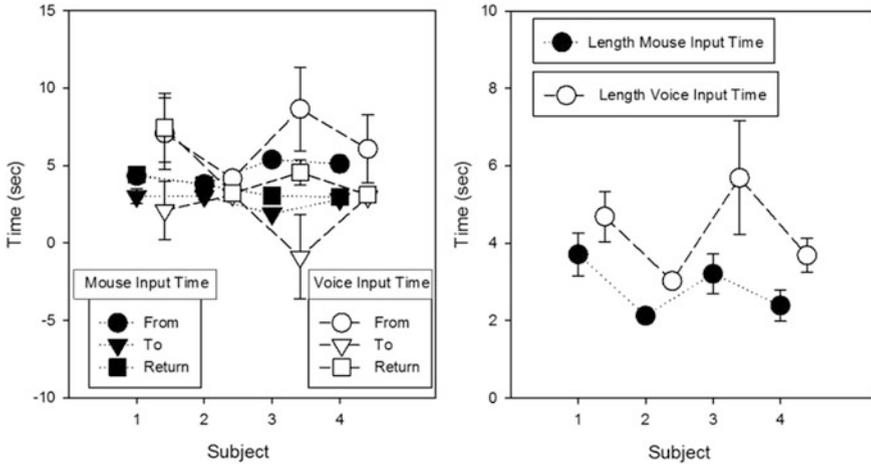


Fig. 2 a Time required for each subject to input popup menu information by input type. Popup menu information was to, from, and return information **b** Time required for each subject to input package length information by input type

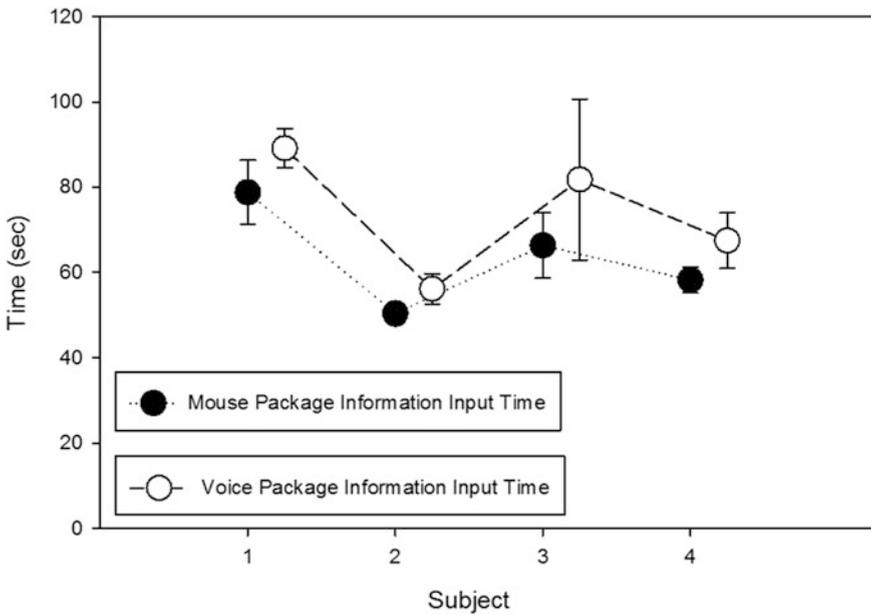


Fig. 3 Total time needed to input all package information for each subject by input type

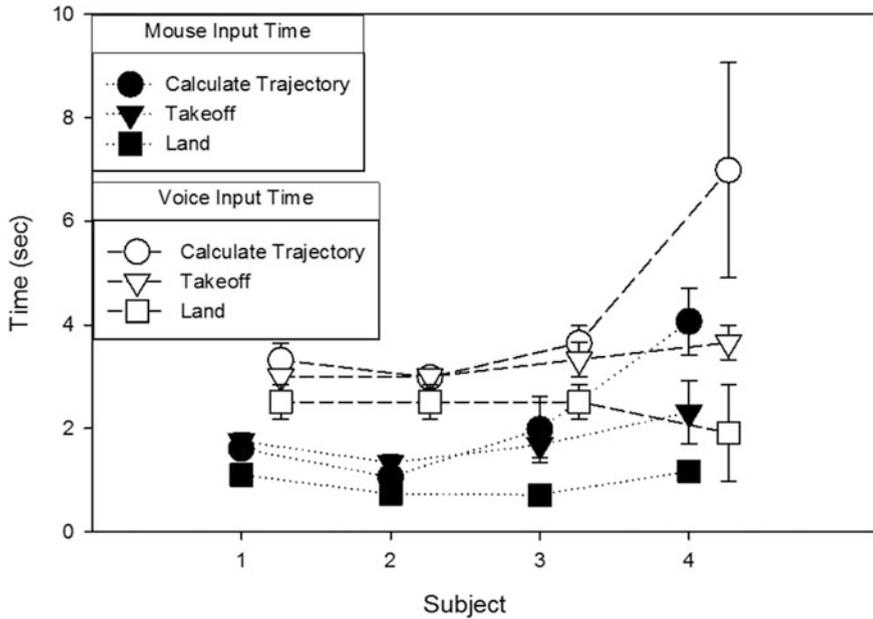


Fig. 4 Input command times for each subject by input type. The input commands consisted of “Calculate Trajectory,” “Takeoff,” and “Land.”

may have been influenced by the length of the phrase. Some of the “From” and “Return” phrases were up to three words while the “To” phrases were always two words. The speech recognition system’s parsing times for these longer phrases may have increased those times overall.

Command Input. The time required by subjects to input commands was generally greater when using voice commands (Fig. 4). Also seen in Fig. 4, the same pattern for each subject when using mouse input was repeated for voice input except it took more time for voice input. Therefore, for commands that are mission critical or safety related (e.g., the requirement to land immediately), it may be necessary to include some type of screen input—such as mouse or touchscreen—in these cases. Lastly, once again the longer phrases (in this case, “Calculate Trajectory”) took more time to register when using voice input.

3.3 Workload

The workload in using mouse and voice commands were relatively equal with the voice commands requiring slightly less workload for inputs (Fig. 5). The temporal demand and frustration level between the two input methods were essentially equal. This may have been due to the time it took the voice recognition system to parse the

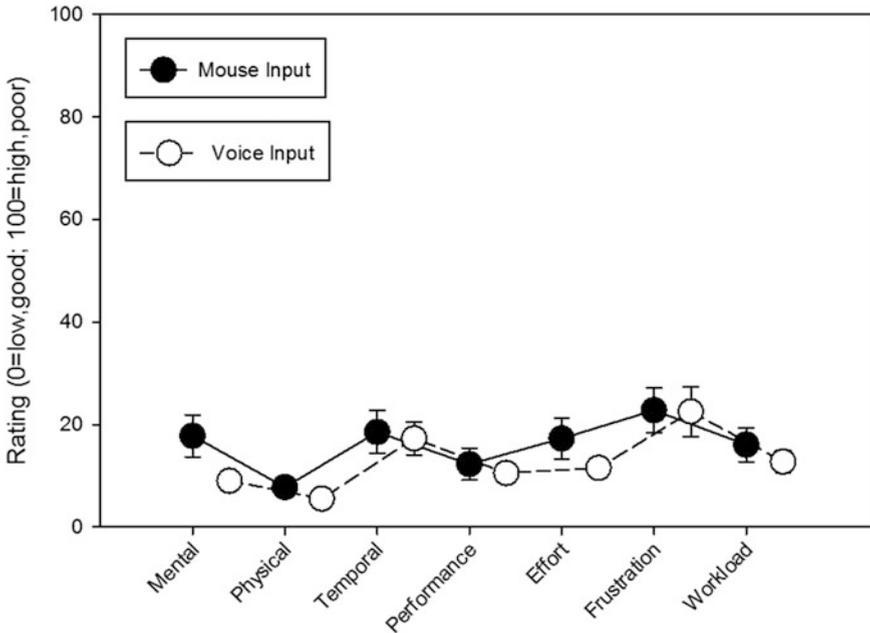


Fig. 5 NASA-TLX ratings across all subjects by input method

voice input and the timing of the input screen in checking if there was a new voice command. Furthermore, the subjects had no immediate indication on the input screen regarding the voice recognition system and its output until it registered on the screen in one of the input fields. Therefore, to decrease the temporal demand and frustration level when using voice input, some type of feedback from the voice recognition system may be required.

3.4 Subjective Preferences

In general, subjects rated using mouse input as slightly easier than voice input when inputting options (*e.g.*, “From”) and commands (*e.g.*, “Takeoff”), and in the general ease of use (Fig. 6). For critical input, such as “Land,” subjects preferred using mouse input. Again, this may have been due to the responsiveness and time required for the speech recognition system. Surprisingly, subjects indicated that the responsiveness of the mouse was slower than that of voice input. Some subjects indicated that it was tedious to move the mouse around to input the package information. This may have contributed to the mouse input being slower than the voice input. Lastly, subjects had a slight preference for using mouse input. However, their preference of using the mouse for critical input may have swayed

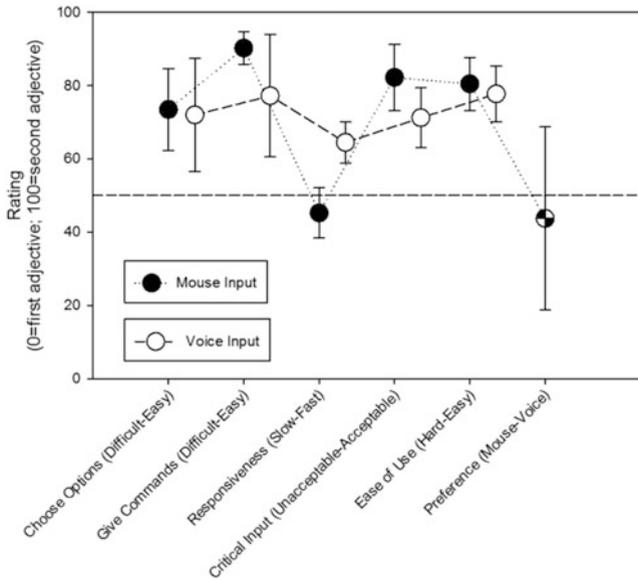


Fig. 6 Questionnaire data

this overall preference. Therefore, voice input may be acceptable for non-critical input such as when vehicles are not yet airborne; whereas, direct screen input may be better for critical commands that must be executed immediately.

4 Coordinated Flight Path Generation and Following

4.1 Trajectory Generation

Once the packages have all been entered into the system, the system must be able to generate multiple trajectories that consider each vehicle's dynamics and operating characteristics, ensure collision-free maneuvers, and guarantee the desired inter-vehicle coordination for the specific mission [26]. An example algorithm that considers each vehicle's dynamics and coordinates the vehicles in space and time in order to generate each vehicle's trajectory is detailed in [27–30] (Fig. 7). This methodology employs Pythagorean-Hodograph Bézier curves that guarantee satisfaction of boundary conditions, dynamic constraints, and timing schedule of each vehicle. Consequently, the trajectories are provably correct and ensure a safety inter-vehicle distance. Also, the path-following and time-coordination algorithms that complement this autonomous framework have known stability guarantees [31, 32]. These guarantees may engender a higher level of trust in the mission operator that the UAVs will safely arrive at their destinations. This trust will enable a higher

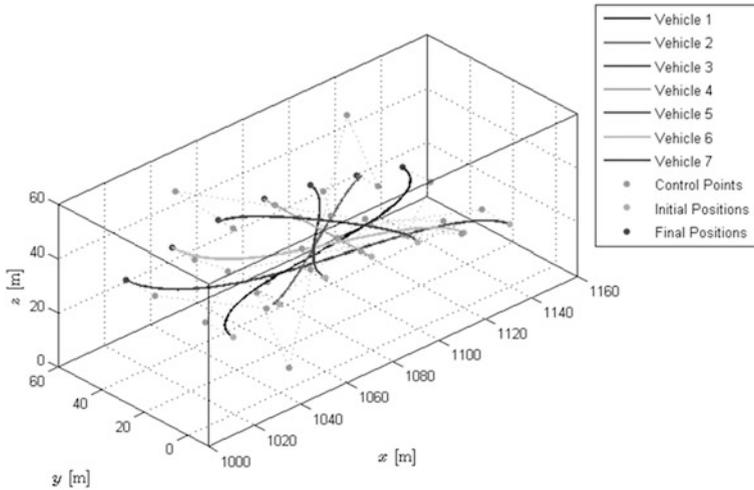


Fig. 7 Three-dimensional temporally deconflicted flight trajectories of 7 multicopters

functioning system and will facilitate teaming amongst all the autonomous agents whether they are human or machine.

4.2 Collision Avoidance

Once the vehicles takeoff, they must have the ability to replan their trajectories avoid obstacles. Example algorithms that guarantee avoidance along with satisfaction of mission constraints and vehicle dynamic constraints are presented in [33–35] (Fig. 8). In general, these algorithms must first predict a collision and then replan the vehicle’s trajectory to avoid the collision. Furthermore, these algorithms are able to avert a possible collision with cooperative or uncooperative obstacles without foreknowledge of the trajectory but with only an online, inaccurate prediction of the obstacle’s trajectory. In fact, collision avoidance can be achieved with only the knowledge of the line-of-sight angle only [33].

4.3 Intervehicle Communication

Vehicles will need to coordinate with one another to arrive at a destination at the same time, at prespecified times, or in a time window so as to meet given temporal separation requirements [36] (Fig. 9). In this case, the communications network must ensure adequate communication between the vehicles [32, 37]. In general, the vehicles communicate over a time-varying network, where the quality of service

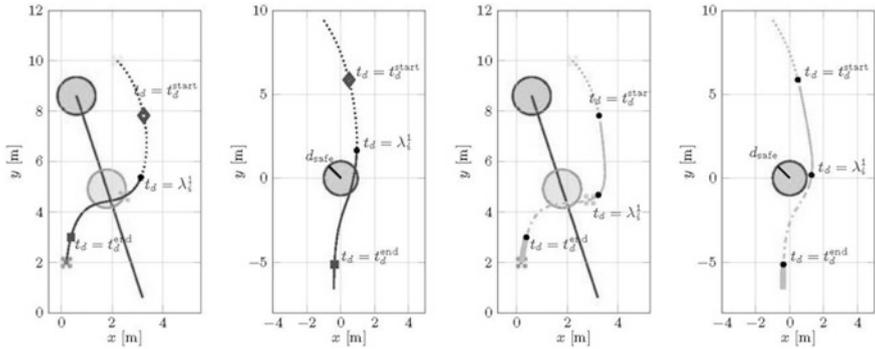
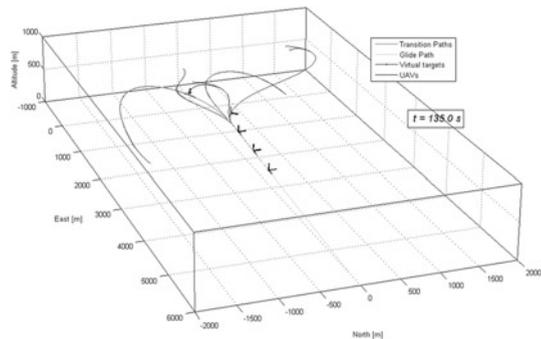


Fig. 8 Illustration of the collision avoidance algorithm: first panel shows the original trajectory (*curved line*) along with that of the obstacle (*straight line*). The separation curve, shown in the second panel, overlaps with the obstacle region (*circle*) indicating a collision. The third panel shows the replanned trajectory (*curved line*) that ensures collision avoidance. The corresponding separation curve is shown in the fourth panel. Notice that this separation curve remains outside the obstacle region

Fig. 9 a Five UAVs arrive at the beginning of glide path within pre-specified arrival windows and separated by approximately 30 s



inevitably determines the performance bounds of the coordination algorithm. Once again, performance guarantees may engender a higher level of trust with the mission operator; thus enabling a higher functioning system.

5 Conclusions

The availability of highly capable, yet relatively cheap, unmanned aerial vehicles (UAVs) is opening up new areas of use for hobbyists and for commercial activities. The goal of this research is a relatively inexperienced operator to define and manage a mission using voice commands. This mission may entail several heterogeneous UAVs flying coordinated patterns or flying multiple trajectories deconflicted in time or space to predefined locations.

In general, the usability of using voice commands is acceptable. With a relatively well-defined and simple vocabulary, the operator can input the vast majority of the mission parameters using simple, intuitive voice commands. However, voice input may be more applicable to initial mission specification rather than for critical commands such as the need to land immediately due to time and feedback constraints.

Furthermore, although the interface is simple, autonomous algorithms that function transparently to the operator allow the mission to proceed with minimal operator input. This methodology employs algorithms that generate trajectories, coordinate vehicles and avoid collisions with cooperative and uncooperative obstacles using only an online, inaccurate prediction of the trajectory of the obstacles. These algorithms provide rigorous proofs of their performance guarantees. To achieve coordination, vehicle will utilize a wireless communication network, the quality of service of which determines the guaranteed performance bounds of the aforementioned time-coordination algorithm. These guarantees may engender a higher level of trust with the mission operator. This trust will enable a higher functioning system and will facilitate teaming amongst all the autonomous agents whether they are human or machine.

Using voice input for mission specification and using either voice or screen input for commanding the mission combined with guaranteed performance bounds for coordinated flight path generation and following will enable mission operators, rather than UAV pilots, to define and oversee UAV missions. Follow-on research will measure the efficiency and acceptability of using voice or screen input for both mission specification and mission control with multiple coordinated sUAVs flying their missions initially in simulation and then in actual flight tests.

References

1. Jenkins, D., Bijan, V.: The Economic Impact of Unmanned Aircraft Systems Integration in the United States. AUVSI, Arlington (2013)
2. Thompson, C.: Unmanned Aircraft to Play Key Role in Future of Agriculture. Southeast Farm Press 2. Available: <http://southeastfarmpress.com/equipment/unmanned-aircraft-play-key-role-future-agriculture> (2013)
3. Corbett, P.: Drone photography catching on in real estate. In: The Arizona Republic. <http://finance-commerce.com/2014/01/drone-photography-catching-on-in-real-estate/>. Finance and Commerce, p. 1 (2014)
4. Johnson, T.: FAA may approve use of drones for hollywood filmmaking. Variety, p. 1 (2014)
5. Barr, A.: Amazon testing delivery by Drone. CEO Bezos Says in USA Today, p. 1. <http://www.usatoday.com/story/tech/2013/12/01/amazon-bezos-drone-delivery/3799021/> (2013)
6. Jansen, B.: FAA approves first commercial drone over land. In: USA Today, p. 1. <http://www.usatoday.com/story/money/business/2014/06/10/faa-drones-bp-oil-pipeline-aerovirement-north-shore/10264197/> (2014)
7. Wald, M.L.: Domestic drones stir imaginations, and concerns. In: The New York Times, p. 1 (2013)
8. Turnigy 9XR User Manual. Turnigy
9. 7CAP/7CHP Instruction Manual. vol. 1m23n13606, Futaba (2003)

10. Hsu, J.: MIT Researcher Develops iPhone App to Easily Control Swarms of Aerial Drones. Popular Science. Available: <http://www.popsci.com/technology/article/2010-04/video-using-smartphones-control-aerial-drones> (2010, 5 April 2010)
11. Smalley, D.: Robocopter: New Technology Brings New Capabilities to the Marine Corps. Available: <http://www.onr.navy.mil/Media-Center/Press-Releases/2014/ONR-Demonstrates-Autonomous-Helicopter-Technology.aspx> (2014, 8 February)
12. Parrot AR.Drone 2.0. Available: <http://ardrone2.parrot.com/> (2015, 8 February)
13. Prioria Hex Mini. Available: <http://www.prioria.com/hex-mini/> (2016, 8 February)
14. Eurofighter Typhoon. The Cockpit. Available: <https://www.eurofighter.com/the-aircraft> (14 February)
15. Trujillo, A.C., Cross, C.D., Fan, H.H., Hempley, L.E., Motter, M.A., Neilan, J.H. et al.: Collaborating with autonomous agents—'I'm a doctor Jim, not an engineer' (AIAA 2015-3033). In: Paper presented at the aviation 2015—15th AIAA aviation technology, integration, and operations conference, Dallas, TX (2015)
16. Chang, J.: Speech Recognition Leaps Forward. Available: <http://research.microsoft.com/en-us/news/features/speechrecognition-082911.aspx> (2011, 10 February)
17. Koetsier, J.: Microsoft Doubles Speech Recognition Speed While Improving Accuracy. Available: <http://venturebeat.com/2013/06/17/microsoft-doubles-speech-recognition-speed-while-improving-accuracy/> (2013, 10 February)
18. Whitney, L.: Google Now Tops Siri at Accuracy, Says Analyst. Available: <http://www.cnet.com/news/google-now-tops-siri-at-accuracy-says-analyst/> (2014, 10 February)
19. Ritchie, R.: Siri 'Crushing' Competitors at Language Accuracy. Available: <http://www.imore.com/siri-crushing-competitors-language-accuracy> (2015)
20. Deng, L., Huang, X.: Challenges in adopting speech recognition. *Commun. ACM* **47**, 7 (2004)
21. McMillian, Y., Gilbert, J.E.: Distributed listening: a parallel processing approach to automatic speech recognition. In: ACL-08: HLT, Columbus, OH, pp. 173–176 (2008)
22. Hart, S.G., Staveland, L.E.: Development of a NASA-TLX (Task Load Index): results of empirical and theoretical research. In Hancock, P.S., Meshkati, N. (eds.) *Human Mental Workload*, Elsevier Science Publishers B. V., Amsterdam, pp. 139–183 (1988)
23. Carnegie Mellon University. CMU Sphinx. <http://cmusphinx.sourceforge.net/Sphinx4-5prealpha> (2015)
24. National Institute of Standards and Technology (NIST). Rich Transcription Evaluation. Available: <http://nist.gov/itl/iad/mig/rt.cfm> (2015)
25. Amazon.com. Amazon Echo. Available: <http://www.amazon.com/Amazon-SK705DI-Echo/dp/B00X4WHP5E> (14 February)
26. Trujillo, A., Fan, H.H., Cross, C., Hempley, L., Navarro, J.P., Mehdi, B.S., et al.: Operator informational needs for multiple autonomous small vehicles. *Procedia Manufact.* **3**, 936–943 (2015)
27. Shanmugavel, M., Tsourdos, A., Zbikowski, R., White, B.A., Rabbath, C.A., Léchevin, N.: A solution to simultaneous arrival of multiple UAVs using pythagorean hodograph curves. In: Paper presented at the American control conference, Minneapolis, MN (2006)
28. Tsourdos, A., White, B.A., Shanmugavel, M.: *Cooperative Path Planning of Unmanned Aerial Vehicles*. American Institute of Aeronautics and Astronautics, Reston (2011)
29. Choe, R., Cichella, V., Xargay, E., Hovakimyan, N., Trujillo, A.C., Kaminer, I.: A trajectory-generation framework for time-critical cooperative missions (AIAA-2013-4582). In: Paper presented at the AIAA infotech@Aerospace conference, Boston, MA (2013)
30. Choe, R., Puig-Navarro, J., Cichella, V., Xargay, E., Hovakimyan, N.: Trajectory generation using spatial pythagorean hodograph bézier curves (AIAA-2015-0597). In: Paper presented at the AIAA SciTech 2015, Kissimmee, FL (2015)
31. Cichella, V., Kaminer, I., Xargay, E., Dobrokhodov, V., Hovakimyan, N., Aguiar, A.P. et al.: A lyapunov-based approach for time-coordinated 3D path-following of multiple quadrotors. In: 2012 IEEE 51st annual conference on decision and control (CDC), Maui, HI, pp. 1776–1781 (2012)

32. Xargay, E., Kaminer, I., Pascoal, A., Hovakimyan, N., Dobrokhodov, V., Cichella, V. et al.: Time-critical cooperative path following of multiple unmanned aerial vehicles over time-varying networks. *J. Guidance Navig. Control Dyn.* **36**, 499–516 (2013)
33. Cichella, V., Marinho, T., Stipanović, D., Hovakimyan, N., Kaminer, I., Trujillo, A.C.: Collision avoidance based on line-of-sight angle. In: Paper presented at the conference on decision and control (CDC), Osaka, Japan (2015)
34. Mehdi, S.B., Choe, R., Cichella, V., Hovakimyan, N.: Collision avoidance through path replanning using bézier curves (AIAA-2015-0598). In: Paper presented at the AIAA SciTech 2015, Kissimmee, FL (2015)
35. Mehdi, S.B., Choe, R., Hovakimyan, N.: Multiple collision avoidance through trajectory replanning using piecewise bézier curves. In: Paper presented at the 54th IEEE conference on decision and control (CDC), Osaka, Japan (2015)
36. Puig-Navarro, J., Xargay, E., Choe, R., Hovakimyan, N., Kaminer, I.: Time-critical coordination of multiple UAVs with absolute temporal constraints (AIAA-2015-0595). In: Paper presented at the AIAA SciTech 2015, Kissimmee, FL (2015)
37. Xargay, E.: Time-critical cooperative path-following control of multiple unmanned aerial vehicles. Ph.D., University of Illinois at Urbana-Champaign, IL (2013)

Adaptive Interaction Criteria for Future Remotely Piloted Aircraft

Jens Alfredson

Abstract There are technical trends and operational needs within the aviation domain towards adaptive behavior. This study focus on adaptive interaction criteria for future remotely piloted aircraft. Criteria that could be used to guide and evaluate design as well as to create a model for adaptive interaction used by autonomous functions and decision support. A scenario and guidelines from the literature, used as example criteria, was presented in a questionnaire to participants from academia/researchers, end users, and aircraft development engineers. Several guidelines had a wide acceptance among the participants, but there was also aspects missing for the application of supporting adaptive interaction for remotely piloted aircraft. The various groups of participants contributed by different aspects supports the idea of having various stakeholders contributing with complementary views. Aspects that the participants found missing includes, predictability, aviation domain specifics, risk analysis, complexity and how people perceive autonomy and attribute intentions.

Keywords Adaptive · Aircraft · Air systems · Interaction · Pilot · Remotely piloted

1 Introduction

There is a long term technical trend towards more automated and autonomous systems and functions within the aviation domain. Also, future remotely piloted aircraft will most likely adapt technical trends towards increased levels of automation and autonomy. This will lead to new design challenges when developing future

J. Alfredson (✉)

Department of Computer and Information Science, Linköping University,
581 83 Linköping, Sweden
e-mail: Jens.Alfredson@liu.se

J. Alfredson

Saab Aeronautics, Bröderna Ugglas Gata, 582 47 Linköping, Sweden

© Springer International Publishing Switzerland 2017

P. Savage-Knepshild and J. Chen (eds.), *Advances in Human Factors in Robots and Unmanned Systems*, Advances in Intelligent Systems and Computing 499,
DOI 10.1007/978-3-319-41959-6_23

manned as well as unmanned aircraft, for instance concerning intelligent adaptive automation. Although design guidance for intelligent adaptive automation has long been available [1] there are major challenges and research questions remaining. For instance, complacency and automation bias occur in the use of imperfect automation [2] and automation has been found to affect pilots through subsystems for management, workload and awareness [3].

There are not only technical trends that call for design challenges. Also, there are operational needs in civil and military aviation towards adapting to dynamic chains of events that might best be met by adaptive aiding. It has been shown that psychophysiological based real-time adaptive aiding significantly aids performance in a complex remotely air vehicle task [4] and also coadaptive aiding (i.e. both system and user adapts) has been demonstrated in a remotely piloted aircraft setting [5].

In this study adaptive interaction criteria for future remotely piloted aircraft has been in focus. That is, criteria for making design decisions related to development of adaptive interaction for future remotely piloted aircraft. The reason for this interest is the idea that if criteria can be identified, structured and prioritized it could be used:

- To guide the design of adaptive interaction
- When evaluating specific design or concepts related to adaptive interaction
- To create a model for adaptive interaction that could be integrated into future unmanned aircraft, to be used by autonomous functions and decision making.

Even though the need for adaptive interaction criteria may be just as high in other domains it is not possible to wait for the corresponding work of other domains to fill the need for identifying, structuring and prioritizing adaptive interaction criteria and hoping to apply the results to the domain of unmanned aircraft since the criteria is domain specific. The reason for it being domain specific is that each domain has unique user needs and requirements and for the aviation domain also specific technical context. For the aviation domain examples could include real time decision making of highly dynamic spatially problems of for example traffic avoidance or collision avoidance. Examples of technical context could be regarding state-of-the-art communication technology for the aviation domain or on-board sensor capabilities.

In a broad sense, adaptive interaction criteria could include aspects such as usability, coherency, transparency, fidelity, validity and performance aspects including function allocation, guidance of multimodal interaction, and decision support, since that is all aspects potentially affecting design decisions that could be applied to the design process of a future remotely piloted aircraft. By applying criteria to the design process lessons learned will be gained and iteratively feedback will be available to refine the identification, structuring and prioritization of better suited criteria. As a starting point of this study a scenario was defined, including some technical assumptions. Also, a literature review was conducted aiming for initial criteria to be used in the study.

The scenario included a high workload reconnaissance mission, inspired by a study of adaptive automation for human supervision of multiple uninhabited vehicles [6]. The system includes adaptive automation to support target identification, route planning, communications and change detection (so that the pilot of the remotely piloted aircraft do not miss anything of importance on the displays), for the remotely piloted aircraft.

Also, the scenario concerned a future where remotely piloted aircraft is flying together with manned aircraft. The remotely piloted aircraft is equipped with a detect and avoid system that know where other aircraft are and have the ability to avoid them, heavily inspired by a former study [7].

As initial criteria a set of eight design principles for adaptive automation and aiding was used proposed by earlier research as design guidelines [8]:

- (1) *“Adaptive function allocation to the operator should be used intermittently. Intermittent allocation can improve performance in monitoring tasks.*
- (2) *Energetic human qualities should be considered in design. For example, degrees of challenge can be automatically adjusted with artificial tasks.*
- (3) *Emotional requirements of the human operator must be considered. The human operator should not feel unnecessary to the system as a whole.*
- (4) *The system should be calibrated to the individual operating it. Individual differences factor into the human operator portion of a human-system pairing and thus should be incorporated into the design.*
- (5) *Task transformation should be used to simplify tasks for operators. A task that is partitioned and transformed can be handled piecemeal instead of as a whole.*
- (6) *The environmental context of the system should be used to determine allocation. Environmental stressors such as heat, vibration, and gravitational force affect human performance and should be addressed.*
- (7) *Tasks should be partitioned when both the human and the system can contribute effectively. A true human-system collaboration operates as a pairing instead of a dichotomy of effort. Performance is improved when the most effective attributes of each part are employed.*
- (8) *Adaptation should be controlled by the system but be open to human intervention when the system fails to recognize new conditions or demands. In order to reduce task load on the human operator and improve general performance, the system should allocate tasks. To improve satisfaction and motivation, the human operator should retain control, or perceived control, of the system.”*

2 Method

A questionnaire was designed to facilitate the discussions with the participants of this study. The questionnaire contained three parts:

- Description of context. Including the scenario described in the introduction section of this paper.
- Questions about the importance of the above guidelines. Each criteria were to be rated on a seven point Likert-scale, where the rating of one corresponded to “not important” and seven corresponded to “very important”. There were also open questions, and comments were asked for.
- Background questions, such as age, gender, education level, and professional experience.

After a group of participants had filled in the questionnaire individually there was an opportunity for a group discussion/mini-workshop, to open up for reflections both on the answers provided, but also on the design of the questionnaire itself to gather feedback useful for potential future workshops.

There were 26 participants in this study, of which 23 were male and three female. They were between 26 and 66 years old, with a mean age of 41.

There were three groups of participants in this study, reflecting the various backgrounds of:

- Academia/research (A/R)
- End users (EU)
- Aircraft development engineers (ADE)

The participants from various backgrounds was used to extract knowledge regarding adaptive interaction criteria for future remotely piloted aircraft, through a qualitative analysis of their response.

The group from academia/research consisted of 10 persons, mainly researchers within computer and information science at various levels (2 B.Sc., 4 Ph.D. Students, 4 Ph.D.). The group of end users consisted of 3 pilots that all were very experienced as pilots, one as a fighter pilot, one as commercial pilot, and the third with experience as both fighter pilot as well as commercial pilot. They had also been former participants in a simulator study of earlier workshops performed in a European detect and avoid project [7], and were familiar with the detect and avoid part of the scenario. The group of aircraft development engineers consisted of 13 persons that were all active professionals.

3 Results

Although the mean ratings (see Fig. 1) did not reveal much differences between the participants, individual ratings were sometimes used in the discussion after filling in the questionnaire.

For instance, one of the pilots (EU3) had ranked both design guidelines #1 and #2 with the lowest rank (1), and in the discussion he explained why. For #1 his

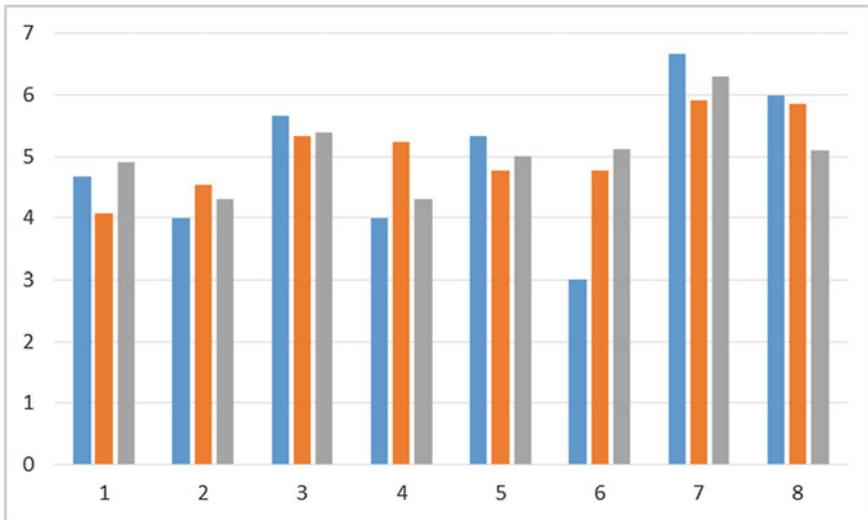


Fig. 1 Mean ratings for each of the eight design guidelines. The *blue bars (left)* represents the response from the end users (EU), the *brown (middle)* represents the aircraft development engineers (ADE), and the *grey (right)* represents the participants from academia/research (A/R)

concern was that it would lead to a less predictable behavior of the system, and for #2 it was the “artificial tasks” that he could not accept when controlling a remotely piloted aircraft. Also, he made an additional comment about #8 and said that he thought it was important that “the human operator should retain control”, but that it should not only be “perceived control”.

Also, the workshop with the aircraft development engineers revealed that they think it is important that the guidelines guide design decisions that are hard to make, for the guidelines to be useful. It was debated if #7 is really fulfilling that or if the fact that “tasks should be partitioned when both the human and the system can contribute effectively” even if it is true might not help to be specifically aware of when performing actual design. Also, remarks were made regarding:

- If it would be possible to apply a principle of automating everything except what is better suited for a remote pilot to perform.
- To let the remote pilot adjust changes of priority in the system.
- That it might take longer time to learn if you as a remote pilot do not perform detailed tasks.
- Would it add stress to the pilot to know that the system is regarding the pilot as stressed?
- Concerns that adaption would not support (or even counteract) drastic transitions between automation levels.
- That it is important that guidelines are well formulated.

Results from the open questions and comments include:

- Missing aspects:
- Complexity (A/R4)
- How people perceive autonomy and attribute intentions (A/R2)
- System transparency and predictability (ADE1, ADE4, EU3)
- Aviation domain specifics (ADE11)
- Level of automation (ADE7)
- Risk analysis (ADE2)
- Biometrics (ADE8)
- How automation affects pilot understanding and need for regular practice (ADE2)
- Comments to Guideline #5: Is it always good to simplify, or is there a risk to loose overall view(ADE3)?

Also, there was comments about ambiguity and wordings for some of the guidelines:

- #2 (EU3)
- #3 (A/R8, EU3, ADE1)
- #5 (A/R3)
- #6 (ADE7, A/R7)
- #8 (EU3)

4 Discussion

The focus of this study has been adaptive interaction criteria for future remotely piloted aircraft, with the idea that criteria can be identified, structured, and prioritized to guide and evaluate design as well as for creating a model for adaptive interaction for future unmanned aircraft, for enhanced decision making potentially aided by autonomous functions. Participants from academia/researchers, end users, and aircraft development engineers has contributed by reflecting on a sample of design guidelines that was used as an examples of possible criteria.

The results show that several of the guidelines had a wide acceptance among the participants, but that there was also aspects missing in the design guidelines for the application of supporting adaptive interaction for remotely piloted aircraft, given the scenario provided to the participants.

The feedback from the pilots, representing the end users, were mostly focused towards sharp end aspects, such as properties of a future system, for example, that the system should be transparent, predictable, and not include artificial tasks and that the pilot should be in control. The reason that the end users stressed properties

of a system might be since they have own experiences of interaction with automated systems to relate to and know its importance by experience.

The feedback from the aircraft development engineers were in a large extent directed towards the guidelines as such, and comments on missing aspects of the set of guidelines. For instance, it was pointed out that the design guidelines should be written so that it supports design decisions, and that aspects such as aviation domain specifics, risk analysis, and putting values into the system concerning learning and training are regarded. The reason that the aircraft development engineers stress these aspects more than the end users might be due to that they are used to regard these type of aspects in their daily work. It supports the idea that criteria can be used to guide design and evaluate concepts for this application.

The feedback from the participants from academia/research were reflecting that this field is not fully matured, such as comments related to complexity and how people perceive autonomy and attribute intentions. The reason that the participants from academia/research stress these matters might be since they are used to do research within this or other fields.

That the various types of participants provided different feedback supports the idea of having various stakeholders contributing with complementary views.

There was comments about ambiguity and wordings from some participants from all the various backgrounds, concerning five of the eight guidelines, and during the workshop some of the aircraft development engineers noted the importance of well formulated guidelines.

It would be interesting for future research to fuse the complementary views from various stakeholders to iteratively continue to study identification, structuring and prioritization of adaptive interaction criteria for future remotely piloted aircraft, to create a model for adaptive interaction that could be integrated into future unmanned aircraft, to be used by autonomous functions and decision making.

If the same or similar criteria that is modelled in such a system is also used to design the system:

- Guidelines could contribute to design so that unwanted situations would not be so frequent or severe, than for a less good design, and
- If unwanted situations occur, a system with adaptive interaction could cope with the situation by adaptive aiding, decision support and display considerations such as multimodal considerations. Also, lessons learned could be fed back to the design of future systems by adaptation of the criteria and design guidelines.

Such a system could include missions monitoring, operator aspects, autonomy, states of the world, and the mission [9]. Biometrics was one of the mission aspects that was brought up, as well as aviation domain specifics. Earlier research concerning operator workload has found that "...guidelines should be created that specify the required measurement technology sensitivity for optimal dynamic task allocation in specific operational settings" [10]. By assessments made on pilot,

system, and context interpretations could be conducted and fed into a situation assessment module. The situation is assessed by comparing the information originated from pilot, system and context with reference situations. When the situation is categorized and the related to a reference situation this information is fed to an adaptive module. The adaptive module assesses the current display options and determines how the information about the situation should be displayed, based on a set of pre-defined plans and updated information from the flight management system. This includes multi modal considerations, and function allocation.

5 Conclusions and Future Work

This study has studied adaptive interaction criteria for future remotely piloted aircraft. Also, the applicability of the criteria to development of future adaptive systems has been studied. The adaptive interaction criteria may be used to aid the design of future piloted aircraft as well as for evaluation of concepts and design. Also, the criteria may contribute to a model for adaptive interaction that could be integrated into future unmanned aircraft, to be used by autonomous functions and decision making.

This study has found that participants from the groups of end users, academia/researchers, and aircraft development engineers widely accepted a set of guidelines provided by earlier research [8]. However, aspects that the participants though was missing includes, system transparency and predictability, aviation domain specifics, risk analysis, complexity, and how people perceive autonomy and attribute intentions. Also, design guidelines should be well formulated and written so that it supports design decisions, and putting values into the system concerning learning and training.

The various group of participants stressed different aspects. End users mostly focused sharp end aspects, such as properties of a future system, the aircraft development engineers mainly reflecting on the guidelines and how they can support design, and the participants from academia/research reflected on that this field is not fully matured. That the various types of participants provided different feedback supports the idea of having various stakeholders contributing with complementary views.

Future work, could include fusion of the complementary views from various stakeholders to iteratively continue to study identification, structuring and prioritization of adaptive interaction criteria for future remotely piloted aircraft, to create a model for adaptive interaction.

Acknowledgments This project is financially supported by the Swedish Foundation for Strategic Research.

References

1. Banbury, S., Gauthier, M., Scipione, A., Hou, M.: Intelligent Adaptive Systems—Literature-Research of Design Guidance for Intelligent Adaptive Automation and Interfaces. CR 2007-075, Defence Research and Development, Canada (2007)
2. Wickens, C.D., Clegg, B.A., Vieane, A.Z., Sebok, A.L.: Complacency and automation bias in the use of imperfect automation. *Hum. Factors* **57**, 397–406 (2015)
3. Durso, F.T., Stearman, E.J., Morrow, D.G., Mosier, K.L., Fischer, U., Pop, V.L., Feigh, K.M.: Exploring relationships of human-automation interaction consequences on pilots—uncovering subsystems. *Hum. Factors* **57**, 965–975 (2015)
4. Wilson, G.F., Russel, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiologicaly determined adaptive aiding. *Hum. Factors* **49**, 1005–1018 (2007)
5. Christensen, J.C., Estep, J.R.: Coadaptive aiding and automation enhance operator performance. *Hum. Factors* **55**, 965–975 (2013)
6. Parasuraman, R., Cosenzo, K.A., De Visser, E.: Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. *Mil. Psychol.* **21**, 270–297 (2009)
7. Alfredson, J., Hagström, P., Sundqvist, B.-G.: Situation awareness for mid-air detect-and-avoid system for remotely piloted aircraft. In: AHFE 2015, Las Vegas, pp. 1014–1021 (2015)
8. Steinhauser, N.B., Pavlas, D., Hancock, P.A.: Design principles for adaptive automation and aiding. *Ergonomics Des.* **17**, 6–10 (2009)
9. Gutzwiller, R.S., Lange, D.S., Reeder, J., Morris, R.L., Rodas, O.: Human-computer collaboration in adaptive supervisory control and function allocation of autonomous system teams. In: HCI 2015, Los Angeles, pp. 447–456 (2015)
10. Johnson, A.W., Oman, C.M., Sheridan, T.B., Duda, K.R.: Dynamic task allocation in operational systems: issues, gaps, and recommendations. In: Aerospace conference 2014, Big Sky, Montana, pp. 1–15 (2014)

Confidence-Based State Estimation: A Novel Tool for Test and Evaluation of Human-Systems

**Amar R. Marathe, Jonathan R. McDaniel, Stephen M. Gordon
and Kaleb McDowell**

Abstract Test and evaluation (T&E) of complex human-in-the loop systems has been a challenge for system developers. Traditional methods for T&E rely on questionnaires given periodically in combination with task performance measures to quantify the effectiveness of a given system. This approach is inherently obtrusive and interferes with natural system interaction. Here, we propose a method to leverage unobtrusive wearable technology to create a system for continuously assessing human state. Previous efforts at this type of assessment have often failed to generalize beyond controlled laboratory environments due to increased variability in signal quality from both the wearable sensors and in human behavior. We propose a method to account for this variability using measures of confidence to create robust estimates of state capable of dynamically adapting to changes in behavior over time. We postulate that the confidence-based approach can provide high-resolution estimates of state that will augment T&E of complex systems.

Keywords Test and evaluation · Human assessment · Confidence · Sensor fusion · State estimation

A.R. Marathe (✉) · J.R. McDaniel · S.M. Gordon · K. McDowell
Human Research and Engineering Directorate, U.S. Army Research Laboratory,
Aberdeen Proving Ground, Adelphi, MD 21005, USA
e-mail: amar.marathe.civ@mail.mil

J.R. McDaniel
e-mail: jmcdaniel@dcscorp.com

S.M. Gordon
e-mail: sgordon@dcscorp.com

K. McDowell
e-mail: kaleb.g.mcdowell.civ@mail.mil

J.R. McDaniel · S.M. Gordon
DCS Corporation, Alexandria, VA, USA

1 Introduction

Test and evaluation (T&E) of complex human-in-the-loop systems has been a challenge for system developers in recent years. In general, traditional methods for T&E rely on the use of questionnaires given periodically in combination with overall task performance measures to quantify the effectiveness and usability of a given system. However, human assessment through surveys and questionnaires is inherently obtrusive and interferes with natural system interaction and task performance. Furthermore, such an approach does little to address the fact that complex, human-in-the-loop systems may fail for a number of reasons including poor system design, system malfunction, or operator error.

Recently, the National Research Council recommended that the T&E community adopt less obtrusive human assessment techniques, such as those provided by wearable technology [1]. These technologies would both mitigate the negative impact on natural system interaction and provide access to less subjective measures of operator state, which may provide a more accurate perspective into specific failure instances. Unfortunately, interpreting data from wearable sensors into valid measures of state remains a challenge. There has been prior work in estimating human state using behavioral and physiological signals [2–13]; however, most of these approaches fail to generalize beyond the controlled environment of typical laboratory-based experiments. When laboratory-based constraints are relaxed, signal variability often increases leading to changes in the relationship between the recorded signal and human state.

Borrowing ideas from the control systems literature, we argue the use of confidence measures can improve overall state estimation for human state assessment systems. We also propose that confidence values are applicable when not only interpreting the raw, measured signals, but also when evaluating the individual components of the system, one of which is the human. Such a measurement system would enable robust measures of state that dynamically adapt over time as signal quality, human behavior, task or environment varies as well as make inferences about the state, such as whether or not the human is even capable of performing well on the given task. We postulate that the confidence-based approach would provide a powerful tool to enhance T&E of human-in-the-loop systems by providing accurate, high resolution estimates of state, which, when combined with overall task performance measures and questionnaire data, will provide a wealth of information; enabling system designers to better identify sources of errors through minimally intrusive techniques.

The remainder of this paper is organized as follows. Sect. 2 provides an overview of the proposed T&E system. Sect. 3 provides an overview of research needed to enable the development of such a system. In particular, this section includes an in-depth discussion regarding the construct of confidence as we propose to apply it to human assessment, and presents proof-of-concept simulations that demonstrate how the use of confidence can improve reliability of state estimates. Finally, Sect. 4 concludes with a discussion about how the confidence-based approach will augment T&E.

2 Towards a Tool for Test and Evaluation

Standard methods for T&E often rely on the use of questionnaires given periodically in combination with overall task performance measures to quantify the effectiveness and usability of a given system. These tools provide a low-resolution estimate of human-system performance. We argue that utilizing modern human assessment techniques in place of periodic questionnaires would allow both the measurement of human state in real-time as well as confidence-based inferences about that state. For instance, we know that systems fail for a number of reasons such as poor system design, system malfunction, or operator error. However, precisely identifying the source of such failures is a challenge. We posit that a system that robustly tracks human operator state (as shown in Fig. 1) will generate high-resolution information regarding the human-system interaction that can help evaluators disambiguate the source of operational failures and precisely identify the time and context of such failures during the T&E process.

The proposed system would be used when an operator is brought into work with a new technology during the T&E process. Traditionally, in this type of scenario, the operator would start with a battery of pre-assessment questionnaires, followed by training on the new system evaluated that day. After training, the operator would fill out another battery of questionnaires and then carry out the mission designed for the current evaluation. During the course of the mission, the operator would be interrupted periodically to provide feedback to the T&E team about specific issues with the new technology. Upon completion of the mission, the operator would then complete a final battery of questionnaires, and provide any additional feedback to

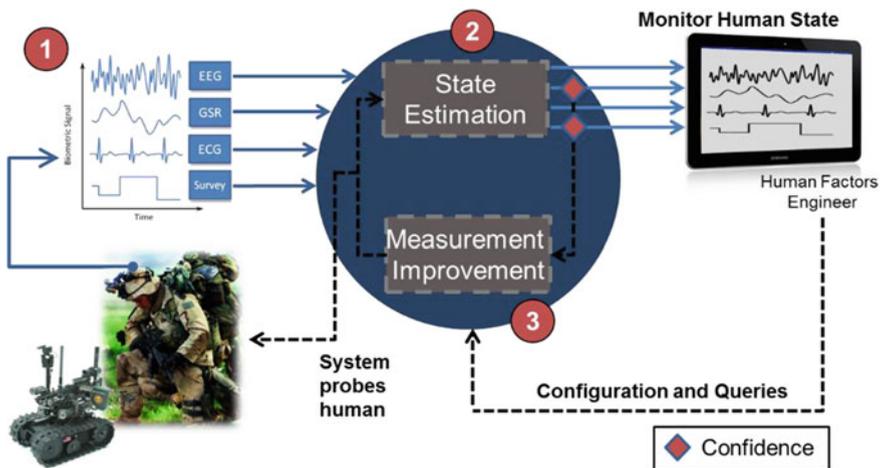


Fig. 1 Proposed T&E System for robustly tracking human operator state to generate high-resolution information regarding the human-system interaction. Critical research areas are highlighted & numbered within red circles

the T&E team that may be pertinent to system operation. The questionnaires, coupled with mission performance metrics would then be used to evaluate the effectiveness of the new technology.

With the proposed system, the T&E session would proceed slightly differently. The operator would start with a series of pre-assessment questionnaires, followed by training on the new technology to be evaluated that day. After completing training and any additional questionnaires, the operator would then be outfitted with an array of behavioral and physiological sensors. These body-worn sensors would be coupled with a suite of environmental and task based sensors (Step 1 in Fig. 1). The operator would carry out the mission designed for the current evaluation. During the course of the mission, data from the entire sensor suite (both body worn and environmental) would be used to estimate changes in the operator (Step 2 in Fig. 1). Information about specific states such as fatigue, attention, workload, stress would be continuously estimated on the basis of the recorded data. Periodically, when the state estimation algorithms were unable to confidently estimate a particular state on the basis of the sensor suite, the operator would be momentarily interrupted to provide feedback to the state estimation system (Step 3 in Fig. 1). This feedback would provide the state estimation algorithms information needed to quickly improve their estimates. By periodically infusing external information about the human's state, the system would be better able to dynamically track state over long periods of time. Additionally, the operator is rarely interrupted to explicitly provide feedback to the T&E team during mission execution, thereby enabling more natural interaction with the technology. Upon completion of the mission, the operator would then complete another set of questionnaires, and provide any additional feedback to the T&E team that may be pertinent to the system operation. The questionnaires, coupled with mission performance metrics and the continuous state estimates derived from the recorded data would then be used to evaluate the effectiveness of the new technology.

We argue that the addition of these state estimates to the data record would provide a wealth of new information regarding the onset and offset of a variety of changes in the operator that may have influenced mission performance. This new information can then be used to better identify design elements or features that also may have significantly impacted performance (positively or negatively). This information can help evaluators disambiguate the source of operational failures and precisely identify the time and context of such failures and, thus, would have a tremendous effect on T&E of future systems. To realize this type of system, three critical research areas must be addressed:

1. State estimation and sensor fusion methods that combine information from multiple sensor modalities to estimate human state.
2. Confidence measures to maintain robustness.
3. Measurement improvement techniques.

These areas are covered in more detail in the following section.

3 Enabling the Future of Test and Evaluation

3.1 *State Estimation and Sensor Fusion*

The first element in our proposed system is estimating state from multi-modal sensors. Recently, the state assessment community has seen unparalleled advancements in sensor and analysis technologies that provide new insights into different facets of human behavior and performance. Methods exist to extract information related to fatigue [2–4], stress [5], workload [6–8], vigilance [9–11], and even target identification [12, 13] from individuals based on data from physiological sensors.

Concurrently, there have been substantial improvements in sensor fusion techniques across a variety of domains. A detailed review of this literature can be found in [14]. In general, the fusion of multiple heterogeneous information sources is divided into two main categories: (i) building a joint model by integrating multiple approaches and (ii) fusing the output of multiple approaches. While modeling the dependencies across multiple approaches that have been built using different design principles is effective [15, 16], it is often infeasible. When modeling dependencies across multiple approaches is not possible, fusion can be performed over the outputs of each approach instead. Approaches have ranged from weighted-sum techniques [17], Bayesian fusion [18], and Dempster-Shafer based approaches [19–21].

3.2 *Confidence*

Once state estimates are generated, confidence measures are needed to quantify the reliability of those state estimates in near real-time. The importance of confidence in systems with low signal-to-noise properties has long been understood in decision theory [22–24] and control communities [25, 26], and thus we believe it is also a critical component for robust human state assessment systems. While it is well understood that human psychological and physiological states vary widely, both across and within individuals [27, 28], it is less well understood how to predict the expected variability given an observed state in a specific operator working within a particular task environment. This is due in large part to an incomplete understanding of how operator states change over time and across subjects.

These challenges are further exacerbated by variability in current human sensing techniques that produce estimates of operator state that are often either inconsistent, invalid, or both when applied in real-world circumstances [29]. As such, the myriad available measures based on human-sensed data, including behavioral and psychophysiological signals, are not yet widely used [27, 30]. Thus, confidence-based approaches should be used when assessing signal quality and integrating, or fusing, multiple modalities.

In addition to this traditional use of confidence, however, we also argue for the purposes of T&E that confidence should be used to reason about the human's state and the potential effects of that state within the context of overall system performance. For instance, human feedback is often treated as having little or no noise or it is constrained to a level that is presumed to be effectively without noise [31–34]. In the context of T&E, this may apply to subjective questionnaires or the human's self-evaluation of the overall system's performance. Therefore, in our proposed framework, we consider confidence on multiple levels:

1. As traditionally used in control systems, confidence can be used to measure the reliability or validity of signals recorded from specific sensors.
2. As we integrate sensor information to estimate state, we can infer confidence in the state estimate based on a multitude of factors including: a combination of the underlying signal reliability, information from sensor modalities not directly associated with the state estimate, and context.
3. As we increase our ability to measure human state, we can infer a confidence value in all the system components (including the human) in order to either best diagnose the causes for specific system successes or failures (e.g. T&E), or dynamically leverage information from each agent over time (Human-Autonomy Interaction applications).

Examples of the first two approaches to confidence can be found in recent advances from the field of brain-computer interaction (BCI) technologies. Efforts in that field have illustrated that state-of-the-art human sensing technologies can be leveraged and that real-time estimates of measurement confidence can support improved integration of human-sensed data [35, 36]. Examples of the third approach to confidence can be found in our own previous work, which pursues a unique extension to the established methods [37, 38]. Much of this work focuses on establishing confidence methods on the basis of neural data on a short time scale; however future extensions to these approaches will consider contributions of non-neural and long time scale factors.

3.3 Confidence Measures from Short Term Neural Data

Figure 2 demonstrates the utility of the confidence-based approach using data from one participant performing a visual target identification task in collaboration with a computer vision autonomy [37] (details of the human experiment can be found in [39]). This exemplar data highlights three important points. First, the confidence-based approach is effective at fusing human-sensed information. In Fig. 2a, the target-detection performance of the confidence-based fusion approach outperforms each of the individual target detectors, as well as the non-confidence-based approach. Second, the confidence-based approach also works well at integrating human and non-human sensed data. In Fig. 2b, the target detection performance of the confidence-based fusion approach outperforms both the human and computer

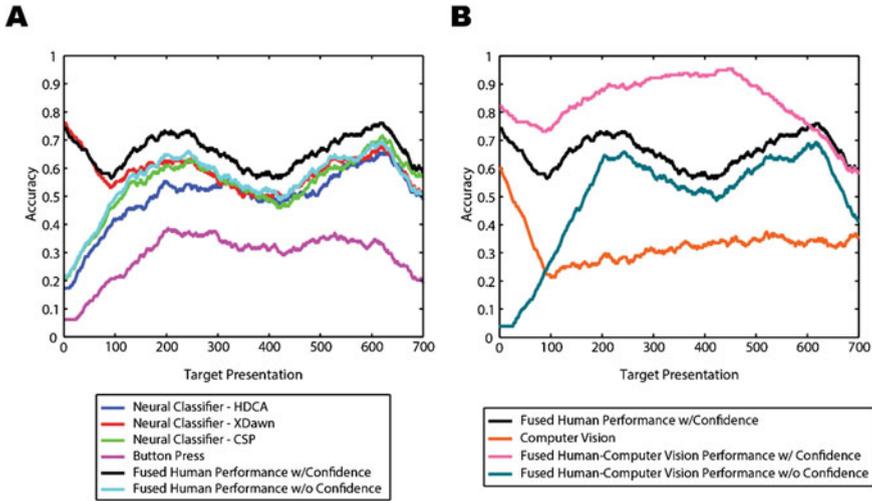


Fig. 2 Comparison of target detection performance over time using confidence and non-confidence based fusion approaches. **a** illustrates the combination of multiple human-sensed measures to detect target stimuli. **b** shows the combination of human target detection with computer vision-based target detection

vision target detection as well as the non-confidence-based fusion method. This result also demonstrates that while confidence can improve the accuracy of the estimates, the accuracy of the underlying data streams will ultimately dictate the accuracy of the fused data. Again, in Fig. 2b, as the performance of both the human and computer vision target detection decrease, the performance of the confidence based fusion approach also decreases.

We believe the example results presented in Fig. 2 highlight the potential utility of a confidence-based approach; however when operating in an open-loop fashion, this initial confidence-based approach cannot provide the needed robustness over time if the underlying signals, themselves, are not robust over time. To address this latter point, there are two methods to improve overall performance. First, the confidence measures themselves can be improved by employing better estimation techniques as well as leveraging data from a greater variety of sources. Second, rather than using the confidence-based approach in an open-loop fashion, the system should be closed loop and seek additional information when confidence decreases (as shown in Fig. 1). Methods for improving confidence measures are discussed in the following section, and following that, we discuss ideas for closing the loop.

3.4 Confidence Measures from Non-Neural and Long Term Factors

The previously discussed study (Fig. 2) employed confidence measures that used only information from neural data on a short (1 s) time scale; however there is a large amount of information from other sources that could potentially improve the estimate of confidence. Ideally, when assessing human state, one should not only measure central nervous activity but also peripheral and autonomic system functions [40]. This includes behavioral or physiological information streams such as eye tracking, heart rate, skin conductance, or respiration rate. For example, if a system were estimating an individual's attention level using electroencephalography (EEG), confidence in the EEG based estimate could be derived from concurrently recorded eye tracking data. In this case, if the eye tracking indicated a gaze at the appropriate location, confidence in the EEG based assessment would be high (unless contraindicated by other factors), while if eye tracking behavior indicated that the individual was not focused on the appropriate location, confidence in EEG based attention measure would be low. Similar approaches may be applied to a wide range of human sensed data. Additionally, long-term factors might further augment estimates of confidence. For example, chronic sleep deprivation often results in a greater probability of fatigue, stress, frustration and a variety of other states. In this case, knowledge of an individual's sleep history could be used as an integral part of deriving confidence in human state estimates.

An accurate estimate of confidence in particular human state estimates will likely require a combination of a number of the approaches above. Ongoing and future studies should examine specific methods for combining different approaches to improve our confidence estimates. When available, improved confidence estimates will enable detection of performance degradation, and enable systems to trigger mechanisms to improve state estimates.

3.5 Measurement Improvement

The final piece of the confidence-based tool for T&E is a measurement improvement component. This component is necessary to enable the system to find ways to improve classification performance when confidence in state estimates begins to decrease. The most straightforward approach to improve the confidence in the measurement is to get more information from the user in order to update the current state estimation techniques. This approach has been successfully leveraged through a technique known as Active Learning [41] which uses an iterative process to query human users to provide minimal amounts of information to maximally improve classifier performance. Active Learning has been successfully applied to several domains including: speech recognition and annotation [42]; image and video classification [43]; document categorization and text classification [44]; and EEG

classification [45–47]. A similar technique, where the system probes the user for the most important information, can be used within the proposed T&E application. The challenge inherent in this approach is to extract enough information from the operator to improve classification of state, while minimizing the frequency and duration of interruption to natural task interaction.

Another similar approach for measurement improvement would be to incorporate more information as a means to estimate state. In this case, rather than querying the user for additional information, the system would actively seek other sources of information to improve the reliability of state estimates. For example, when using such minimally intrusive techniques as EEG or heart rate, large movements from increased activity may contaminate the signals and introduce artifacts into the data. This would lead to low confidence in EEG- or HR-based estimates while movement persists. However, the source and frequency of the movement artifacts may provide information relating to subject stress that can be used as a surrogate while the primary measures (EEG and HR) are considered invalid. While the movement artifacts persist, the state assessment system would continue to produce stress estimates by ignoring the contaminated signals and focusing on the new information. Thus, by appropriately leveraging and dynamically integrating multiple signal modalities, the system can provide a more robust measure over time.

Another example of incorporating additional information within team-based tasks, would be to adaptively switch between multiple human agents to account for variability in human performance. Figure 3 shows examples of how such a system might work. These data represent target detection accuracy from several subjects as they moved through a virtual environment [48, 49]. In this experiment, participants would periodically enter an environment with dense fog that would impede target detection performance. Across all subjects, target detection accuracy in the presence of fog would decrease by an average of 23.26 % (Std Dev 5.7 %) and reaction time

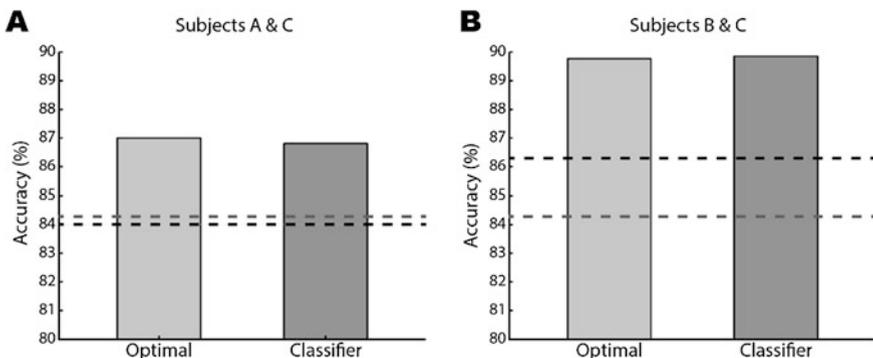


Fig. 3 Demonstration of the effectiveness of a measurement improvement system that incorporates environmental information to dynamically integrate information from multiple sources. **a** and **b** each show performance of a target detection task that leverages information from two subjects across normal and degraded visual environments. *Dotted gray line* represents Subject C in both plots. *Dotted black line* represents Subject A and B in **a** and **b** respectively

would slow by 225 ms (Std Dev 41.12 ms). As expected, some participants performed better at target detection in the normal visual environment, but struggled in the fog. Other participants performed average in the normal visual environment, but excelled relative to their peers in the fog environment. Using this data, we can employ measures of confidence in each individual's ability to perform the task as a means to dynamically combine target detection performance across multiple individuals to leverage their unique strengths. This dynamic combination of multiple individuals demonstrates how the confidence-based approach coupled with a measurement improvement system can improve overall target detection performance. Figure 3 illustrates two examples of the joint performance of two participants using an optimal fusion method (prior knowledge of when the fog turns on/off) and classifier based fusion method (classifying the presence of the fog based on changes in the recorded EEG signals from the participants). For both pairs of participants, fusing the information based on the presence of the fog improves target detection performance over either individual (Dotted gray line represents performance of Subject C in both a and b. Dotted black line represents performance of Subject A and Subject B in a and b respectively). In this situation, the measurement improvement system enables the system to achieve higher performance across changes in environmental conditions.

4 Conclusion

This paper proposes a framework for a novel tool to augment T&E of complex human-in the loop systems. The framework couples measures of confidence to human-sensed data with an array of closed-loop approaches for dynamically improving classifier reliability to create robust measures of state capable of adapting to changes in the signals over time. As such, this framework can provide a powerful tool to enhance T&E of human-in-the-loop systems by providing accurate, high resolution estimates of state, which, when combined with overall task performance measures and questionnaire data, will provide a wealth of information enabling system designers to better identify sources of errors through minimally intrusive techniques.

The potential viability of the individual components of the confidence-based system for T&E are demonstrated through exemplar data collected from various studies. While these data provide evidence that such a system should be possible, future work will focus on experimentally validating these claims and developing a proof-of-concept system capable of robustly estimating human state, and thus creating a novel tool for T&E of human-systems.

Acknowledgments This project was supported by the Office of the Secretary of Defense ARPI program MIPR DWAM31168 and by US Army Research Laboratory's Cognition and Neuroergonomics/Collaborative Technology Alliance #W911NF-10-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Council, N.R.: Human-system integration in the system development process: a new look. The National Academies Press, Washington, DC (2007)
2. Stikic, M., Johnson, R.R., Levendowski, D.J., Popovic, D.P., Olmstead, R.E., Berka, C.: EEG-derived estimators of present and future cognitive performance. *Front Hum. Neurosci.* **5** (2011)
3. Shen, K.-Q., Li, X.-P., Ong, C.-J., Shao, S.-Y., Wilder-Smith, E.P.V.: EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clin. Neurophysiol.* **119**, 1524–1533 (2008)
4. Lin, C.-T., Wu, R.-C., Jung, T.-P., Liang, S.-F., Huang, T.-Y.: Estimating driving performance based on EEG spectrum analysis. *EURASIP J. Appl. Signal Process.* 3165–3174 (2005)
5. Hosseini, S.A., Khalilzadeh, M.A., Changiz, S.: Emotional stress recognition system for affective computing based on bio-signals. *J. Biol. Syst.* **18**, 101–114 (2010)
6. Hope, R.M., Wang, Z., Wang, Z., Ji, Q., Gray, W.D.: Workload classification across subjects using EEG. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **55**, 202–206 (2011)
7. Wilson, G.F.: An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *Int. J. Aviat. Psychol.* **12**, 3–18 (2002)
8. Kothe, C.A., Makeig, S.: Estimation of task workload from EEG data: new and current tools and perspectives. In: Presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society 30 September 2011
9. Duta, M., Alford, C., Wilson, S., Tarassenko, L.: Neural network analysis of the mastoid EEG for the assessment of vigilance. *Int. J. Hum-Comput. Interact.* **17**, 171–195 (2004)
10. Hord, D.J.: An EEG predictor of performance decrement in a vigilance task (1982)
11. St John, M., Risser, M.R., Kobus, D.A.: Toward a usable closed-loop attention management system: predicting vigilance from minimal contact head, eye, and EEG measures. *Found. Augment Cogn* 12–18 (2006)
12. Gerson, A.D., Parra, L.C., Sajda, P.: Cortically coupled computer vision for rapid image search. *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**, 174–179 (2006)
13. Marathe, A.R., Ries, A.J., McDowell, K.: Sliding HDCA: single-trial EEG classification to overcome and quantify temporal variability. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22**, 201–211 (2014)
14. Khaleghi, B., Khamis, A., Karray, F.O., Razavi, S.N.: Multisensor data fusion: a review of the state-of-the-art. *Inf. Fusion.* **14**, 28–44 (2013)
15. Kwon, J., Lee, K.M.: Tracking by sampling trackers. In: IEEE International Conference on Computer Vision (ICCV 2011), IEEE, 1195–1202 (2011)
16. Wu, S., Bondugula, S., Luisier, F., Zhuang, X., Natarajan, P.: Zero-shot event detection using multi-modal fusion of weakly supervised concepts. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2665–2672 (2014)
17. Kim, T., Lee, H., Lee, K.: Optical flow via locally adaptive fusion of complementary data costs. In: Proceedings of the IEEE International Conference on Computer Vision, 3344–3351 (2013)

18. Liu, C., Yuen, J., Torralba, A.: Sift flow: dense correspondence across scenes and its applications. *Pattern Anal. Mach. Intell. IEEE. Trans.* **33**, 978–994 (2011)
19. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* 325–339 (1967)
20. Shafer, G., others: *A mathematical theory of evidence*. Princeton University Press Princeton (1976)
21. Lee, H., Kwon, H., Robinson, R.M., Nothwang, W. d, Marathe, A.M.: Dynamic belief fusion for object detection. *ArXiv Prepr. ArXiv151103183*. (2015)
22. Pascal, B., Krailsheimer, A.J.: *Pensees*: Translated with an introduction by AJ Krailsheimer. Penguin (1968)
23. Bernoulli, D.: Exposition of a new theory on the measurement of risk. *Econom. J. Econom. Soc.* 23–36 (1954)
24. Lehmann, E.L.: *Some principles of the theory of testing hypotheses*. Springer (2012)
25. Olson, E., Strom, J., Goeddel, R., Morton, R., Ranganathan, P., Richardson, A.: Exploration and mapping with autonomous robot teams. *Commun ACM* **56**, 62–70 (2013)
26. Tsiligkaridis, T., Sadler, B., Hero, A.: Collaborative 20 questions for target localization. *IEEE Trans. Inf. Theory.* **60**, 2233–2252 (2014)
27. Christensen, J.C., Estep, J.R., Wilson, G.F., Russell, C.A.: The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage* **59**, 57–63 (2012)
28. Ratcliff, R., Philiastides, M.G., Sajda, P.: Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc. Natl. Acad. Sci.* **106**, 6539–6544 (2009)
29. McDowell, K., Lin, C.-T., Oie, K.S., Jung, T.-P., Gordon, S., Whitaker, K.W., Li, S.-Y., Lu, S.-W., Hairston, W.D.: Real-world neuroimaging technologies. *IEEE Access.* **1**, 131–149 (2013)
30. Parasuraman, R., Wickens, C.D.: Humans: still vital after all these years of automation. *Hum. Factors J. Hum. Factors Ergon. Soc.* **50**, 511–520 (2008)
31. Fong, T., Thorpe, C., Baur, C.: Multi-robot remote driving with collaborative control. *IEEE Trans. Ind. Electron.* **50**, 699–704 (2003)
32. Fong, T., Thorpe, C., Baur, C.: Robot, asker of questions. *Robot. Auton. Syst.* **42**, 235–243 (2003)
33. Hayati, S., Venkataraman, S.: Design and implementation of a robot control system with traded and shared control capability. In: *IEEE International Conference on Robotics and Automation*, IEEE 1310–1315 (1989)
34. Sellner, B., Simmons, R., Singh, S.: User modelling for principled sliding autonomy in human-robot teams. In: *Multi-Robot Systems. From Swarms to Intelligent Automata Vol. III*, pp. 197–208. Springer (2005)
35. Sajda, P., Pohlmeier, E., Wang, J., Parra, L.C., Christoforou, C., Dmochowski, J., Hanna, B., Bahlmann, C., Singh, M.K., Chang, S.-F.: In a Blink of an eye and a switch of a transistor: cortically coupled computer vision. *Proc. IEEE.* **98**, 462–478 (2010)
36. Huang, Y., Erdogmus, D., Mathan, S., Pavel, M.: A Fusion approach for image triage using single trial erp detection. In: *3rd International IEEE/EMBS Conference on Neural Engineering CNE '07*, 473–476 (2007)
37. Marathe, A.R., Lance, B.J., Nothwang, W., Metcalfe, J.S., McDowell, K.: Confidence metrics improve human-autonomy integration. In: *Presented at the Human Robot Interaction, Bielefeld, Germany 3 March 2014*
38. Marathe, A.R., Ries, A.J., Lawhern, V.J., Lance, B.J., Touryan, J., McDowell, K., Cecotti, H.: The effect of target and non-target similarity on neural classification performance: a boost from confidence. *Front. Neurosci.* **9**, 270 (2015)
39. Touryan, J., Apker, G., Kerick, S., Lance, B., Ries, A.J., McDowell, K.: Translation of EEG-based performance prediction models to rapid serial visual presentation tasks. In: *Foundations of Augmented Cognition*. 521–530. Springer (2013)

40. Oie, K.S., Gordon, S.M., McDowell, K.: The multi-aspect measurement approach: rationale, technologies, tools, and challenges for systems design. In: Martin, J., Lockett, J.I., Allender, L. E., Savage-Knepshield, P. (eds.) *Designing soldier systems: current issues in human factors*. Ashgate, Burlington, VT (2013)
41. Settles, B.: Active learning literature survey. *Univ. Wis. Madison*. **52**, 11 (2010)
42. Zhu, X.: Semi-supervised learning literature survey (2005)
43. Joshi, A.J., Porikli, F., Papanikolopoulos, N.P.: Scalable active learning for multiclass image classification. *Pattern Anal. Mach. Intell. IEEE. Trans.* **34**, 2259–2273 (2012)
44. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. *J. Mach. Learn Res.* **2**, 45–66 (2002)
45. Marathe, A., Lawhern, V., Wu, D., Slayback, D., Lance, B.: Improved neural signal classification in a rapid serial visual presentation task using active learning. *IEEE Trans. Neural Syst. Rehabil. Eng.* 1–1 (2015)
46. Wu, D., Lance, B.J., Parsons, T.D.: Collaborative filtering for brain-computer interaction using transfer learning and active class selection. *PLoS ONE*. **8**, e56624 (2013)
47. Wu, D., Lance, B., Lawhern, V.: Transfer learning and active transfer learning for reducing calibration data in single-trial classification of visually-evoked potentials. In: *IEEE International Conference on Systems, Man and Cybernetics (SMC 2014)*, IEEE, 2801–2807 (2014)
48. Gordon, S.M., McDaniel, J.R., Metcalfe, J.S., Passaro, A.D.: Using behavioral information to contextualize BCI performance. In: *Foundations of Augmented Cognition*. 211–220, Springer (2015)
49. Metcalfe, J.S., Gordon, S.M., Passaro, A.D., Kellihan, B., Oie, K.S.: Towards a translational method for studying the influence of motivational and affective variables on performance during human-computer interactions. In: *Foundations of Augmented Cognition*. 63–72, Springer (2015)

Human Robots Interactions: Mechanical Safety Data for Physical Contacts

Alberto Fonseca and Claudia Pires

Abstract In a world that relies heavily on technology, the industry invests heavily to developing solutions that focus on the positive interaction between people and machines and the isolation of physical or immaterial infrastructure as a method of protection. An approach based on strategy must be followed to change the paradigm of human machine interaction, researchers must look to the machine as a co-worker and as such it may not pose a risk to other colleagues. So the challenge for designers and machine developers must therefore turn to “minimization of force” as a key to reach the safety of the machine/equipment. Considering these questions and having into account that existing data about this issue is scattered, focused in specific applications and cannot easily be transferred to different or more complex applications the International Technical Committee ISO/TC 199—Safety of machinery decided to create a Study Group—ISO/TC 199 SG01—with a purpose of prepare an International Standard that would support the design, development and use of machines that will interact with people.

Keywords Safety of machinery · Human-machine interactions · Safe contact · Safety data

1 Introduction

Market and industrial needs are forcing technological developments and leading to a new working paradigm in which humans and machines will be sharing the same working place at the same time collaborating on the same tasks.

A. Fonseca (✉) · C. Pires
Centro de Apoio Tecnológico à Indústria Metalomecânica, Rua dos Plátanos n 197,
4100-414 Porto, Portugal
e-mail: afonseca@catim.pt

C. Pires
e-mail: claudia.pires@catim.pt

This may sound not very new but if we add that for doing those tasks will be not possible to have physical or immaterial barriers, between people and the machine in order to prevent human injuries due to the contact and ensure the adequate level of safety, we will realise that in a near future humans and machines have to find out how to work side by side in a friendly and safely way.

This raises a need to develop new machine design concepts based on *active safety* and define a set of vital data and test methods to help machine designers and manufacturers on achieving the adequate safety level for the new human-machine collaborative working situation.

Although there is some information available it is very scattered and not always supported by scientific studies to see it as reliable and applicable to the safety objectives needed.

Based on this finding it appears important to work in the compilation of information that allows a set of a statistically or experimentally proven data in view of the development of harmonized solutions to prevent the risks associated with human/machine contact.

Shown that there is a remarkable interest in the human/machine contact issue from the market and industrial sides the International Technical Committee for Standardization ISO/TC 199—Safety of Machinery is about of setting up a new Working Group in charge of elaboration a new standardization document to establish mechanical safety data for physical contacts.

The Study Group has defined the following work phases: identify information on the subject in other fields; definition of parameters to be considered when designing a machine (speed, power, contact dimensions, contact location, geometry...); identify possible solutions centered on human-robot interaction and considers broader aspects, such as human behavior a towards collaboration with robotic systems.

The Group has already proposed the approach to define Safe Contact and introduced these two essential aspects: risk level and machine complexity.

Now, the project transcended of possible contribution to an effective creation of norm. Perspective to future meetings to will be identify the most relevant aspects to be included in this framework as the definition of contact forces against a set of different parameters.

2 Literature Review

The first step in the development of a normative framework is the investigation of other references that could serve as a basis. In this situation there was the existence of the following references:

- ISO/WD. 8—Safety of Machinery [1]
- ISO 10218-1:2011 [2]
- ISO 10218-2:2011 [3]

- ISO 12100: 2010—Safety of machinery [4]
- BS EN 16005: 2012—Power operated pedestrian doorsets [5]
- ISO 14120: 2002—Safety of machinery. Guards [6]
- ISO/TS 15066:2016—Robots and robotic devices—Collaborative robots [7]
- EN 71-1:2014—Safety of toys [8]
- EN 71-8: 2011 Safety toys [9]
- CEN/TC 122—prCEN/TR 20442 [10]

The parameters to be considered in the development of the standard are identified through an analysis process. This process includes the examination of organizational management factors, facility layout, job design, tools, equipment machinery, information transfer and personal factors. Besides these aspects it is still considered the physical impact of collaborative robots in humans and their effects. Desmoulin and Anderson [11] conducted a study to know the effects of mechanical action of bruises on living beings. Thus, it was possible to determine the parameters involved and their impact of mechanical action in humans.

The ISO 13482:2014 [12] relates situations or hazardous events. However, warns of dangers related to the impact are not yet known, because not exist until the time of publication of the standard data recognized and validated about this situation.

The contact possibility in the interaction between man and robot exists and can be considered normal. However, what is necessary to safeguard is that there is no harm to humans. The ISO/TS 15066:2016 specifies safety requirements for industrial robot systems and the collaborative work environment, being fundamental at this time the definition of the contact forces between humans and robots.

3 Description of the Problem/Situation

Currently there is the presence of robots in companies of any industrial sector.

This constant presence raises a number of questions for researchers. The first question to ask is: work together (man-robot) will be safe? We must ward off humans from robots? Or should we consider robots as part of employees of a company?

With this and other questions, the International Technical Committee for Standardization ISO/TC 199—Safety of machinery constituted in 2014 a Study Group—designated ISO/TC 199—Study Group 1 in order to reflect on the man-machine interaction, their implications and identifying possible limits forces to respect whenever foreseeable contact between the machine and the person to be regarded as dangerous.

From this study aims to define a regulatory framework to support the design, development and use of machines that interact with humans in order to prevent hazards associated with the man-machine contact. The purposes of the normative document that this working group aims to develop relate to the fulfillment of a

fundamental principle of security machines, established in the ISO 12100:2010, which calls for the identification of risks and the definition of measures to prevent the amount, i.e., at the design stage and construction of machinery.

The definition of the safety limits of the contact forces with the human body is not as linear as it sounds. You need to consider a number of factors that influence the final result. The age, physical condition, pain perception, as well as the area of the body affected alters the limit values of the contact force. Apart of the related situations the human being, one has to consider factors associated with equipment such as the frequency and duration of contact, geometry, strength of materials, among others (Safety of Machinery—Mechanical safety data for physical contacts between moving machinery and people—ISO TC 199/SC, 2015).

The development of the regulatory framework aims to provide to designers and manufacturers of machinery, values force of contact to ensure safety of collaborator, without creating excessive constraints or undermine productivity.

4 Methodology

The development of this normative reference includes the following steps: formation of the study group; the feasibility and relevance of normative reference; creation of the working group and development of normative reference.

The ongoing project is to develop a normative reference for defining the contact forces between man and robot (Fig. 1).

The first step was the definition of the study group; ISO/TC 199—Study Group 1; which holds regular meetings in order to aggregate information collected by members, share ideas and bring together the basic elements for decision-making regarding the possibility of developing the reference normative.

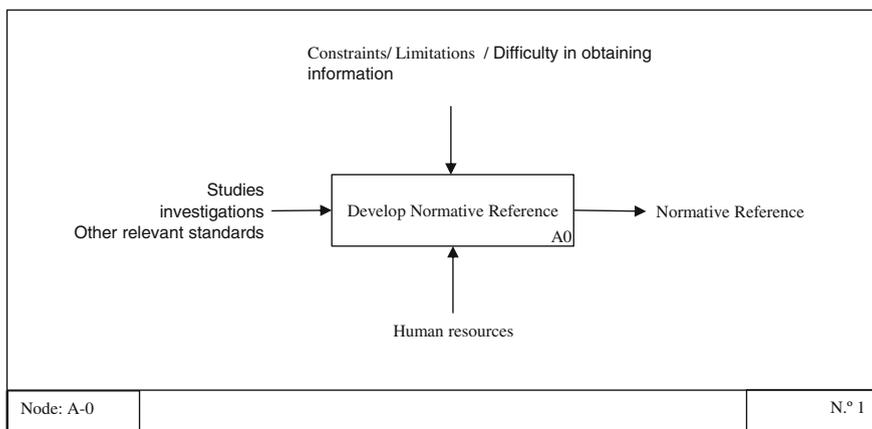


Fig. 1 Level 1—IDEF0 representation of the phases of the project

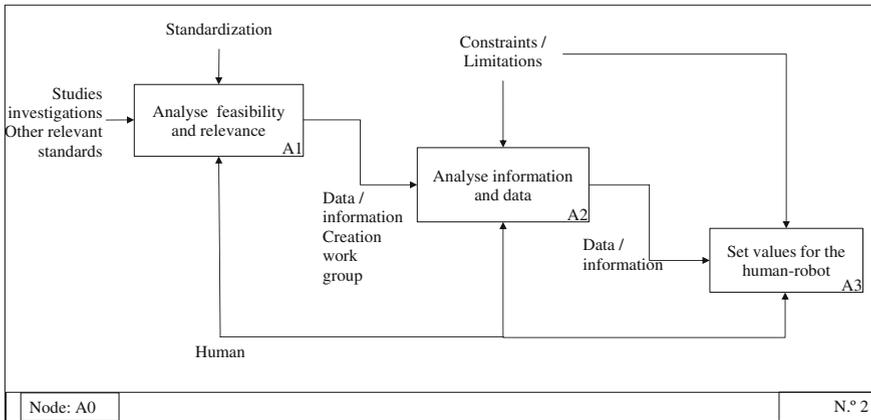


Fig. 2 Level 2—IDEF0 representation of the phases of the project

After this step, the study group will become a work group that will conduct a compilation of data and information; it is the beginning of the development of the reference normative.

The Work Group focuses on the collection of information on international standards from different sectors (some of them referred to below) of industrial activities and on works and scientific studies in this area.

For that the SG1 contacted several other international standardization committees (almost 30), may be interested in this work, in order to achieve the largest possible participation and contribution of different areas of knowledge and make.

The study takes into account parameters associated with human-machine interaction, particularly the human condition (age, gender, affected part of the body, pain perception) and conditions of use of equipment (strength of materials, force, duration and frequency of the contact force application, contact surface geometry) and different forms of contact such as sliding, abrasion, pinching and crushing.

The Figs. 1 and 2 represent, using the IDEF0 tool, the functions or activities of the system developed to obtain the desired result.

The last phase is to get the values of physical contact forces between man and robot be translated on a reference normative, which will provide support to manufacturers and users of equipment.

5 Results

The ISO working group of TC 199 was preceded by a study group that researched and concluded about the existence of information on mechanical contact forces between people and machines. Also concluded that the existing data related with

human/machines contact is available in several published standards but all of them are specifically targeted to certain types of machines.

There are also publications with physical contact data that designers can use but this data is not yet harmonized and readily accessible in a usable form.

In order to make the data harmonized and usable worldwide the working group expects:

- Aggregate the data scattered across multiple sources and create a unifying document.
- Support new types of mobile/autonomous robots that will be deployed outside of industrial settings as well as within industry, including new types of human/machine contacts.
- Specify safe contact limits for new types of deployment, while still covering the traditional approach of preventing contacts.

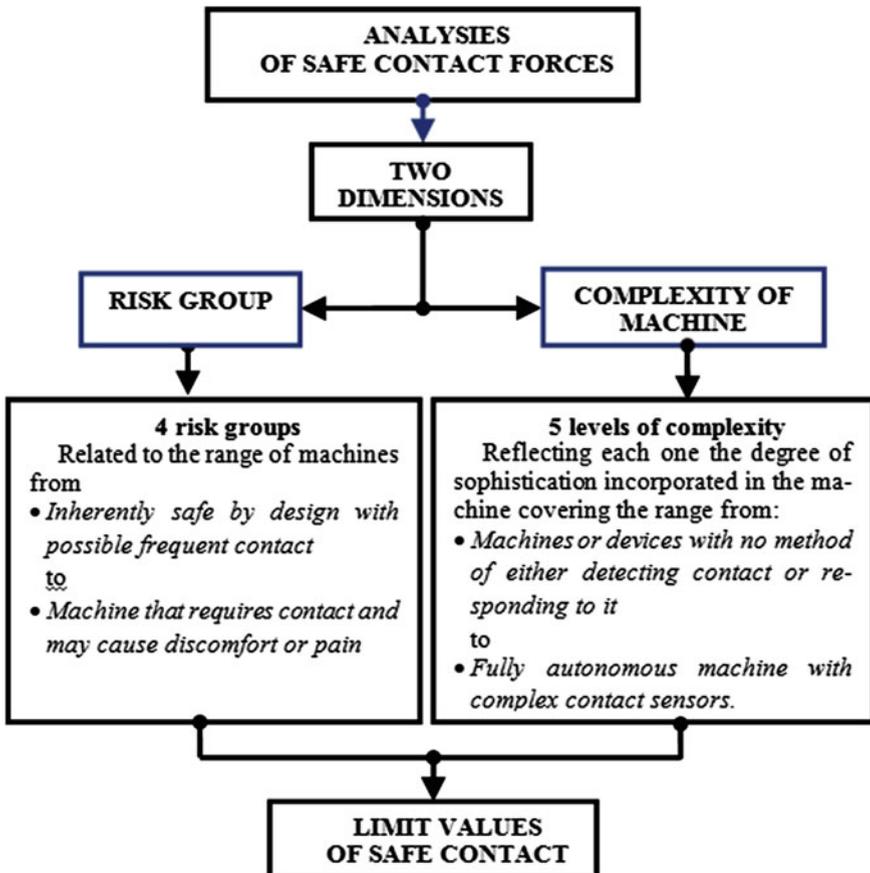


Fig. 3 Dimensions for analysis of safe contact forces

- Supplement existing standards that include human contact information that has no known root in the scientific literature with well-founded and reference data.

At the moment the reference standard is in development stage.

The limit values of contact forces are being set, based on studies and research carried out at existing values and weighting and consideration of different results.

With a view to defining Safe Contact the Work Group presents a proposal for analysis of safe contact forces, based on two dimensions (see Fig. 3): *Risk Group* and the *Machine Complexity Level* (ISO/WD. 6, 2015).

The group see with great interest the definition and inclusion in the standard of a type test—*safety test*—which aims to complement existing information, particularly in complex situations where the type of the contact is not clear, or it can be an alternative to an analysis based on established limits. Within the group there is a consensus that this test must be simple, easy to implement and incorporate material that simulates the response of a person to contact such as “artificial flesh” or other “instrumented biological material” [13].

The study of a set of standards that has being the basis for defining the limit values of contact forces suggests that these values may be expressed in the way that is exemplified in Fig. 4:

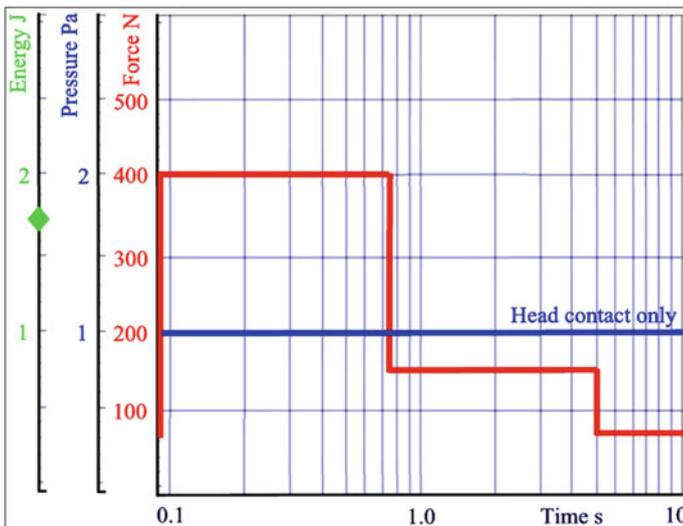


Fig. 4 Example of limit values of contact forces. (ISO TC 199/SC N; Date 2015-03-05; ISO/WD .8 ISO TC 199/SC/WG)

6 Conclusion

Since the study group shown that there is a need of coherent, well-founded and harmonized information about human-machine mechanical contacts the members of International Standardization Committee ISO/TC 199—Safety of Machinery approved by majority to set up a new working group for developing a standard in this subject. Under the ISO rules this working group has to do it within the next 48 months.

The work program is to be prepared and agreed in the kickoff meetings but is foreseen that the first step will be the compilation of the existing data scattered across different sources and prove the consistency of the values by scientific or statistical data.

Is also intended that the standard should give test methods to prove and consolidate by experiments the limit values for human machine contacts.

By the end is expected to publish a standard with vital information to help the designers and manufactures to designing and making machines more friendly-using, inherently safe and increase significantly their productivity.

After preparation of the standard is expected to have a standardized framework that provides reference values to machine designers, in order to promote the safe use of production equipment by users and ensure the development and production of safe machines.

In the future it is intended that this project transposing the theoretical field barrier and be carried out the consolidation/validation of the results obtained by experiments, using measuring instruments and production equipment.

Acknowledgments This article is based on work of the ISO/TC 199—Safety of machinery/Study Group 1. The authors would like to thank the group members the support and interest shown in the presentation and promotion of this paper.

References

1. ISO/WD.8—Safety of Machinery. Mechanical safety data for physical contacts between moving machinery and people (ISO TC 199/SC N, 2015, document from Study Group 1)
2. ISO 10218-1:2011—Robots and robotic devices—Safety requirements for industrial robots—Part 1: Robots
3. ISO 10218-2:2011—Robots and robotic devices—Safety requirements for industrial robots—Part 2: Robot systems and integration
4. ISO 12100: 2010—Safety of machinery. General principles for design—Risk assessment and risk reduction
5. BS EN 16005: 2012—Power operated pedestrian doorsets. Safety in use. Requirements and test methods
6. ISO 14120: 2002—Safety of machinery. Guards—General requirements for the design and construction of fixed and movable guards
7. ISO/TS 15066:2016—Robots and robotic devices—Safety requirements for industrial robots—Collaborative operation

8. EN 71-1:2014—Safety of toys—Part 1: Mechanical and physical properties
9. EN 71-8: 2011 Safety toys—Part 8: Active toys for domestic use
10. CEN/TC 122—prCEN/TR 20442—Safeguarding crushing points by means of a limitation of the active forces
11. Desmoulin, Geoffrey T., Anderson, Gail S.: Method to investigate contusion mechanics in living humans. *J. Forensic Biomech.* Vol. 2, Ashdin Publishing (2011)
12. ISO 13482:2014—Robots and robotic devices—Safety requirements for personal care robots
13. Fonseca, A., Pires, C.: Smart prevention for sustainable safety—human-machine interactions: mechanical safety data for physical contacts between moving machinery and people. In: WOS 8th International Conference—Book of Abstracts. WOS2015 Scientific Committee (2015)

Part VI
An Exploration of Real-World Implications
for Human-Robot Interaction

Droning on About Drones—Acceptance of and Perceived Barriers to Drones in Civil Usage Contexts

Chantal Lidynia, Ralf Philipsen and Martina Ziefle

Abstract The word “drone” is commonly associated with the military. However, the same term is also used for multicopters that can be and are used by civilians for a multitude of purposes. Nowadays, drones are tested for commercial delivery of goods or building inspections. A survey of 200 people, laypersons and active users, on their acceptance and perceived barriers for drone use was conducted. In the present work, user requirements for civil drones in different usage scenarios with regard to appearance, routing, and autonomy could be identified. User diversity strongly influences both acceptance and perceived barriers. It was found that laypeople rather feared the violation of their privacy whereas active drone pilots saw more of a risk in possible accidents. Drones deployed for emergency scenarios should be clearly recognizable by their outward appearance. Also, participants had clear expectations regarding the routes drones should and should not be allowed to use.

Keywords Civil drones · Usage contexts · User requirements · Barriers · Technology acceptance · Piloting experience · Human factors

1 Introduction

Sometimes, there are many names for the same concept. Unmanned aerial system, unmanned aerial vehicle (UAV), remotely piloted aircraft system, multicopter, or drone; they all describe the same or at least very similar pieces of airborne technology.

C. Lidynia (✉) · R. Philipsen · M. Ziefle
Human-Computer Interaction Center (HCIC), RWTH Aachen University,
Campus-Boulevard 57, 52074 Aachen, Germany
e-mail: lidynia@comm.rwth-aachen.de

R. Philipsen
e-mail: philipsen@comm.rwth-aachen.de

M. Ziefle
e-mail: ziefle@comm.rwth-aachen.de

The word “drone,” almost regardless of continent, be it Europe, North America, or even Australia, is often associated with the military and even (covert) observation missions or espionage, e.g., [1, 2]. And while this technology has its roots in military and warfare, see, e.g., [3–5], its applications have entered the civilian realm of useful tools and gadgets, especially since the technological development facilitated ever decreasing sizes of drones that are no longer lethal. Consequently, the term “drone” is also used for multicopters that can be and are used by civilians for a multitude of non-threatening purposes. For one, scale model pilots use them as a different kind of remote-controlled model aircraft. Sport enthusiasts document their own outdoor-activities, especially in regard to extreme-sports. Unmanned aerial systems with either rotors or fixed wings are deployed for precision farming, e.g., [6, 7], or wildlife preservation [8].

Additionally, multicopters have attracted the interest of commercial users who plan to deploy these gadgets to transport and deliver goods, e.g., [9, 10], inspect the structural integrity of buildings, e.g., [11], record documentaries or movie sequences, see, e.g., [12], or even send them out in cases of natural disasters to locate survivors or remaining hazards.

All of these different uses have an impact on the general population. The attitude toward and acceptance of unmanned flying aircrafts is instrumental in establishing this technology as valid option for future civilian applications.

As far as the authors of this paper know, up until now, there has been very little research done on human factors that influence the acceptance of this technology. Sandbrook [8] also commented on the lack of research on social factors concerning drone deployment in non-military contexts. For the most part, technological aspects are the focal points of research, e.g., [11, 13, 14]. Of almost equal interest is the legal side of drone usage: Where am I allowed to deploy my multicopter, how much altitude is allowed for these aircrafts, and what zones are completely out of question for me to fly over? This varies considerably across different countries. Even though the European Union has guidelines on civil drone use above 150 kg, each individual nation has instated their own regulations and laws, some more lenient than others, when it comes to UAVs lighter than 150 kg, see among others [4, 15, 16].

2 Related Work

Only little research has been conducted about the requirements that civilians, i.e., the general public, have so that drones or multicopters can be used as non-threatening “helpers” or useful tools. The author of [17] used differently worded questionnaires to gauge the social acceptance of drones as means of cargo or passenger transportation. Her study showed that the information given to the participants had an impact on how this new technology was perceived. The better informed about the risks and possible benefits, the more favorable the acceptance of remotely piloted aircrafts.

In [1], the authors investigated the perception of risks pertaining the comparison of unmanned aerial vehicles and conventionally piloted vehicles in Australia. They came to the conclusion that the risks associated with either type of aerial vehicle did not differ. They also examined if a different naming of the system had any impact on the perceived risk, using the options ‘manned aircraft,’ ‘drone,’ ‘unmanned aircraft,’ ‘autonomous aircraft,’ and ‘remotely piloted aircraft’ in their study. Again, they could find no difference in the risk associated with each of these systems.

Boucher, see [2] and [18], has conducted qualitative studies to establish a basis of what people associate when first confronted with the topic of civil drones and their non-military deployment. To do so, he used several focus groups, in the UK and Italy, to get an insight into the first associations people have when asked about drones. Two of the main aspects he found out were that privacy, especially fear of its loss, plays a major role in the perception of drones, and use of drones that yields benefits to others (society, wildlife, etc.) was accepted while benefits to the pilot alone (fun, mementos, etc.) were considered unacceptable and this use needs careful monitoring and regulations.

The overview of existing research showed that although larger payloads, such as passengers or bigger deliveries, are better suited for drones with fixed wings, multi-rotor systems are more readily available and used by hobby and commercial pilots. Also, they possess more maneuverability and require less space for take-off and landing. Therefore, these systems are also more likely to be encountered by laypersons.

So far, there is still a decided lack of research on the acceptance of and requirements for drones and the factors that might influence these concepts. One German research centers on a specific user-group, namely firefighters, and their acceptance of drones as tools for their job, especially disaster management and prevention [19]. It might also be of interest if there are significant differences in the evaluation of drones and drone use by those who actively use it, early adopters still, or those who have had none or only passive exposure to drones, especially considering that technological developments and changes in the legal regulations indicate a spread of drone use in civil contexts in the not so distant future [15]. Therefore, these are the starting points of the present study.

3 Method

The method section is structured as follows: First, the findings from previous focus group studies on which the present work is based will be shortly introduced. Second, the development of the measurement instrument, i.e., the questionnaire, will be presented. Subsequently, there will be a summary of the data acquisition and analyses, followed by a description of the gathered sample.

3.1 *Previous Focus Group Studies*

In the run-up to the present study, focus group discussions with experienced drone users and laypeople had been conducted, because, apart from diverse privacy concerns, little was known about concrete requirements on and attitudes toward civil drone technology. Consequently, an explorative approach was needed as a first step. The focus groups aimed at identifying potential usage scenarios, requirements on design and control of drones, and perceived barriers and benefits.

Usage Scenarios. Participants distinguished between three usage contexts: (1) drone use as leisure activity, e.g., model flight or private aerial photography, (2) commercial use, e.g., courier services or inspection of large technical installations, and (3) emergency use, e.g., locating missing persons or aerial reconnaissance at disaster scenes. Depending on the scenario, the participants' requirements on drones varied.

Requirements on Drones. Requirements regarding two main aspects of drones could be identified: (1) flight characteristics and control, e.g., the used route, the possible flight radius, or the level of automation regarding the piloting, and (2) identifiability, as in adapting the structural shape and the color to the purpose of use, as well as giving the drone the means to identify itself by, for example, sending an identification code per wireless communication if requested by anyone.

Perceived Benefits and Barriers. Although many advantages of civil drones were mentioned in the focus groups, e.g., pleasure of flying, new perspectives for photography, or access to hardly accessible places, several barriers to drones could be identified that mainly concern privacy issues. In particular, a possible surveillance and unwarranted intrusion into the private sphere were perceived as barrier to acceptance. Also, the overflight of private territory was considered as critical.

3.2 *Questionnaire*

To quantify the findings from the focus group discussions with a larger sample, a questionnaire was developed. It consisted of four parts, one dealing with user factors and three concerning thematic issues.

Demography and Further User Factors. The first part of the questionnaire collected demographical information, e.g., age, gender, educational level, or occupation. Furthermore, Beier's inventory to measure technical self-efficacy [20] was used to gain an impression of the participants' technical affinities. At last, previous experience with civil drones in terms of usage and passive contact, i.e., having watched flying drones, was gathered for both private and professional contexts.

Evaluation of and Requirements in Different Usage Scenarios. The questionnaire's second section addressed the participants' requirements on drones in three different usage contexts using a within-subject design. Each participant had to

answer questions regarding the requirements on drones in the following contexts: (1) *hobby*, (2) *commercial*, and (3) *emergency*. Short scenario-based introductions with examples derived from the focus groups were given for each context to handle unequal knowledge levels and possible gaps regarding the awareness of concrete purposes of drone use.

The requirements on drones were gathered in a dichotomous way by using questions with polarized, mutually exclusive answer options, whereby one was generally restrictive and one offered more freedom regarding appearance and flights of drones. Table 1 gives an overview of the drone attributes and the related selectable characteristics.

General Evaluation of Civil Drone Technology. The second thematic section dealt with a general evaluation of civil drone technology regardless of a concrete usage scenario. Exemplarily, the following statements used for evaluation shall be mentioned: “*Basically, I find drone technology useful*” or “*What bothers me is that there is no way to identify whether the drone is filming.*” A full listing of presented statements will only be shown in the result section to avoid redundancies. Even numbered Likert-scales (min = 0: “Strongly disagree”; max = 5: “Strongly agree”) were used to compel participants to make more differentiated choices.

Barriers. In the last section, participants had to rank several barriers to drones derived from focus groups from 1 (“most important”) to 5 (“least important”). The questioned barriers included *violation of privacy, pilot’s anonymity, risk of accidents, noise*, and the *missing inference to the purpose of use from the appearance*.

3.3 Participant Acquisition, Data Preparation, and Analyses

The survey was realized as online questionnaire. Participants were acquired at the university, in the social environment of the authors, and by using social networks. In addition, expert forums dealing with civil drones, in particular multicopters, were used to address experienced users.

A total of 253 participants started to fill in the questionnaire. The dropout rate was 20.1 %. Therefore, 53 participants who did not answer at least the first thematic section of the questionnaire were removed from the dataset. The remaining data was

Table 1 Queried drone attributes and given answer options

Drone Attribute	Unrestrictive option	Restrictive option
Form	Customizable	Standardized concerning usage
Color	Customizable	Standardized concerning usage
Flight route	Free choice of user	Approval by authorities
Flight radius	Outside the pilot’s visual range	Within the pilot’s visual range
Piloting	(Temporarily) autonomous	Permanent human control
Identification	No identification needed	Wireless identification ability

analyzed by both parametric and non-parametric statistical methods. Two-tailed tests were used for significance testing and the level of significance was set to $\alpha = 0.05$.

3.4 Sample

A total of $N = 200$ participants have completed the questionnaire. 126 (63.0 %) of them were male, 74 (37 %) female. The age in the sample ranged between 15 and 74 years, the mean was 38.21 years ($SD = 13.38$). The participants' level of education was rather high: the most-often stated educational attainment was a university degree (42.5 %, $n = 81$), followed by graduation from high school (27.0 %, $n = 53$) and vocational trainings (23.0 %, $n = 46$). The remaining participants completed secondary school (9.0 %, $n = 18$) or had not achieved a school-leaving qualification yet (.5 %, $n = 1$). 62.5 % ($n = 125$) of the participants pursued an occupation, while 27.0 % ($n = 54$) were pupils, students, or apprentices. The technical self-efficacy in the sample was rather high with $M = 3.63$ ($SD = 1.33$, scale min = 0, scale max = 5).

Drone Usage. 45 participants (22.5 %) had previously used civil drones in private or professional contexts. The sample's typical drone user is male (95.6 %, $n = 43$) and has a high technical self-efficacy ($M = 4.62$, $SD = 0.62$). Concerning this attribute, the difference between users and non-users ($M = 3.35$, $SD = 1.35$) was significant with $t(198) = -6.117$; $p < 0.001$; $d = -1.036$. In contrast, no significant differences between the user groups could be revealed ($p > 0.05$) regarding age and the level of education. 60.6 % ($n = 94$) of the non-users stated that they have watched the use of a civil drone at least once before.

4 Results

The presentation of the results is structured as follows: First, the requirements on drones in the three exemplary usage contexts will be presented. Second, the general assessments of and attitudes toward civil drone use will be outlined. The section closes with the presentation of findings gathered from the ranking task.

4.1 Requirements on Drones in Different Usage Contexts

The requirements on drones varied to a great extent depending on the usage context. The differences between them were significant with $p < 0.05$ for all requirements, except where otherwise specified in the following. As can be seen in Fig. 1, a majority of the participants stated that drones in the hobby context can have a

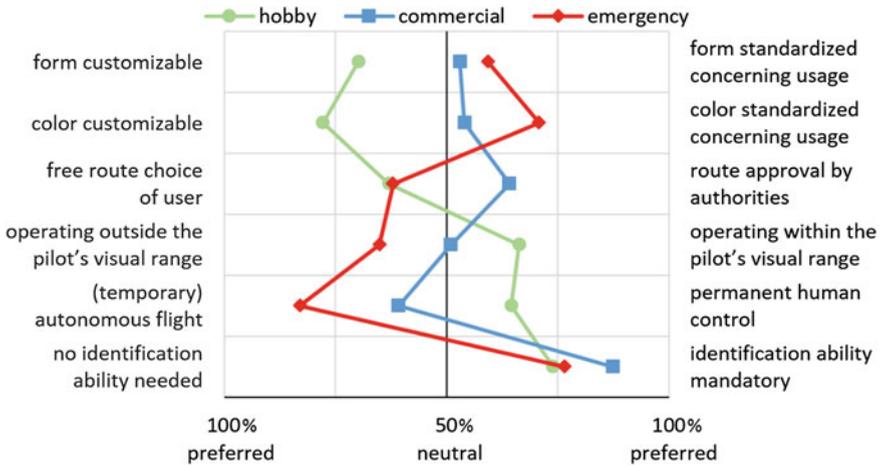


Fig. 1 Percent approval rates of all participants to requirements on drones in different usage contexts

customizable color and form and thus pleaded for unrestrictive rules. In contrast, there was a slight preference for a standardized appearance of commercial and emergency drones. There were no significant differences regarding the latter two use cases on requirements for the drone’s form.

Concerning the flight paths, participants would allow both private and emergency drones (n.s. differences) to use routes freely chosen by their pilots, whereas flight plans of commercial drones should be approved by authorities. With regard to the other flight parameters, a reversed response behavior compared to the requirements on the drone’s appearance was revealed. More specifically, participants requested more restrictive rules for the flights of hobby drones. They were required to fly within the pilot’s visual range and under permanent control of a human pilot. While the participants’ opinions regarding these requirements on commercial drones were rather indecisive, a majority stated that emergency drones should be treated less restrictive and be allowed to operate at least temporarily autonomously and outside the visual range of a pilot.

Finally, one requirement was completely independent from the respective usage context (n.s. differences): the drone’s ability to identify itself per wireless communication was stated as mandatory by a majority of participants for all contexts.

Concerning previous experience with drones, several differences between users and non-users were revealed in their individual requirements on the usage in the presented scenarios. Most of them were significant with $p < 0.001$. The only exceptions to this trend were piloting in the hobby context; flight radius and color in commercial context; and flight radius, form, and color in the emergency use case, for which users’ and non-users’ requirements did not differ significantly.

A full overview of users’ requests can be found in Fig. 2. A representation of non-users’ requirements will be forgone here because it did not significantly differ

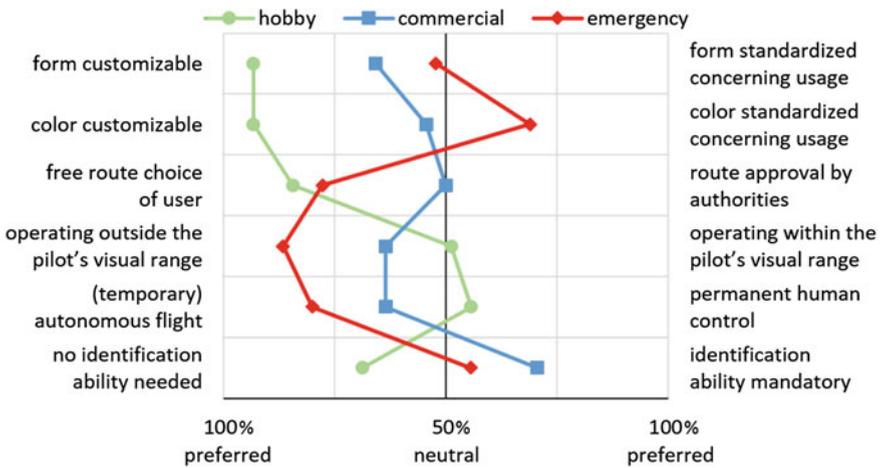


Fig. 2 Percent approval rates of active drone pilots to requirements on drones in different usage contexts

from the total sample (see Fig. 1). Two aspects of users’ response behavior stuck out: First, drone users tended to less restrictive requirements. Although the main effects of the scenarios are comparable, the majority of users leaned towards more freedom regarding drones’ appearance and flight parameters in all usage contexts. Second, in contrast to non-users, the majority of drone users did not request identification capability in the hobby context, whereas such ability is wanted for commercial and emergency use.

4.2 General Evaluation and Attitudes

The general assessment of civil drone technology revealed that the participants did neither fear injuries from accidents with drones nor drones in general. The fear related statements were the only ones rejected on average. In contrast, for the remaining statements, a rather neutral to slightly consenting attitude must be assumed. This applies in particular to concerns about the anonymity of the pilots, the missing possibility to infer the drone’s usage from its looks as well as for the willingness to generally permit drone use and private piloting. Higher approval rates were revealed regarding the unwillingness to accept flights over one’s own property, the perceived usefulness of drones, and the wish for identifiability of ongoing film or photography activities. See Fig. 3 for a complete overview of participants’ average evaluations.

When looking at the differences between users and non-users, it becomes clear that opinions differ significantly regarding all presented items ($p < 0.001$ and absolute values of Cohen’s $d > 0.7$ for all t -tests). These differences are particularly evident in

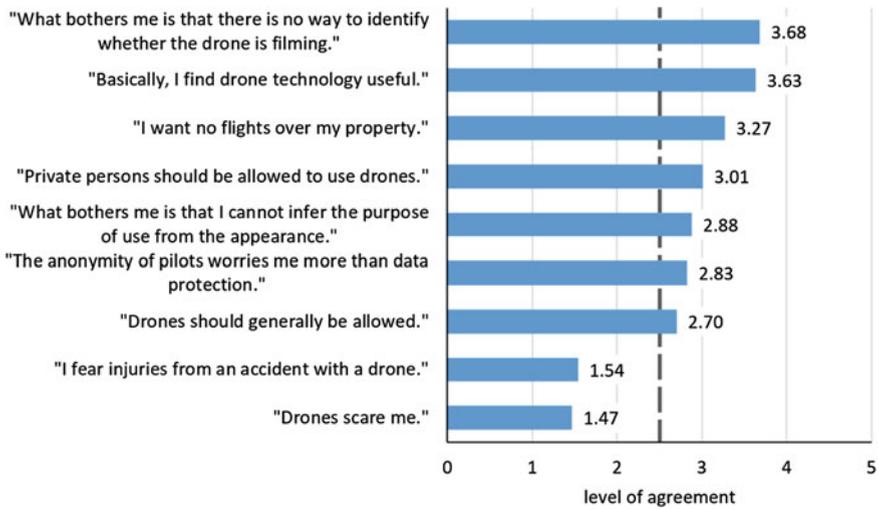


Fig. 3 Average agreement on evaluation statements (min = 0, max = 5). The dashed line indicates the arithmetical neutral level of agreement

the following statements: First, users strongly support the permission of private and general drone use (M approaches the maximum agreement rating), while non-users take a rather neutral position ($t_{\text{general_use}}(196) = -6.566; p < 0.001; d = -1.113$ respectively $t_{\text{private_use}}(197) = -7.528; p < 0.001; d = -1.276$). Second, users considered the missing identifiability of the pilot, the purpose of use, and ongoing film activities as unproblematic, while non-users are worried about these issues ($t_{\text{pilot}}(194) = 4.204; p < 0.001; d = 0.72$, $t_{\text{purpose}}(197) = 6.892; p < 0.001; d = 1.168$, and $t_{\text{filming}}(197) = 9.805; p < 0.001; d = 1.662$). Last, non-users do not want air traffic over their private territory, whereas users would allow it ($t_{\text{air_traffic}}(194) = 5.439; p < 0.001; d = 0.924$). See Table 2 for the complete comparison of the user groups.

4.3 Barriers

Table 3 gives an overview of barriers to the acceptance of civil drones and their median ranks for the complete sample. The potential *violation of privacy* was revealed as most important and significantly differing from the other barriers ($Z = -5.916; p < 0.001$), followed by the *pilot's anonymity* and *missing possibility to infer the drone's purpose of use from its appearance*. In the latter two cases, there was no significant difference between the barriers' rankings ($p > 0.05$). The *risk of accidents* and the *drone's noisiness* were rated even more unimportant. Although the median ranks of these items did not differ, there was a significant difference concerning the ranking ($Z = -3.370; p < 0.001$);

Table 2 Average agreement on evaluation statements (min = 0, max = 5) and related standard deviations with regard to experience groups

Statement	Users		Non-users	
	M	SD	M	SD
“Private persons should be allowed to use drones.” *	4.53	0.97	2.56	1.67
“Basically, I find drone technology useful.” *	4.42	0.99	3.40	1.35
“Drones should generally be allowed.” *	4.07	1.30	2.30	1.66
“I want no flights over my property.” *	2.16	1.80	3.60	1.49
“The anonymity of pilots worries me more than data protection.” *	1.98	1.64	3.07	1.50
“What bothers me is that there is no way to identify whether the drone is filming.” *	1.93	1.86	4.19	1.18
“What bothers me is that I cannot infer the purpose of use from the appearance.” *	1.36	1.76	3.32	1.66
“I fear injuries from an accident with a drone.” *	0.80	1.06	1.75	1.60
“Drones scare me.” *	0.22	0.60	1.84	1.62

A * indicates significant differences with $p < 0.001$

Table 3 Perceived barriers and median ranks (1 = “most important; 5 = “least important”)

Barrier	Median rank
Violation of privacy	2
Pilot’s anonymity	3
Missing inference to the purpose of use from the appearance.	3
Risk of accidents	4
Noise	4

Dashed lines indicate significant differences between the ranked items

Table 4 Most and least important barriers with regard to experience groups

Importance	Barrier	
	Users	Non-users
Most important	Risk of accidents (Mdn = 2)	Violation of privacy (Mdn = 1)
Least important	Missing inference to the purpose of use from the appearance. (Mdn = 4)	Noise (Mdn = 4)

Concerning previous experience, privacy loses importance. As shown in Table 4, drone users ranked the risk of accidents as most important barrier, whereas non-users did not differ from the complete sample. Both the *risk of accidents* ($U = 1705$; $p < 0.001$; $r = -0.35$) and the *drone’s noisiness* ($U = 2297$; $p = 0.002$; $r = -0.22$) were significantly ranked as more important by users than by non-users. In contrast, the *violation of privacy* ($U = 1973$; $p < 0.001$; $r = -0.30$) and *missing possibility to infer the drone’s purpose of use from its visual appearance*

($U = 2068$; $p < 0.001$; $r = -0.27$) were rated more important by non-users than users. There was no significant difference between the user groups concerning the evaluation of the *pilot's anonymity*.

5 Discussion

The ever increasing development of technology for unmanned aerial vehicles and their possible deployment in civil usage contexts prompts questions about the social acceptance of these aircrafts, possible barriers to their use, and likely factors that influence these issues. As current research is still sparse or even nonexistent, the presented study aimed at providing a first insight into the influence of different usage contexts and expertise or prior experience with drone technology on the perceived barriers and requirements people have.

A comparison of different usage scenarios for drones showed that the context of the deployment is important. Different types of usage result in different requirements concerning, for example, looks or flight path. This holds also true for legal regulations that vary depending on the usage context. While this study could not detect a clear position concerning commercially used drones, UAVs used in hobby or emergency contexts evoke clear opinions on what is wanted or needed for their acceptable use. While drones for leisure usage can be less restricted in their look and design, their handling and flight paths are to be carefully observed. For emergency drones, the opposite holds true with their design clearly identifying their purpose but given *carte blanche* as to where they are allowed to fly.

In general, though, this study could not detect any fear of this technology. On the contrary, its beneficial uses were appreciated. Nevertheless, it was also shown that drones evoke a lot of concerns about the safety of one's privacy, especially in regard to visual recordings, which mirrors results from other studies such as [18].

The factor of expertise, the status of active drone user or non-user, yielded a lot of opposing opinions and results. Non-users are very concerned for their privacy, about who is piloting the drone, and if they are filmed without their knowledge. Although they do not fear the aircraft itself, they are not sure about its use, especially in the hands of hobby pilots. Another big problem for laypeople is the overflight of private property.

Drone pilots, on the other hand, are less concerned for their privacy. They are more worried about the risk of accidents, even though they are more accustomed to the technology. Unsurprisingly, active drone operators have less restrictive requirements on drones in civil usage contexts than the general public with little or no contact to this technology. To facilitate a better acceptance of this novel technology use, in accordance to previous results, [17], a better information of the public and framing of drones is necessary. Clear regulations, which are either recently in place or at least in development, and a more transparent use of drones are needed to satisfy both advocating pilots and concerned laypeople who are warier of unpermitted recordings of their person and property than they are of injuries caused by crashing aerial vehicles.

6 Limitations and Outlook

The questioned sample was rather educated and should have been more heterogeneous. Based on this and the source of acquired participants, the results presented here cannot accurately describe the opinion of the general public. Although possible barriers to civil drone use have been ranked, the results based on this method do not disclose absolute significance. Therefore, further studies need to be conducted. For one, tradeoffs for drone deployment in civil usage contexts should be examined via conjoint analysis: Under what terms are users and/or civilians willing to accept impediments, especially concerning privacy issues? Is it necessary to have information about the pilot's identity or the reason for UAV deployment? How important are personal benefits/advantages?

Furthermore, other user factors such as age (especially the status of digital native or digital immigrant), gender, or the personal attitude concerning privacy and data security should be examined in regard to their influence on the acceptance of UAVs. Does the opinion change when it is no longer only early adopters who use the technology and the people coming into contact with drones are more diverse regarding, for example, gender, education, or technology self-efficacy?

The most controversially viewed possibility of UAVs is that of autonomous flight. Our sample mostly rejected it for hobby pilots and, in Germany it is currently prohibited, but the delivery of goods, for example, would rely heavily on this aspect and, therefore, a change in its acceptance is needed. For this, possible factors could be the pervasiveness but also a more reliable technology to prevent in-air collisions or crashes. This should be examined in further studies.

Another very interesting component is the cultural bias. The present study only included participants from Germany. Other qualitative studies, [2, 18], have already shown different privacy concerns from citizens of the UK, who are used to CCTV surveillance, and Italians, who do not have a lot of CCTV and therefore are a lot more skeptical about drones and the possibility of being filmed or recorded, with or without their knowledge.

Acknowledgments The authors thank all participants for their patience and openness to share opinions on a novel technology. Furthermore, thanks go to Dennis Lohse for his research assistance.

References

1. Clothier, R.A., Greer, D.A., Greer, D.G., Mehta, A.M.: Risk perception and the public acceptance of drones. *Risk Anal.* **35**(6), 1167–1183 (2015)
2. Boucher, P.: Domesticating the drone: the demilitarisation of unmanned aircraft for civil markets. *Sci. Eng. Ethics* **21**(6), 1393–1412 (2015)
3. Villasenor, J.: Drones and the future of domestic aviation. *Proc. IEEE* **102**(3), 235–238 (2004)
4. Finn, R.L., Wright, D.: Unmanned aircraft systems: surveillance, ethics and privacy in civil applications. *Comput. Law Secur. Rev.* **28**(2), 184–194 (2012)

5. Bracken-Roche, C., Lyon, D., Mansour, M.J., Molnar, A., Saulnier, A., Thompson, S.: Surveillance drones: privacy implications of the spread of unmanned aerial vehicles (UAVs) in Canada. In: Surveillance Studies Centre, Queen's University, Kingston, Ontario, Canada, (2014)
6. Ross, P.E.: Open-source drones for fun and profit. *Spectr. IEEE* **51**(3), 54–59 (2014)
7. Grenzdörffer, G.J., Engel, A., Teichert, B.: The Photogrammetric Potential of Low-cost UAVs in Forestry and Agriculture. *Int. Arch. Photogrammetry, Rem. Sens. Spat. Inform. Sci.* **31**(B3), 1207–2014 (2008)
8. Sandbrook, C.: The Social implications of using drones for biodiversity conservation. *Ambio* **44**(Suppl 4), 636–647 (2015)
9. DHL: unmanned aerial vehicles in logistics. A DHL perspective on implications and use cases for the logistics industry. In: DHL, Troisdorf, (2014)
10. Agatz, N., Bouman, P., Schmidt, M.: Optimization approaches for the traveling salesman problem with drone. In: ERIM Report Series, vol. Reference No. ERS-2015-011-LIS. (2015)
11. Hallermann, N., Morgenthal, G.: Unmanned aerial vehicles (UAV) for the assessment of existing structures. *IABSE Symp. Rep.* **101**(14), 1–8 (2013)
12. Ravich, T.R.: Commercial drones and the phantom menace. *J. Int. Media and Entertainment Law* **5**(2), 175–213 (2014)
13. Colomina, I., Molina, P.: Unmanned aerial systems for photogrammetry and remote sensing: a review. *ISPRS J Photogrammetry and Rem. Sens.* **92**, 79–97 (2014)
14. Perritt Jr., H.H., Sprague, E.O.: Drones. (2014)
15. Juul, M.: Civil drones in the european union. In: European Parliamentary Research Service (ed.). European Union, (2015)
16. House of Lords: Civilian Use of Drones in the EU. In: 7th Report of Session 2014-15. HL Paper 122. European Union Committee, London, (2015)
17. MacSween-George, S.L.: A public opinion survey- unmanned aerial vehicles for cargo, commercial, and passenger transportation. Paper presented at the 2nd AIAA, San Diego, California, September 15–18, 2003
18. Boucher, P.: You wouldn't have your granny using them: drawing boundaries between acceptable and unacceptable applications of civil drones. *science and engineering ethics* (2015)
19. Hermanns, A.: Anwender-Akzeptanz und Bewertung unbemannter Flugsysteme (Drohnen) im Katastrophenschutz. *Theorie, Empirie, regulatorische Implikationen. Zivile Sicherheit*, vol. 6. LIT Verlag, Münster (2013)
20. Beier, G.: Kontrollüberzeugung im Umgang mit Technik [Locus of control when interacting with technology]. *Rep. Psychol.* **24**(9), 684–693 (1999)

Factors Affecting Performance of Human-Automation Teams

Anthony L. Baker and Joseph R. Keebler

Abstract Automated systems continue to increase in both complexity and capacity. As such, there is an increasing need to understand the factors that affect the performance of human-automation (H-A) teams. This high-level review examines several such factors: we discuss levels and degrees of automation, the reliability of the automated system, human trust of automation, and workload transitions in the H-A system due to off-nominal events. The influence that each of these factors has on the H-A team dynamic must be more completely understood in order to ensure that the team can perform to its maximum potential. Thorough understanding of this dynamic is especially important to ensuring that H-A teams can succeed safely and effectively in critical contexts.

Keywords Automation · Human-systems integration · Human-Automation teams · Team performance · Reliability · Trust of automation · Off-nominal events

1 Introduction

Since the dawn of the industrial revolution, automation has held the promise of vastly improving the work efficiency of humankind. Within the last few decades, we have seen the human-automation (H-A) relationship change, moving the role of automation from tool to teammate in order to drive and sustain this change. The proliferation of automation has come with the task of understanding how automation fits into the existing puzzle of human working relationships, and there has been much research to guide the process of placing that puzzle piece. This review discusses some of that research.

A.L. Baker (✉) · J.R. Keebler
Embry-Riddle Aeronautical University, 600 S Clyde Morris Blvd,
Daytona Beach, FL, USA
e-mail: BakerA19@my.erau.edu

J.R. Keebler
e-mail: KeeblerJ@erau.edu

To put this research into an appropriate context, we will consider the environment that can perhaps place the most demand on an H-A team: the void of space. Considered an ICE—*isolated, confined, and extreme*—environment, space and future travel through it will require a shift in the way we think about H-A teams. The next frontier in human space exploration is a mission to Mars. With current technology, it would take a human crew about six months to get there, and then six months to return. A mission to land crew on Mars may have to last for several months at the minimum, perhaps more, depending on when the orbits of Earth and Mars provide for favorable launch windows. This long-duration spaceflight (LDSF) means that mission parameters for the human operators will be different from previous missions. The long communication time between mission control on Earth and the astronauts (radio signals can take up to 20–30 min to travel between the planets) means that astronauts will be expected to perform within “bounded autonomy”, meaning that they are free to perform most functions as they see fit, with lightly-interspersed input from mission control at critical junctures [1]. Any LDSF mission must consider all of these extra constraints when designing the H-A system, in order to achieve safe, effective, efficient performance.

To this extent, this paper will review several factors that affect the performance of H-A systems. Levels and degrees of automation will be reviewed and considered and the performance impact of reliability of automated systems will be assessed. The human side of the H-A team will be considered, with specific emphasis on factors guiding human trust of automation. We will also consider the consequences of failures in the H-A system, and we will investigate factors that improve performance outcomes after failures. Finally, we will draw conclusions about ways to improve the overall performance of H-A teams, and we will provide directions for future research.

2 Stages and Levels of Automation

Automation is defined by the manner in which it carries out its tasks, and by the extent to which it is given certain types of tasks. Before going further, it would be best to define what we mean by automation, because the term can be used many different ways. We will use “automation” to refer to a computerized or mechanical system used to carry out a role or a type of work performed by humans.

Automated systems can be differentiated in a few ways. There are generally two schools of thought when it comes to describing levels of automation. The first school of thought arose several decades ago with a seminal paper by Sheridan and Verplank [2], which discussed the teleoperation of submersible vehicles and work platforms. The article further discusses control hardware (such as sensors, communication, controls, and the workstation) and how it affects performance of the human operator. This was one of the earlier works which assessed the performance of H-A teams. In order to characterize the automated assets used by the operators, the authors outlined a model to describe different levels of automation that were

possible, with each level of automation providing a different level of support to the human’s operation of the system. The model is provided in Table 1.

Our original definition of automation referred to the complete or partial replacement of human operation of a task with an automated system. In contrast, the model in Table 1 implies that automation is not all-or-none, but rather that there are distinct levels with various amounts of automation. As the level of automation increases, the amount of work entrusted to the human operator is reduced, as task demands are increasingly offloaded to the automated system. At the 10th level, an automated system is in full control of all decisions and does not inform human operators. Rarely are systems automated to this extent; generally, some level of input from a human, or some ability to inform human operators of task outcomes, is always useful to have.

The second school of thought in defining automation is more recent, and coincides with the rise of information-processing research. Notably, Parasuraman et al. [3] created a model of automation that grounds automation levels in an information-processing paradigm. It is helpful to first consider a simplified model of an information-processing task. An example is provided in Fig. 1.

In the first stage of the model, acquisition, information about the environment and the state of the system is gathered and synthesized from multiple sources. In a human operator, this is done via the senses, while an automated entity will make

Table 1 10 levels of automation, from Sheridan and Verplank [2]

Automation level	Automation description
1	The computer offers no assistance; human must take all decision and actions
2	The computer offers a complete set of decision/action alternatives, or
3	Narrows the selection down to a few, or
4	Suggests one alternative, and
5	Executes that suggestion if the human approves, or
6	Allows the human a restricted time to veto before automatic execution, or
7	Executes automatically, then necessarily informs humans, and
8	Informs the human only if asked, or
9	Informs the human only if it, the computer, decides to
10	The computer decides everything and acts autonomously, ignoring the human



Fig. 1 Information-processing model. While this is a gross simplification of the complexity of human (or machine) information processing, it is useful in understanding the process of going from data acquisition to action execution

use of sensors. This stage includes the allocation of attention and cognitive pre-processing of information. The second stage of the model, analysis, involves working memory to a large extent. Here, the human or automated system will consciously perceive, manipulate, and process retrieved information. In the third stage, cognitive processing is used to derive an appropriate response about the information gathered. In the fourth stage, the decision is acted upon.

What Parasuraman, Sheridan, and Wickens did was to take this four-stage model and describe how each of the stages could have its own levels of automation. This stands in contrast to Sheridan and Verplank's model, which only considers how the automated systems come to their decisions. In this newer model, the entire information-processing process is considered, and each stage can be automated at a different level. This accounts for a multitude of modern computerized systems that are specialized for acquiring and analyzing massive amounts of information very quickly, as well as synthesizing it into a set of choices to be made for a human operator to decide on. As one example, the proliferation of internet-usage data allows companies to collect large amounts of data about how their customers use their sites. Systems are able to harness this information, analyze traffic and purchase patterns, and provide information about what parts of the site are making money, so that the operators can decide on how to capitalize on this information.

Returning to our LDSF context, let us imagine the existence of an automated system which can control power allocation to various systems of a hypothetical spaceship. To what extent should the system be automated? In other words, should all power allocation actions be made as the system deems appropriate? Should actions only be taken when there is a near-perfect chance that the power allocation will not result in failure? What if those conditions are not always met, and the automated system is not able to do much in the way of allocating power, despite its tasking?

The question of the extent of automation is a difficult one. The best solution (and, unfortunately, the one that provides the least amount of guidance at face value) is that a system needs to be automated "just the right amount". If the automated system is only capable of very little, or is only entrusted with menial tasks, human operators are not likely to trust the automated system [4]. However, if the automated system has a very large amount of responsibility, there is potential of a significant "lumberjack effect" [5]. The system is like a large tree in a forest: the bigger it is, the harder it will fall, or in automation terms, the more responsibility the automation has, the greater the performance decrement when the system fails. Onnasch et al. [5] evidence for the existence of a point, called a , at which automation should not be given further responsibility, as crossing this point results in significantly worse post-failure performance. The authors provide a chart, reproduced in Fig. 2, which shows the relationships between human operator situational awareness, operator workload, system failure performance, and system routine performance, each as a factor of degree of automation.

Being cognizant of a is not enough. System designers for LDSF are given the difficult task of getting as close to a as possible without unduly jeopardizing the performance of the human-automation team when a failure in the automated system

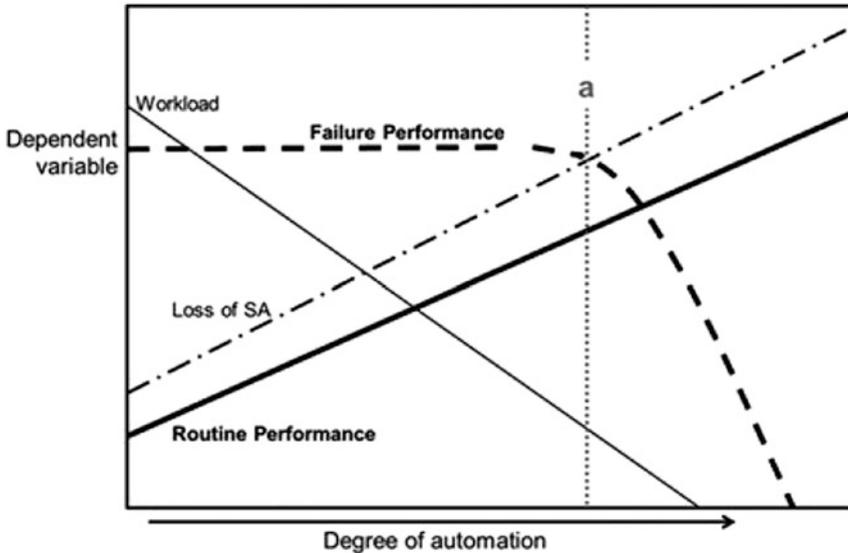


Fig. 2 Several variables as a factor of degree of automation [5]. Note the sharp drop in failure performance after the system is automated past *a*

occurs. The system performance benefits and reductions in operator workload are non-negligible and are the drivers that demand that the degree of automation used is as extensive as is safely possible.

3 Reliability of Automated Systems

Reliability of the automated system is a factor that plays a large role in how the human operator actually uses the system, and in turn, how the system is able to perform. In essence, reliability is the rate at which an automated system performs properly and predictably. Understandably, greater automation leads to greater performance by an H-A team. However, an unreliable automated system places greater task demands on the human operator, who must then compensate for potentially incorrect information, analyses, decisions, or executions of action. Yeh and Wickens [6] assessed the performance of participants on a target-detection task using a cue-detection system that changed in its reliability. Starting off as reliable, the system became unreliable at a certain point in the task. The authors found that participants adjusted their usage of the system to compensate for perceived flaws in the system, with users relying more on their own judgments than those of the system when they believed that the system was unreliable.

User adjustments are not the only outcome of an unreliable system. Rovira et al. [7] conducted a study assessing participants' response times in which they were

tasked with deciding whether targets were enemy or friendly. Participants had the support of an automated system, which aided their identification of the targets. This system became similarly unreliable at certain points. The authors found that participant response times were slower when dealing with the unreliable level of automated support, which provides more evidence for the idea that an automated system that is not consistently reliable induces a performance decrement on the human-automation system.

4 Trust of Automation

Predictability of the automated system's choices, actions, and capabilities is important to the human side of the team: as the human's understanding of the automated asset's purpose and abilities increases, the potential performance of the team increases as well [8]. This is referred to as having a shared mental model of the task at hand, in that the user's mental model of the task and the automation fits with the model of what the automation perceives and is capable of. Lee and See [9], who reviewed the existing literature on human trust of automated systems, further inform this congruence between automation capability (i.e. trustworthiness) and operator understanding of the automation. The authors illustrated several concepts that are not new to the field, but which are very useful in understanding the complex relationship of the H-A team (Fig. 3).

User trust of an automated system is considered calibrated when it matches the system's capabilities, and calibrated trust is conducive to effective performance of

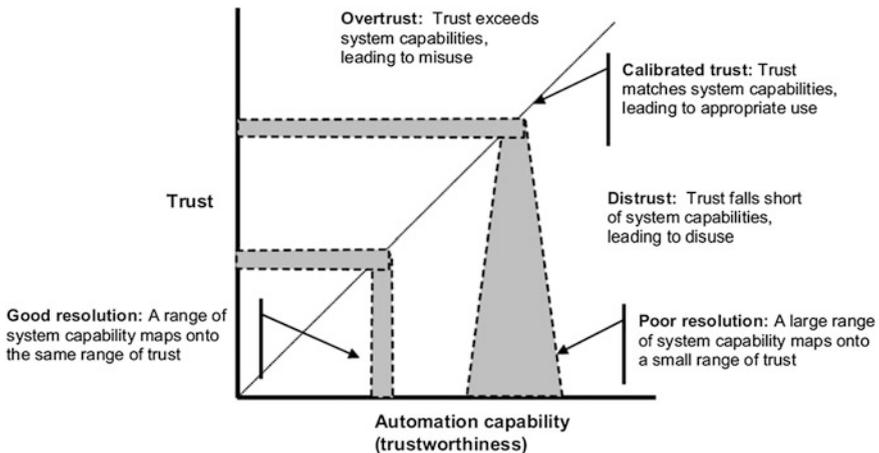


Fig. 3 Relationship between automation capability and user trust [9]. The *diagonal* indicates an appropriate calibration of trust. Areas above and below the diagonal result in overtrust or distrust of the system, respectively

the human-automation team, as it results from a good understanding of the automated system's capabilities. Calibrated trust has good resolution, as a certain range of system capabilities matches with a certain range of user trust. Poor resolution (and poor calibration) results when capability does not match user trust. Parasuraman and Riley [10] provide more insight into the errors committed by human operators with poorly calibrated trust. The authors explain that humans can make inappropriate use of an automated system via misuse or disuse. Misuse refers to detrimental overreliance on the automated system (as when the system is incapable of performing to the operator's expectations). Disuse refers to detrimental underreliance on the automated system (as when the human is incapable of performing to expectations, and needs the automated system to perform better). Thus, a strong H-A team demands that operators have a clear understanding of the automated system's capabilities, and that they understand the situations in which the system's is most useful.

5 Workload Transition: When Automation Fails

Automation failure is not a question of if, but when. If LDSF is to succeed, the H-A system must be capable of handling these failures swiftly, appropriately, and effectively. The point at which automation fails is referred to as a workload transition [11] referring to the transition of workload that was previously managed by the automated system onto the human operator. This process is also euphemistically referred to as an off-nominal event. Understandably, workload transition places a large demand on the human operator, who must now manage not only the automated system's tasks but the repair procedures as well.

The CODDMAN Factors that determine performance of the H-A system after a workload transition are largely related to the design of the automated system itself: going back to the lumberjack effect, a system that is highly automated and has very much responsibility will fail in a more catastrophic way than will a system with less tasking or automation. In addition, as explained earlier, systems that are better understood by their operators are better able to manage workload transitions. However, several human factors affect performance after workload transition, and most of those factors are related to cognitive ability and performance (such as working memory capacity, knowledge of repair procedures, resistance to stress, etc.).

Sebok et al. [12] investigated the process of a workload transition, as well as how various human-automation interaction (HAI) factors were affected by automation at each stage of the information-processing model. The model further considers how fatigue affects various operator tasks and abilities (Fig. 4).

The CODDMAN model [12] provides a simple way of representing a large number of factors that relate to various stages of information processing. In this model, a workload transition occurs between the Detection stage and the Diagnosis stage. Notably, we can see a few of the effects that we have so far covered. System reliability significantly reduces operator monitoring of the system. DOA refers to

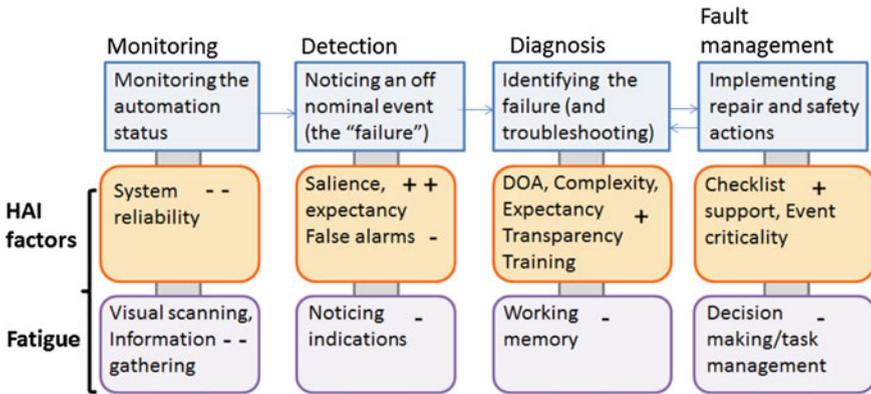


Fig. 4 Complacency Effects on Detection, Diagnosis, and Fault Management (CODDMAN) [12]. Pluses indicate that the relevant stage (e.g. monitoring, detection, etc.) of human tasking is improved by the features with pluses. Minuses indicate that those features reduce the effectiveness of the stage. For the Fatigue row, the effects of fatigue reduce each of the abilities or activities listed in each stage

degrees of automation, and in this case, is not in contrast with the lumberjack effect: rather, this refers to research which has shown that more highly-automated systems (independent of how much responsibility they are tasked with) can provide better support for operators after a failure, which improves fault diagnosis [5]. As a further point, the authors of the CODDMAN model note that the SEEV model [13], which predicts general performance of human operators in multi-modality situations, further validates several of the factors within the CODDMAN model. In sum, each of the factors within the CODDMAN model, and how they relate to the performance of an H-A team undergoing a workload transition, is critical to informing the development of appropriate H-A teams and tasking for LDSF.

6 Conclusions: Designing Automation for Effective H-A Team Performance

We have reviewed several of the factors that affect performance of the H-A team, especially as applied to the context of long-duration spaceflight. In order to prevent or mitigate the risks of off-nominal events, each of these factors must be thoroughly considered during the design of automated systems. Our concluding recommendations for H-A system design are as follows:

1. An automated support system must have an appropriate level of automation so as not to put the team at excessive risk when it fails, in line with the lumberjack effect.

2. The system must be reliable, which will inspire calibrated trust of it by the human operators, which in turn will allow for better performance of the H-A team due to the congruence of their shared mental models.
3. The system should be designed to avoid causing operator misuse or disuse.
4. The system must be designed to allow the operators to swiftly and accurately diagnose and manage faults in the event of workload transition. Such steps may include improving transparency of the system (via improving display ecology), adding checklist support to the fault management step, or improving operator training on system repair and management.

With appropriate consideration of each point, we can give a team of humans with automated assets the best chance to perform to their fullest capabilities and survive the unforgiving demands of LDSF. While LDSF is a special case where the H-A system must be implemented with extreme care, these points can be applied to any H-A system in order to best support the performance of the H-A team.

References

1. Roma, P.G., Hursh, S.R., Hienz, R.D., Brinson, Z.S., Gasior, E.D., Brady, J.V.: Effects of autonomous mission management on crew performance, behavior, and physiology: insights from ground-based experiments. In: Vakoch, D.A. (ed.) *On orbit and beyond*, pp. 245–266. Springer, Heidelberg (2013)
2. Sheridan, T., Verplank, W.: *Human and computer control of undersea teleoperators*. Man-Machine Systems Lab, Massachusetts (1978)
3. Parasuraman, R., Sheridan, T., Wickens, C.D.: A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. A*. **30**, 286–297 (2000)
4. Desai, M., Medvedev, M., Vazquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggerman, C., Yanco, H.: Effects of changing reliability on trust of robot systems. In: 7th ACM/IEEE International Conference on Human-Robot Interaction, pp. 73–80. IEEE Press, New York (2012)
5. Onnasch, L., Wickens, C.D., Li, H., Manzey, D.: Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Hum. Factors* **56**, 476–488 (2014)
6. Yeh, M., Wickens, C.D.: Attention and trust biases in the design of augmented reality displays. DTIC Document: Prepared for the US Army Research Laboratory, Interactive Displays Federated Laboratory (2000)
7. Rovira, E., McGarry, K., Parasuraman, R.: Effects of imperfect automation on decision making in a simulated command and control task. *Hum. Factors* **49**, 76–87 (2007)
8. Schuster, D., Ososky, S., Jentsch, F., Phillips, E., Lebiere, C., Evans, W.A.: A research approach to shared mental models and situation assessment in future robot teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 55, pp. 456–460. SAGE Publications, California (2011)
9. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* **46**, 50–80 (2004)
10. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. *Hum. Factors* **39**, 230–253 (1997)
11. National Research Council: *Workload transition: implications for individual and team performance*. National Academies Press, Washington DC (1993)

12. Sebok, A., Wickens, C., Clegg, B., Sargent, R.: Using empirical research and computational modeling to predict operator response to unexpected events. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 58, pp. 834–838. SAGE Publications, California (2014)
13. Wickens, C.D.: Multiple resources and mental workload. *Hum. Factors* **50**, 449–455 (2008)

A Neurophysiological Examination of Multi-robot Control During NASA's Extreme Environment Mission Operations Project

John G. Blich

Abstract Previous research has explored the use of an external or “3rd person” view in the context of augmented reality, video gaming, and robot control. Few studies, however, involve the use of mobile robot to provide that viewpoint, and fewer still do so in dynamic, unstructured, high stress environments. This study examined the cognitive state of robot operators performing complex search and rescue tasks in a simulated crisis scenario. A solo robot control paradigm was compared with a dual condition in which an alternate (surrogate) perspective was provided via voice commands to a second robot employed as a highly autonomous teammate. Subjective and neurophysiological measurements indicate an increased level of situational awareness was achieved in the dual condition along with a reduction in workload and decision oriented task engagement. These results are discussed in the context of mitigation potential for cognitive overload in complex and unstructured task environments.

Keywords Human robot interaction • Cognitive state • Situational awareness • Workload • Decision making • Robot assisted rescue

1 Surrogate Perspective(s) for Robot Control

The integration of an omniscient or “3rd Person” perspective into the video game and augmented reality industry has inspired a variety of research that is directly relevant to multi-robot control. Salamin and colleagues, for example, report that many video gamers prefer to use a third person perspective for moving an avatar through an artificial environment, but usually switch back to a 1st person viewpoint from within the avatar itself to perform “thin” tasks where dexterous manipulation is required [1]. Similar preferences have been reported in teleoperation of mobile

J.G. Blich (✉)

AFRL 711th HPW/RHC, Wright Patterson AFB, Dayton, OH, USA
e-mail: john.blich@us.af.mil

© Springer International Publishing Switzerland 2017

P. Savage-Knepshild and J. Chen (eds.), *Advances in Human Factors in Robots and Unmanned Systems*, Advances in Intelligent Systems and Computing 499,
DOI 10.1007/978-3-319-41959-6_28

341

robots [2, 3] and the use of augmented reality for remote grasping and manipulation [4, 5].

Despite an abundance of research published on Human Robot Interaction (HRI) there appears to be a dearth of literature documenting the use of a robotic team mate to provide this 3rd “person” perspective. Note that since an external offset of this nature can be provided by any number of non-human devices, the term “surrogate perspective” is used in lieu of the “3rd” (or 4th) person label. In turbulent conditions where sediment, dust or other obscurants interfere with the primary robot’s sensors, a mobile offset or “surrogate” perspective can potentially save the mission from total failure by finding a gap in the obscurant cloud to exploit and guide the primary platform and/or its manipulator(s) from a different angle. This approach has the potential to enable and expand upon an otherwise ineffective object identification or retrieval task. Multiple perspectives also present the potential for stereo-sensing approaches that offer far more capability in handling dynamic targets and environments than mono-sensing.

Given these considerations, this investigation reflects a seedling investment in the SPAARC, or Surrogate Perspective for Adaptive and Agile Robot Control concept. Its goal was to determine quantitative measures of surrogate perspective value (or burden as the case may be) derived from teaming a highly autonomous robotic teammate with a human teleoperator in a challenging survey and manipulation task. This particular effort explored the SPAARC concept in an underwater context with Remotely Operated Vehicles (ROVs) in order to slow down human robot interactions for close examination in a relatively viscous operating realm while exploiting the inherent risks and stressors associated with NASA’s Extreme Environment Mission Operations (NEEMO) project.

2 Research Environment and Expectations

This investigation took place at sea aboard a 40’ research vessel supporting the Aquarius Reef Base (ARB)—a unique underwater habitat operated by Florida International University and often used by NASA’s NEEMO program office to simulate the confined quarters that astronauts often have to endure while working in space. Unlike other simulations and Mars analog research endeavors, NEEMO astronauts living and working aboard the ARB do so under the substantial risk and stress associated with saturation diving and confined quarters—often for weeks at a time. That level of risk and discomfort, while manageable and minimal compared to combat, presented the military with an excellent opportunity to investigate cognitive function under stress in a dynamic, unstructured environment.

The simulation vignette used in this research was a modified version of the robot assisted rescue scenario enacted by Blich and colleagues during the 2001 Mars Flashline project [6]. This simulation involves a stationary habitat (simulated by the Aquarius Reef Base) and a manned rover (simulated by a stack of air tanks located approximately 50’ away from the ARB) that were simultaneously hit by a

micro-meteorite storm that punctured protective shielding and caused precious gases and fluids to leak outside at an alarming rate thereby threatening the humans inside. Meteorite shrapnel from the impact also hit several crewmembers in each environment. Subsequent life-threatening injuries obviously demanded immediate attention by fellow crewmembers making robotic assistance crucial to crew survival. Small rovers with sophisticated end effectors (simulated by the Video Ray submersibles and two jaw grippers) were called upon to survey confined spaces too small for humans to penetrate. Their task was to locate the leaks (simulated by punctured bicycle tubes placed strategically around the Aquarius Reef Base) by support divers, and then complete an ad hoc repair of each leak by applying a rapid curing epoxy patch through break-away tubes imbedded in each gripper jaw (simulated by multi-colored electrical tape wrapped around the puncture site on each bicycle inner tube).

Returning now to relevant HRI literature, an abundance of research on cognitive demand associated with multi-robot control calls into question which way the cost benefit scale might tip if surrogate perspective(s) were implemented in the form of mobile robots [7–9]. The majority of this literature, however, endorses the role of adaptive automation as a powerful ally in mitigating the mental workload and attentional demands associated with multi-robot control [9, 10]. With that in mind, the second robot intermittently injected into this rescue scenario was simulated at a high level of autonomy and artificial intelligence sufficient to respond effectively to voice commands.

Given the naturalistic interface between human and robotic team mate implemented above, it was expected that the ROV operators attempting this challenging rescue mission in the SOLO condition would work harder, exhibit reduced situational awareness, and present a much more intensive engagement profile for choice selection and decision making than when the surrogate perspective was made available in the DUAL condition. It was obviously expected that overall mission performance would improve in the DUAL condition as well.

3 Method

3.1 *Participants*

Four male military personnel from the 88th Security Forces Squadron at Wright Patterson Air Force base were selected from a field of 13 other volunteers to participate in this research based on their performance during a four hour assessment and selection process described below. As active duty military service members, these participants ranged in age from 21–30 years of age, had normal hearing and normal or fully corrected eyesight, and received no monetary incentive beyond standard travel reimbursements.

3.2 Apparatus

The mobile robot manipulators used for this experiment were Video Ray Pro-III Remotely Operated Vehicles (ROVs). Each of these was equipped with a simple two Degree Of Freedom (DOF) gripper for object retrieval. Each ROV was also upgraded with powerful GTO thrusters prior to deployment for enhanced mobility in high currents. Neurophysiological data was collected with a wireless brainwave (or EEG for ElectroEncephaloGraphy) monitoring device called the X-10 B-Alert system manufactured by Advanced Brain Monitoring Inc. EEG and ECG data were recorded on ruggedized Getac V110 laptops. Video data of participant control activity was captured on 2ea QSEE color cameras mounted on each ROV control console as well as Google Glass headsets worn by participants (as indicated in Fig. 1).

3.3 Measures

This investigation endeavored to examine the influence of a secondary robot's surrogate perspective in terms of the operator's cognitive state and workload. For the purposes of this research, cognitive state was taken to represent how involved the operator is with the task environment whereas workload represented the nature and intensity of effort applied to the task. Given the non-invasive real time advantages of neurophysiological measurement promoted by Parasuraman and Wilson [11] as well as recent findings by Matthews and colleagues that measures of



Fig. 1 Data collection and ROV control setup at sea aboard the R/V Sabina support vessel

cognitive load don't necessarily converge [12], a combination of subjective and EEG based methods were used to examine cognitive state and workload.

Neurophysiological data was collected using the B-Alert system mentioned above. This headset acquires 9 channels of EEG collected along the scalp at standard locations: Fz, Cz, POz, F3, F4, C3, C4, P3, and P4. The proprietary acquisition software used in this process includes artifact decontamination algorithms for eye blink, muscle movement, and environmental/electrical interference such as spikes and saturations. This software also contains a cognitive state and workload classification package that compares EEG collected in field conditions with baseline data acquired while participants perform a series of computer based vigilance tasks lasting from 15–30 min under nominal conditions in a controlled setting.

The first baseline task requires the participant to remain vigilant while choosing between different symbols presented on a laptop display. The second task requires a simple keyed response to a single stimulus (a red circle appearing on the screen) without any choice or decision-making process involved. The third task requires the participant to respond in a similar fashion to the second task, but with an auditory stimulus presented while their eyes were closed.

EEG classification for this experiment was conducted post hoc via comparisons with three baseline profiles and a large EEG database of typical human sleep and distraction patterns. These classifications are presented as second by second probabilities that each epoch (or second) of data matched the participant's Hi (or decision based) engagement state established in the first (3-choice) baseline task, a Lo (or awareness based) engagement state observed during the second (eyes open) baseline task, or a distracted state determined by the third (eyes closed) baseline task. EEG workload classification was conducted in a similar fashion by comparing elements from each participant's 3-choice baseline task with a large data base of typical human performance on forward and backward span tasks. See Berka et al. [13] for more detail on this classification process.

Subjective measures of cognitive state were collected using a popular situational awareness survey validated by Adams and colleagues on behalf of the U.S. Navy called the China Lake Situational Awareness (CLSA) scale [14]. This survey essentially asks the participant to rate their awareness of what was going on around them during recent activity based on a variety of criteria categorized on a 5 point scale from 1 (Very Good) to 5 (Very Poor).

Subjective measurement of workload was measured in a similar fashion using the NASA Task Load Index or TLX—a popular instrument in the human factors field for more than two decades [15]. Both of these instruments were administered to participants immediately upon completion of each ROV training/evaluation trial and field session. Scores were then normalized to a common 100-point system following a higher-as-better paradigm for consolidated graphing with other metrics.

3.4 *Design and Procedure*

Upon completion of informed consent documentation, participants were administered a standard spatial ability battery [16] and spatial orientation survey developed [17]. Evaluations were conducted commensurate with the findings of Blitch and colleagues that these instruments were effective in predicting robot teleoperation effectiveness for semi-autonomous mobile robots engaged in a similar rescue oriented task during the NASA Haughton Mars Project in 2001 [6]. Participants were then outfitted with a B-Alert device and asked to perform the baseline tasks described above before performing 8 training/evaluation trials in an underwater obstacles course set up in a local pool. The top four performers from that session were chosen for deployment to the FIU facility in Florida.

Upon embarkation aboard the research vessel and arrival in the vicinity of the Aquarius Reef Base (ARB), operators were provided with a visual orientation of the area and allowed a single familiarization dive lasting approximately 15 min in order to identify various obstacles and hazards in the area before beginning their rescue mission sessions. Each mission commenced with a simulated radio call (conveyed via a role player on board) from the habitat and rover explaining the nature of the crisis and asking for immediate assistance. The role player then worked through a series of scripted radio calls (with increased urgency and emotion in order to inject stress into the scenario) describing which leaks were progressing at what rate on various gauges, as well as health status of injured crew members, etc.

Each participant completed two missions in a solo or dual condition (counter balanced in sequence and lasting approximately 45 min each) before switching roles with the other operator. The secondary ROV operator was required to simulate a highly autonomous and intelligent robot. As such, the surrogate operator was not allowed to speak to the primary operator at any time, and was only allowed to activate the controls on his ROV in the DUAL condition when the primary operator spoke instructions from a set of simple scripted commands such as descend, ascend, turn right/left, forward/backward, etc.

4 Results

Given that data was collected on the same participants across two temporal phases, initial analysis was performed using a paired t-test with a relaxed alpha set at 0.10 to accommodate the inherent variance involved with data collected in dynamic unstructured environments. The EEG data was averaged across all epochs in each simulation session and analyzed alongside subjective survey data (as illustrated in Figs. 2, 3 and Table 1).

Considering cognitive state first, participants recorded a significantly higher probability of engagement in the “Hi” or “choice” based metric for the SOLO condition than when a teammate with a surrogate perspective was present in the

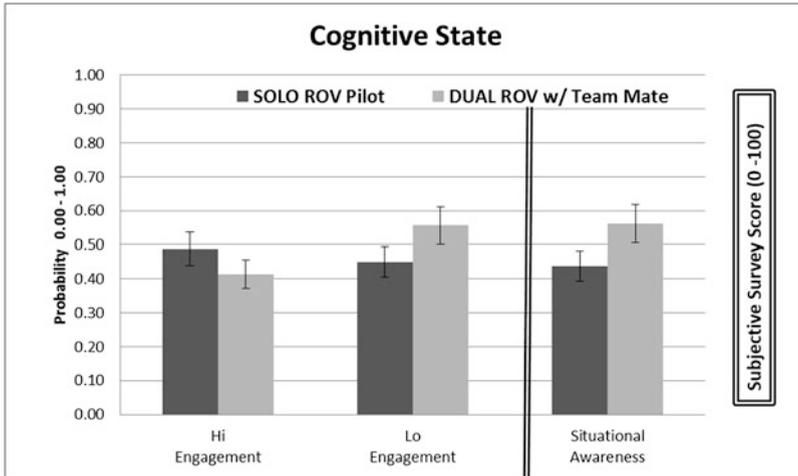


Fig. 2 EEG based engagement and subjective situational awareness (with error bars plotted on 90 % confidence intervals)

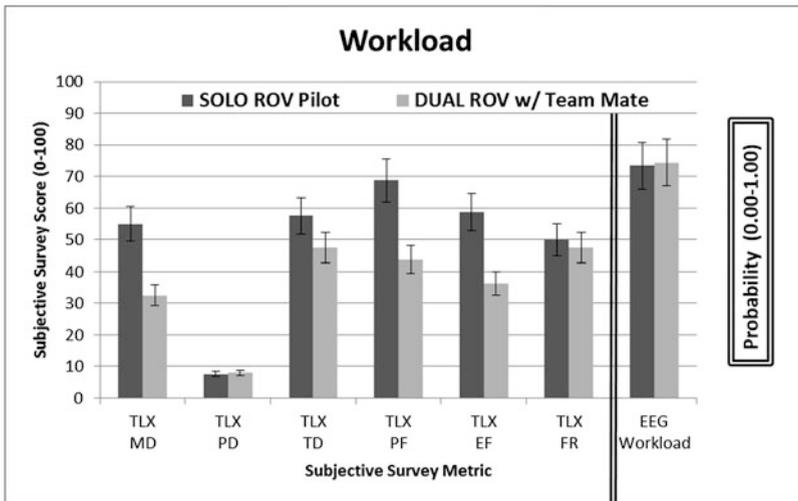


Fig. 3 Workload measured with the NASA TLX and B-Alert EEG system (with error bars plotted on 90 % confidence intervals)

DUAL condition. This data was also accompanied by a large effect size, suggesting a high magnitude of difference between conditions. Although Lo Engagement data fell short of statistical significance, average CLSA scores increased substantially after the surrogate perspective was made available in the DUAL condition suggesting that a beneficial increase in situational awareness was obtained. Analysis of

Table 1 Engagement, awareness and workload statistics

	SOLO		DUAL		<i>t</i> (3)	<i>p</i>	<i>d</i>
	M	SD	M	SD			
EEG Hi Eng	<u>0.488</u>	<u>0.102</u>	<u>0.414</u>	<u>0.122</u>	<u>-1.786</u>	<u>0.086</u>	<u>0.66</u>
EEG Lo Eng	0.450	0.122	0.556	0.152	-1.290	0.144	-0.79
CLSA Score	<u>44</u>	<u>38</u>	<u>56</u>	<u>24</u>	<u>-1.615</u>	<u>0.099</u>	<u>0.70</u>
TLX MD	<u>55</u>	<u>25</u>	<u>33</u>	<u>19</u>	<u>-3.855</u>	<u>0.009</u>	<u>0.94</u>
TLX PD	8	5	8	4	0.961	0.196	-0.09
TLX TD	<u>58</u>	<u>33</u>	<u>48</u>	<u>35</u>	<u>-2.507</u>	<u>0.033</u>	<u>0.31</u>
TLX Pf	<u>69</u>	<u>34</u>	<u>44</u>	<u>27</u>	<u>-3.429</u>	<u>0.013</u>	<u>0.79</u>
TLX Ef	<u>59</u>	<u>35</u>	<u>36</u>	<u>29</u>	<u>-1.546</u>	<u>0.099</u>	<u>0.70</u>
TLX Fr	50	36	48	19	-0.961	0.196	0.07
EEG Wkld	0.734	0.123	0.745	0.107	-1.008	0.194	-0.18

Significant values shown underlined in bold type with a relaxed alpha of 0.10

workload data showed a prominent drop in Mental Demand, Time Demand, Performance, and Effort whenever the surrogate perspective was made available in the DUAL condition. There were no statistically significant differences observed, however, in Physical Demand, Frustration, or EEG based probability of a Hi Workload classification. Unfortunately radical current shifts and subsequent umbilical entanglement challenges made it impractical to score comprehensive mission trials equitably so hypothesis testing of mission effectiveness had to be abandoned.

5 Discussion

The EEG and subjective data presented above support all three of the hypothesis expectations expressed above. TLX scores provide evidence of significantly higher workload recorded in the SOLO condition across four of the six components measured, thereby supporting the notion that a surrogate perspective can substantially moderate the demand for cognitive resources in complex and challenging situations where risk of cognitive overload is prominent. Higher CLSA scores recorded in the DUAL condition also suggest that a substantial benefit was achieved in terms of increased situational awareness provided by the robotic teammate's surrogate perspective.

It is interesting that the EEG based workload metric did not show any substantial difference between conditions. As such, these results are consistent with those of Matthews and colleagues in finding that cognitive load metrics do not always converge across measurement instruments [12]. Given that the workload classifier for this particular EEG device is derived primarily from human performance on high and low digit span tasks, it is possible that it only captures the working

memory component of mental demand and thus leaves other aspects of cognitive load unaddressed.

The EEG based cognitive state metrics provide perhaps the most interesting results of all. The “Hi” or decision based engagement metric decreased significantly when the surrogate perspective was introduced in a trend similar to the subjectively measured workload metric, while the “Lo” or awareness based engagement metric followed an opposite trend in the same direction as the situational awareness scores (albeit with a variance that fell beyond an acceptable level of statistical significance).

6 Conclusions and Future Work

Taken across both neurophysiological and subjective measures, the data collected in this investigation provide a modicum of support for the stated hypothesis on all three counts. Not only did the addition of a surrogate perspective provided by the second ROV result in reduced workload across four of the six TLX components, but it also increased participants’ situational awareness scores in the CLSA data. The EEG based Hi Engagement data indicate a much lower choice intensity was involved in the DUAL condition which presents an important advantage when controlling robots under stress in a critical scenario that could easily be imagined to induce cognitive overload. That said, it must be pointed out that statistical support for this claim is weak in terms of sample size, a relaxed alpha, and the lack of any correction for familywise error. Despite this acknowledged increase in likelihood of committing a Type I error, the relatively large effect sizes observed suggest that replication would be quite worthwhile in a manner consistent with Wickens’ Common Sense Statistics [18].

Although real time cognitive state classifications were not used in this study, the B-Alert system is capable of providing them in near real time (with a 1–5 s delay depending on signal characteristics and display options). Detection of brain state changes in this manner presents a powerful mitigation option in that intervention strategies can be cued and implemented *before* performance suffers from cognitive overload or *before* complacency becomes evident in operator behavior.

It may also be of interest to NASA and others that the magnitude of cognitive workload, engagement, and stress measurements collected during this research suggest that the relatively viscous and occluded underwater environment that submersible ROVs operate in presents a far more daunting challenge to mobile robots than sample collection conducted via “wireless” aerial and ground platforms in terms of current effects, umbilical management, sediment interference, etc. As such, the procedures adopted for even the most realistic and appropriate space exploration analogs like this one may actually turn out to be overly conservative in terms of robot effectiveness. In others words, NASA may be too hard on itself when conducting operational simulations underwater when it comes to mobile robot effectiveness and agility evaluations—especially in consideration of the human factors involved.

Assuming that the SPAARC concept is eventually validated beyond this initial investigation and/or expanded to include air-ground collaborative systems, the implications can be profound for multi-robot control. Not only do these findings provide ample room for optimism regarding resilient and adaptive human robot team performance in stressful, unstructured environments but they also show promise for the pursuit of enhanced trust and confidence between humans and machines championed by Hancock and colleagues [8].

Despite such optimism, an abundance of confounding issues lurk in the shadows. Fatigue, fitness, comfort, emotional valence, and a number of other variables will continue to threaten the validity of future “naturalistic” research just as it did here. It is only with continued ventures beyond the cozy and predictable laboratory environment, however, that we can truly understand the most complicated aspects of human cognition as they adapt to chaotic factors that abound in the “real” world.

Acknowledgments This work was sponsored by the Warfighter Interface Division of the 711th Human Performance Wing at the Air Force Research Laboratory. The author would like to extend an especially warm and profound expression of gratitude to Bill Todd and Jason Poffenberger from NASA/JSC for their outstanding support, as well as Ethan Blackford, Jeff Bolles, and James Christensen for their tremendous prowess in handling complex data collection and participant management issues under daunting conditions and an extremely tight schedule.

References

1. Salamin, P., Thalmann, D., Vexo, F.: The benefits of third-person perspective in virtual and augmented reality?. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology. ACM (2006)
2. Milgram, P., Rastogi, A., Grodski, J.J.: Telerobotic control using augmented reality. In: Robot and Human Communication, Proceedings 4th IEEE International Workshop on RO-MAN'95 TOKYO. IEEE (1995)
3. Okura, F., et al.: Teleoperation of mobile robots by generating augmented free-viewpoint images. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2013)
4. Leeper, A.E., et al.: Strategies for human-in-the-loop robotic grasping. In: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. ACM (2013)
5. Hashimoto, S., et al.: Touchme: an augmented reality based remote robot manipulation. In: Proceedings of ICAT2011 21st International Conference on Artificial Reality and Telexistence. (2011)
6. Blich, J.G., et al.: Correlations of spatial orientation with simulation based robot operator training. In: 4th International Conference on Applied Human Factors and Ergonomics (AHFE). San Francisco CA (2012)
7. Adams, J.A., Kaymaz-Keskinpala, H.: Analysis of perceived workload when using a PDA for mobile robot teleoperation. In: Proceedings IEEE International Conference on Robotics and Automation ICRA'04. IEEE (2004)
8. Hancock, P.A., et al.: A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* **53**(5), 517–527 (2011)

9. Parasuraman, R., Cosenzo, K.A., De Visser, E.: Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Mil. Psychol.* **21**(2), 270 (2009)
10. Scerbo, M.: Adaptive automation. In: Parasuraman, R., Rizzo, M. (eds.) *Neuroergonomics: The Brain at Work*, pp. 239–252. Oxford University Press, 198 Madison Ave, 10016, New York NY. (2007)
11. Parasuraman, R., Wilson, G.F.: Putting the brain to work: neuroergonomics past, present, and future. (Cover story). *Hum. Fact.* **50**(3), 468–474 (2008)
12. Matthews, G., et al.: The psychometrics of mental workload multiple measures are sensitive but divergent. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* **57**(1), 125–143 (2015)
13. Berka, C., et al.: Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *Int. J. Hum-Comp. Inter.* **17**(2), 151–170 (2004)
14. Adams, S., Kane, R., Bates, R.: Validation of the China Lake Situational Awareness Scale with 3D SART and S-CAT. Naval Air Warfare Center Weapons Division (452330D), China Lake, CA (1998)
15. Hart, S.: NASA-task load index (NASA-TLX); 20 years later. *Annu. Meet. Hum. Fact. Ergon. Soc.* **50**(9), 904–908 (2006)
16. Newton, P., Bristoll, H.: Psychometric Success Spatial Ability Practice Test 1, pp. 1–12 (2010)
17. Kozhevnikov, M., Hegarty, M.: A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cogn.* **29**(5), 745–756 (2001)
18. Wickens, C.D.: Statistics. *Ergon. Des. Q. Hum. Fact. Appl.* **6**(4), 18–22 (1998)

A Comparison of Trust Measures in Human–Robot Interaction Scenarios

Theresa T. Kessler, Cintya Larios, Tiffani Walker, Valarie Yerdon and P.A. Hancock

Abstract When studying Human–Robot Interaction (HRI), we often employ measures of trust. Trust is essential in HRI, as inappropriate levels of trust result in misuse, abuse, or disuse of that robot. Some measures of trust specifically target automation, while others specifically target HRI. Although robots are a type of automation, it is unclear which of the broader factors that define automation are shared by robots. However, measurements of trust in automation and trust in robots should theoretically still yield similar results. We examined an HRI scenario using (1) an automation trust scale and (2) a robotic trust scale. Findings indicated conflicting results coming from these respective trust scales. It may well be that these two trust scales examine separate constructs and are therefore not interchangeable. This discord shows us that future evaluations are required to identify scale appropriate context applications for either automation or robotic operations.

Keywords Human–Robot interaction · Trust · Trust scale · Trust measures

1 Introduction

Robots are now pervasive across diverse occupational fields. These range from manufacturing operations, assistive surgical units, search and rescue, space exploration, and unmanned vehicle agents to name but a few [1–5]. Their capabilities have been widely influential in human exploration, across within experimental and applied settings. Parasuraman and Riley reminded us that automation is not merely replacing human labor, but rather, changing the nature of it in which humans work alongside these entities [6]. The overlap across automation and robots is considerable and often difficult to disentangle [7]. To date, robots have primarily been grouped under the umbrella of automation. Thus, automation-related measures have consistently been applied to robot experimentation. However, it is possible

T.T. Kessler (✉) · C. Larios · T. Walker · V. Yerdon · P.A. Hancock
University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816, USA
e-mail: theresakessler@knights.ucf.edu

that erroneous conclusions are being drawn to human–robot scenarios from these automation based metrics. We here evaluate theories and metrics of automation that have and are being applied to robot experimentation. Most especially we compare robot metrics related to trust. Among the presence or absence of a malfunction (i.e., performance reliability), and transparency (level of information) provided to the operator. These factors were tested due to their importance in facilitating effective human–robot interaction [8]. Metrics of trust related to both automation and robots also exist, and comparisons across these are crucial to establish further empirical understanding of their relationship.

2 Trust

Trust has been defined as “*the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability*” [9], and as “*the reliance by an agent that actions prejudicial to their well-being will not be undertaken by influential others*” [10]. However, trust can adopt many different meanings and can be interpreted differently within diverse contexts. Empirically, trust has been noted as an important dimension in preserving relationships among humans and robots. It becomes a critical issue when robots are being assigned critical roles in perilous circumstances in which an operator’s life is at stake, and the success of the mission is contingent upon the cooperation between man and machine. There may also be a heightened ease of trusting robots when the human is physically and cognitively incapacitated [8]. Parasuraman and Riley elaborated on inappropriate levels of trust that can lead to detriment when using automation. Misuse brings about overreliance on the system, thereby, operators may fail to monitor and recognize warning signs when the automation is wrong [6]. Disuse brings about distrust to the system, thus, operators fail to let automation perform at the expense of exhausting their own vigilance and amplifying their task load. Abuse brings about erroneous control of automation and poor human performance. Applying these automation precepts, trust must therefore be appropriately calibrated to match the intentions of the designer and user of the robot [9].

Hancock and colleagues conducted a quantitative meta-analysis examining precursors of trust in human–robot collaboration [8]. They recognized that these factors fall within three respective categories: (a) Robot-related, (b) Human-related, and (c) Environment-related. They found that Robot-related factors were consistently associated with trust in HRI across all measures. More specifically, Robot-related factors were subcategorized into two groups: (1) robot performance-based factors and (2) robot attribute-based factors. Robot performance-based factors were more significantly related to the development of trust and maintenance of that trust. This trend implies that trust can be regulated through the influence of a robot’s performance. Although trust is an important concept to consider, operators sometimes doubt the efficacy of the robot’s behavior and intentions due to difficulties in perceiving its state [11, 12]. With respect to this

concern, our work also focuses on dimensions of robot performance that involve the robot's level of transparency.

3 Transparency

Transparency is the level of information presented to the operator, guiding comprehension about the agent's reasoning process and potential future actions [13]. Transparency has been listed as one of the factors that influences trust development [8]. There have been several notions that pertain to transparency in automation and robot interaction. A more recent theory associated with autonomous agents is the Situation Awareness-based Agent Transparency model (SAT model) [13]. The SAT model consists of three main categories. In SAT Level 1, the operator is able to acknowledge the agent's goals and suggested actions. In SAT Level 2, the operator is provided with information about the agent's reasoning behind its actions. In SAT Level 3, the operator has an awareness the agent's possible success. Should a robot exhibit all three levels or only one based on the nature of the task, the operator will have an adequate understanding of the robot's intentions and will allow the operator to make informed decisions.

The goal of designing a transparent interface is to attain proper use of the system. With transparent interfaces, having knowledge of the system when it is and isn't reliable can assist in shaping the operator trust. Having appropriate trust calibration is linked to high reliability [14]. It is critical, however, to note that higher levels of transparency do not equal higher levels of trust as appropriate levels of information need to be assessed [9, 13]. For example, providing extensive amounts of extraneous information in an interface might overwhelm the operator [15, 16]. Also, should an interface provide an abundance of information about its unreliability, then the operator reacts competently by *not* trusting the robot [13]. Therefore, transparency should engender appropriate levels of trust. Besides transparency, manipulating the behavior of a robot can also facilitate trust. Without meaningful action, transparency fails to sustain an operator's trust.

4 Malfunction

Within automation, various characteristics of performance affect reliability, predictability, and capabilities of the system [13]. Performance reliability has been constantly identified as a primary influence for trust in robots [8]. Therefore, it is critical to understand what kinds of reliability change contributes to the loss of trust. One example would be to study the effects of mechanical malfunctions in a robot to observe whether a user experiences a decrease or increase in trust in the robot. Sometimes, the user might perceive a behavior to be a malfunction when, in fact, it was not. If the user is not given enough information concerning the robot's actions,

then the user may interpret the robot's actions to be an error [17]. Therefore, transparency is important in mitigating these misconceptions. On the other hand, if a user is provided with redundant information, they might interpret the robot's behavior to be erroneous [18]. Transparency must then provide adequate amounts of information to support user trust, but avoid repetitive information. It is crucial to evaluate these constraints among different robotic systems in order to facilitate future robot design. We therefore need qualitative and quantitative assessments in order to gauge the attitude of users. One such method is the use of subjective perception of scales.

5 Convergent Validity of Scales

In the taking of evaluative scales, to ensure that calculation errors remain at a minimum, we must determine whether a scale is accurate and stable [19]. Here, we examine convergent validity, which is "*meant to establish the extent to which the new measure adheres to other related indicators of the construct it is designed to epitomize*" [20]. Convergence in validity is able to help clarify whether the construct has a strong relationship between the measures and it authenticates the scale by providing an adequate assessment criteria of the related construct. Thus, scales with convergent validity are able to translate important theoretical and empirical predictions about the related construct based on results that are collected from the measure.

In trying to formulate a survey related to trust in robots, one would expect to relate to other measures that co-vary with trust in robots. If this covariation does not hold, then (1) the scale is not an adequate measure of the construct, or (2) the scale is valid, but the theory underlying these covariates does not hold, thus, no relationship should be expected [20]. If the relationship holds, then there is a likelihood that this supports the scale's convergent validity. Of relevance here, we review two primary scales that are used in human-robot interaction to measure trust. Their convergent validity is being called into question due to their precise nature in testing for robot-specific and automation-specific trust. Robots have consistently been grouped within automation metrics, but can these metrics adequately test robot trust? If two scales converge upon the same construct, then it stands to reason that they can be used interchangeably. Provided below is an overview of these trust scales.

6 Trust Scales

One of the most significant challenges for successful collaboration between humans and robots is the development of appropriate levels of mutual trust [21, 22]. Due to fast paced technological advancement, the need for understanding trust in human-robot interactions is crucial to future success. Our present study focuses on similarities and scores elicited by two individual trust scales: (1) the Human Robot Trust

Scale [23] and (2) the Trust in Automation Scale [24]. The Trust in Automation Scale and the Human Robot Trust Scale can be seen to have similar properties; however, they also differ in important ways. The Trust in Automation Scale focuses more on the operator, while the Human Robot Trust Scale focuses on the change in an individual's trust over time. Below we describe each scale, providing further comparison.

The Trust in Automation Scale [24] extended from previous trust scales. Jian and colleagues took to an empirical approach rather than a theoretical approach to construct a multidimensional scale in the hopes of identifying various components of trust. The main focus of their scale was to develop a measure that enabled researchers to better predict trust patterns toward automation, based on the assessment of the operators. The scale distinguished between three potentially different types of trust: human-human trust, human-machine trust, and trust in general [24]. Our overall goal was to determine what were the potential similarities and/or differences between these concepts. Jian and colleagues found similar patterns among all three trust relationships that indicated trust is perceived equally across these multiple domains. Therefore, the Trust in Automation Scale provides a baseline model for assessing trust between human and machines in regard to the operator's perception. This scale is thought to be most reliable when it is used to understand trust in automated systems.

Regardless of the domain of application, the environment or task, a human's trust in their non-human collaborator is an essential element required to ensure that any functional relationship will ultimately be effective [23]. Schaefer focused on many of these contributing factors believed to affect trust as well as their design. The Human Robot Trust Scale derives from Scholtz's [25] theory that there are five different types of human roles within HRI. In addition, it reflects Jian's [26] Trust Scale. The Human Robot Trust Scale was created in order to provide a method that would measure an individual's trust towards robots specific to HRI that focused on the antecedents and measurable factors mentioned above. In addition, it measures the changes in trust over time. It acknowledges both the physical form and functional capabilities that determine the classification of the robot which are essential to trust development. The overall goal of the scale is to derive a more sensitive and accurate test score as compared to Jian et al's Human Automation Trust Scale specifically as it relates to human–robot interaction [23].

7 Automation Versus Robotics

When looking at the aspects of trust in HRI, the lines defining a machine as automation or robot need to be clarified. Specifically, the question arises in how we define the terms automation and robotics and where the line of differentiation resides. Sheridan [27] defined levels of automation using a scale from 1 to 10. The scale begins at level one where a machine has no decision making power, and extends to level 10 where the machine makes decisions and takes actions without

conferring the user [and see 28, 29]. These describe levels of autonomy, but not all robots are autonomous, so this would not be ubiquitously applicable in defining a robot apart from automation. In use, automation is known for completing repetitive, structured, and controlled tasks efficiently and effectively [23]. Merriam Webster broadly defines *automation* as controlled operation of an apparatus process, or system by mechanical or electronic devices taking the place of human labor [30]. However, this does not necessarily distinguish between these two terms which one often used interchangeably. Hancock and colleagues [8] highlight the shared fundamental similarities, but recognize a number of inherent differences in human-automation and human-robot trust. These observations provide the foundation for the present comparison.

In order to more appropriately define what a robot is and differences with automation, we look to an early known use of the word robot. One of these examples is the use of the term “*roboti*” [31], which was used to describe autonomous artificial people made from synthetic organic material. By 1923 the word was translated into thirty languages and the term “*robot*” entered the vernacular and psyche of cultures across the globe. This same theme was repeated in *I, Robot*, which added the Three Laws of Robotics: (1) A robot may not injure a human being or, through inaction, allow a human being to come to harm; (2) A robot must obey orders given it by human beings except where such orders would conflict with the First Law; (3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Law [32]. While the Laws of Robotics were the work of science fiction, for now Law Two persists as robots cannot currently program themselves. In addition to this, Siciliano pointed out it is hard to determine how one definition of the term robot could be expansive enough to cover the broad cultural use of the term, especially considering its different applications in research and development [33]. More recently, the word robot has been defined by Wikipedia as “*a mechanical or virtual artificial agent, usually an electro-mechanical machine that is guided by a computer program or electronic circuitry,*” [34] and as a mechanical or virtual artificial agent guided by a computer program or electronic circuitry, [35].

In analyzing the concepts of automation and robotics, a quadrant design was used to categorize technology through intelligence and autonomy [7]. The X axis represented automation and the Y intelligence level. The four quadrants represent functions of high/low automation paired with high/low intelligence. Quadrant IV holds the characteristics of high automation and high intelligence, but specifically describes robots as having lower levels of both characteristics versus artificial intelligence. Thus, they describe robots as existing at a specific point of automation and intelligence which theoretically could be parsed out from other types of automation.

Following the same theme, we investigate the differences in trust scales which pose questions specific to automation (the Trust in Automation Scale) and to robots (the Human Robot Trust Scale). The purpose of this study is to uncover the similarities and differences in trust ratings resulting from the use of these scales in our HRI study.

8 Methods

8.1 Participants and Materials

Study participants consisted of 58 University students (15 males and 43 females) who were ($M = 20.43$, $SD = 5.99$) years old. Participants were recruited through the University participant pool, Sona System. Compensation for participation was provided in the form of extra credit in class. All participation was voluntary.

For the purposes of this experiment, a Lego Mindstorms EV3STORM robot was used. The robot was located in a 32 square foot environment with a large filing cabinet (see Fig. 1). Finally, a desktop computer with internet access was used to deploy questionnaires. Questionnaires consisted of a demographics questionnaire, the Trust in Automation Scale [24], the Negative Attitudes Towards Robots Scale (NARS) [35], as well as the Human Robot Trust Scale [23].

8.2 Design and Procedure

The study design consisted of a 2 (Transparency Level: High vs. Low) \times 2 (Malfunction Presence: Present versus Absent) \times 2 (Autonomy Level: High vs. Low) between participants design with trust as the dependent measure. When participants arrived at the test facility they were greeted and asked to read and sign the informed consent. Next, they completed the demographic and NARS questionnaires. Following this, participants were shown the robot that they would be working with, as well as asked to complete the Trust in Automation Scale [24] and the Human Robot Trust Scale (Pre) [23]. They were informed of their objectives

Fig. 1 Experimental set up from the view of the participant



and trained on their required performance tasks. For this, they were required to assist the robot in completing its mission, which was to drive around the operational space while scanning the area for threats and notifying the participant of its status. Participants were asked to assist the robot if it encountered any problems and was unable to resolve any such issue on its own. The experimenter answered questions any the participants may have had about their duties and the trial began.

In the malfunction condition, the robot incurred an error and either corrected its problem (high autonomy) or required assistance (low autonomy), while providing high levels of information (high LOI) or low levels of information (low LOI) to the participant about its status. In the no malfunction conditions, the same things occurred except that the robot did not incur any errors. The experiment was completed when the robot reached the back of the area it was scanning. Participants were then again asked to complete the Trust in Automation Scale and the Human Robot Trust Scale.

9 Results

Two separate analysis of covariance were performed using a 2 (Malfunction Presence vs. Absence) \times 2 (Transparency Levels) \times 2 (Autonomy Levels) design measuring change in the Trust in Automation Scale and the Human Robot Trust Scale. The covariates consisted of the three subscales of the Negative Attitudes Toward Robots Scale: (i) Negative Attitude toward Situations of Interaction with Robots, (ii) Negative Attitude toward Social Influence of Robots, and (iii) Negative Attitude toward Emotions in Interaction with Robots. No assumptions were found to be violated and the covariates were moderately correlated with one another and to both of the dependent variables.

The first analysis of covariance was conducted using the Trust in Automation Scale as the dependent variable. This demonstrated that difference in trust scores varied significantly with malfunction, $F(1,47) = 7.39$, $p < 0.01$, $\eta_p^2 = 0.136$. In addition, the difference of the scores between pre and posttests of trust also varied significantly with level of information in conjunction with malfunction presence or absence, $F(1, 47) = 4.90$, $p < 0.05$, $\eta_p^2 = 0.094$. Finally, pairwise comparison showed the Trust in Automation Scale difference scores proved to be higher when no malfunction was present ($M = 15.02$) versus when there was a malfunction present ($M = 7.68$) (see Fig. 2). There were no significant effects pertaining to the autonomy level of the robot.

The second analysis of covariance used the Human Robot Trust Scale difference scores. They were found to change significantly with level of information and malfunction, $F(1, 47) = 10.19$, $p < 0.01$, $\eta_p^2 = 0.178$. Additionally, pairwise comparisons showed that trust scores, were higher when the level of information was high with a malfunction present ($M = 5.71$) versus when no malfunction was present ($M = 2.41$). Finally, trust scores were also higher when the level of information was low and there was no malfunction ($M = 7.30$) versus when a malfunction was

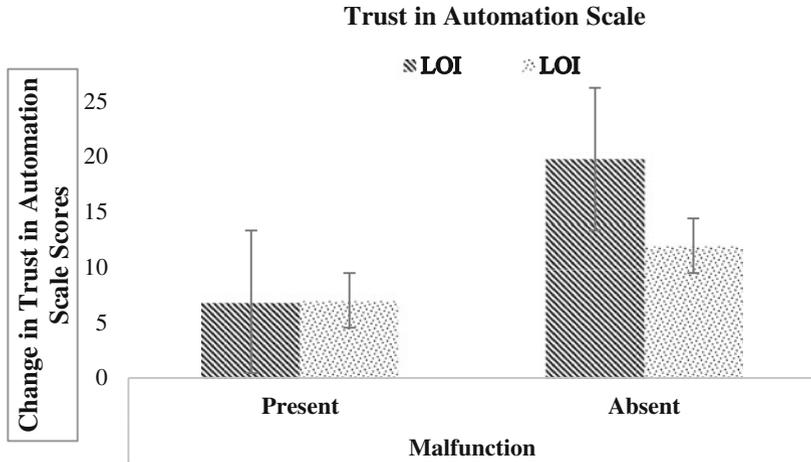


Fig. 2 Change in trust in automation scores for level of information (LOI) and malfunction Presence at the $p < 0.05$ significance level

present ($M = 2.50$). It is interesting to note that the two different trust measures yielding conflicting results. In particular, the interaction of level of information and malfunction presence demonstrated differing results (see Fig. 3). There were no significant differences contingent upon the autonomy level of the robot.

The Human Robot Trust Scale showed larger increases in trust when the level of information provided by the robot was high versus low during a malfunction condition. Additionally, trust was also higher when the level of information provided was low versus high during the no malfunction condition. In contrast with

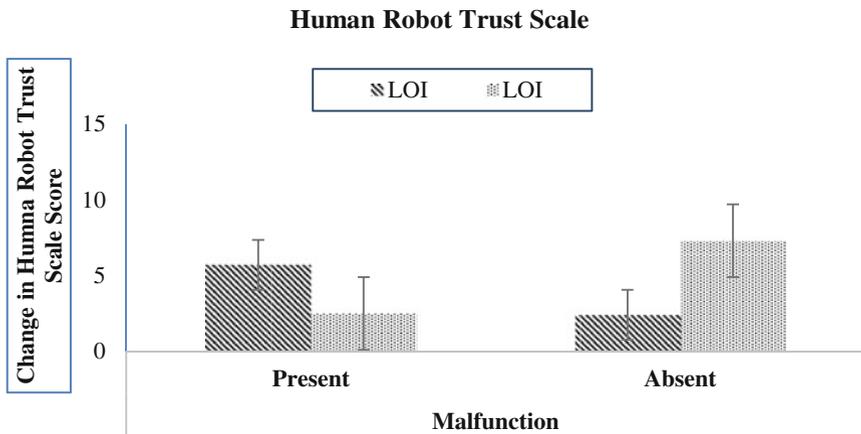


Fig. 3 Change in human robot trust scale scores (%) for level of information (LOI) and malfunction Presence at the $p < 0.01$ significance level

this, the Trust in Automation Scale demonstrated small differences in trust when there was a robot malfunction, despite the level of information provided. However, this scale did show trust levels to be highest when there was no malfunction and high levels of information versus low.

10 Discussion

Our results demonstrate that the Human Robot Trust Scale and the Trust in Automation Scale exhibit differing effects on the manipulations on trust. Results suggest that these scales are measuring constructs that are sufficiently different from one another that the scales cannot be used interchangeably without consequence. This may be in part due to the nature of each scale's questions. For example, the Human Robot Trust Scale features questions based more directly on the behaviors of the robot which are predominately quantitative (e.g. What percent of the time did this robot: act consistently, meet the needs of the mission, and perform as instructed). Conversely, the Trust in Automation Scale features questions which are more qualitative in nature (e.g., Rate the following on a scale of one to seven: the system has integrity, I am familiar with the system, and I am suspicious of the system's intent, action, or outputs). In addition to the differing nature of the questions on the two scales, others have proposed that the concept of robots exist on a continuum as a function of automation and intelligence [7]. Automation rather, is a broad and encompassing construct that can be used to describe all machines and programs, but the definitions of those machines and programs do not effectively describe automation in its entirety. This further supports the notion that different scales should be used to measure trust in robots versus trust in automation as a whole.

Specific to this aim, we recommend the Trust in Automation Scale be used for the broader context of automation that cannot be defined specifically as a robot, whereas the Human Robot Trust Scale should be used in contexts where the user interacts with a definable robot. In order to achieve this, in cases of automation ambiguity, we should use both scales while also collecting data on the participants' perceptions of whether the device is in the category of automation or more specifically robotic in nature.

We recommend that future investigations should consider the evolving nature of the concept of a *robot* in our global culture and perception of the term which is clearly varying across time. Thus, any scale which measures human-robot trust will need on-going re-appraisal in order to assess the validity of its outcomes across place and time.

Acknowledgments The research reported in this document was performed in connection with Contract No. W911NF-10-2-0016 with the U.S. Army Research Laboratory, under UCF, P. A. Hancock, Principal Investigator. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position,

either expressed or implied, of the U.S. Army Research Laboratory or the U.S. government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

References

1. Chen, J.Y., Barnes, M.J.: Supervisory control of multiple robots effects of imperfect automation and individual differences. *Hum. Factors J. Hum. Factors Ergon. Soc.* **54**(2), 157–174 (2012)
2. Heerink, M., Krose, B., Evers, V., Wielinga, B.: Assessing acceptance of assistive social agent technology by older adults: the almere model. *Int. J. Soc. Robots* **2**, 361–375 (2010)
3. Hinds, P.J., Roberts, T.L., Jones, H.: Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Hum-Comput. Interac.* **19**, 151–181 (2004)
4. Parasuraman, R., Cosenzo, K.A., de Visser, E.: Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. *Mil. Psychol.* **21**, 270–297 (2009)
5. Tsui, K.M., Yanco, H.A.: Assistive, Surgical, and Rehabilitation Robots from the Perspective of Medical and Healthcare Professionals. *The AAAI Workshop on Human Implications of Human-Robot Interaction*, pp. 34–39. AAAI Press, Vancouver, Canada (2007)
6. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. *Hum. Factors J. Hum. Factors Ergon. Soc.* **39**(2), 230–253 (1997)
7. Yagoda, R.E., Gillan, D.J.: You want me to trust a robot? The development of a human–robot interaction trust scale. *Int. J. Soc. Robot.* **4**(3), 235–248 (2012)
8. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **53**(5), 517–527 (2011)
9. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors J. Hum. Factors Ergon. Soc.* **46**(1), 50–80 (2004)
10. Hancock, P.A., Billings, D.R., Schaefer, K.E.: Can you trust your robot? *Ergon. Des. Q. Hum. Factors Appl.* **19**(3), 24–29 (2011)
11. Bitan, Y., Meyer, J.: Self-Initiated and respondent actions in a simulated control task. *Ergonomics* **50**(5), 763–788 (2007)
12. Linegang, M., Stoner, H.A., Patterson, M.J., Seppelt, B.D., Hoffman, J.D., Crittendon, Z.B., Lee, J.D.: Human-Automation Collaboration in Dynamic Mission Planning: A Challenge Requiring an Ecological Approach. In: *Proceedings of the 50th Human Factors and Ergonomics Society Annual Meeting; HFES: San Diego, CA*, pp. 2482–2486 (2006)
13. Chen, J.Y., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M. Situation Awareness-Based Agent Transparency (No. ARL-TR-6905). Army Research Laboratory, Aberdeen Proving Ground, MD. Human Research and Engineering Directorate (2014)
14. Ross, J.M., Szalma, J.L., Hancock, P.A., Barnett, J.S., Taylor, G.: The effect of automation reliability on user automation trust and reliance in a search-and-rescue scenario. *Proceed. Hum. Factors Ergon. Soc. Annu. Meet.* **52**(19), 1340–1344 (2008) (SAGE Publications)
15. Cook, M., Smallman, H.: Human factors of the confirmation bias in intelligence analysis: decision support from graphical evidence landscapes. *Hum. Factors* **50**, 745–754 (2008)
16. Neyedli, H.F., Hollands, J.G., Jamieson, G.A.: Beyond identity incorporating system reliability information into an automated combat identification system. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **53**(4), 338–355 (2011)
17. Kim, T., Hinds, P.: Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In: *Proceedings of the 15th IEEE International Symposium on*

- Robot and Human Interactive Communication. ROMAN 2006, pp. 80–85. IEEE, Hertfordshire, United Kingdom (2006)
18. Parasuraman, R., Miller, C.A.: Trust and Etiquette in high-criticality automated systems. *Commun. ACM* **47**(4), 51–55 (2004)
 19. Field, A.: *Discovering Statistics Using IBM SPSS Statistics*, 4th edn. SAGE Publications Ltd., Thousand Oaks, California (2013)
 20. Crano, W.D., Brewer, M.B., Lac, A.: *Principles and Methods of Social Research*, 2nd edn. Psychology Press, Abingdon, England (2002)
 21. Desai, M., Stubbs, K., Steinfeld, A., Yanco, H.: Creating trustworthy robots: lessons and inspirations from automated systems. In: *Proceedings of the AISB Convention, New Front. Hum–Robot. Interac.* (2009)
 22. Groom, V., Nass, C.: Can robots be teammates? Benchmarks Hum–Robot Teams. *Interac. Stud.* **8**(3), 483–500 (2007)
 23. Schaefer, K.E.: *The Perception and Measurement of Human Robot Trust*. (Doctoral Dissertation), University of Central Florida, Orlando, FL (2013)
 24. Jian, J.Y., Bisantz, A.M., Drury, C.G.: Foundations for an Empirically Determined Scale of Trust in Automated Systems. *Int. J. Cogn. Ergon.* **4**(1), 53–71 (2000)
 25. Scholtz, J., Young, J., Drury, J.L., Yanco, H.A.: Evaluation of Human-Robot Interaction Awareness in Search and Rescue. *ICRA'04. IEEE Int. Conf. Robot. Autom.* **3**, 2327–2332 (2004)
 26. Jian, J., Bisantz, A., Drury, C.G., Llinas, J.: Foundations for an Empirically Determined Scale of Trust in Automated Systems (AFRL-HE-WP-TR-2000-0102). Air Force Research Laboratory, Wright-Patterson AFB, OH (1998)
 27. Sheridan, T.B., Verplank, W.L.: *Human and Computer Control of Undersea Teleoperators*. Institute of Technology Cambridge, Man-Machine System Lab, Massachusetts (1978)
 28. Sheridan, T.B.: *Humans and Automation: System Design and Research Issues*. Wiley/Human Factors and Ergonomics Society, New York (2002)
 29. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. *Syst. Man Cybern. Part A: Syst. Hum.* *IEEE Trans. on* **30**(3), 286–297 (2000)
 30. Automation. Merriam-Webster. <http://www.merriam-webster.com/dictionary/automation>
 31. Čapek, K., Selver, P.R.U.R.: *(Rossum's Universal Robots): A fantastic melodrama*. Doubleday, Garden City, N.Y (1923)
 32. Asimov, I.I.: *Robot*. Fawcett Publications, Greenwich, Conn (1950)
 33. Siciliano, B., Oussama, K.: *Springer Handbook of Robotics*. Springer, Berlin (2008)
 34. Robot. Wikipedia. <https://en.wikipedia.org/wiki/Robot>
 35. Nomura, T., Kanda, T., Suzuki, T.: Experimental Investigation into Influence of Negative Attitudes Toward Robots on Human–Robot Interaction. *AI & Soc.* **20**(2), 138–150 (2006)

Human-Robot Interaction: Proximity and Speed—Slowly Back Away from the Robot!

Keith R. MacArthur, Kimberly Stowers and P.A. Hancock

Abstract This experiment was designed to evaluate the effects of proximity and speed of approach on trust in human-robot interaction (HRI). The experimental design used a 2 (Speed) \times 2 (Proximity) mixed factorial design and trust levels were measured by self-report on the Human Robot Trust Scale and the Trust in Automation Scale. Data analyses indicate proximity [$F(2, 146) = 6.842, p < 0.01$, partial $\eta^2 = 0.086$] and speed of approach [$F(2, 146) = 2.885, p = 0.059$, partial $\eta^2 = 0.038$] are significant factors contributing to changes in trust levels.

Keywords Human factors · Human-robot interaction · HRI · Human-robot trust · Proximity · Proxemics · Speed · Psychological experiments · Human robot trust scale · Trust in automation scale

1 Introduction

Human-robot interaction, the branch of science that evaluates how people interact with autonomous robotic entities, is becoming increasingly relevant as robotic platforms are progressively being integrated into many social environments. For example, robots now play a critical role in healthcare, entertainment, and education. As such, we must examine the acceptance and integration of these robotic entities across domains. A critical part of this examination is the consideration and evaluation of key factors such as social norms, relationships, and trust [1].

K.R. MacArthur (✉) · K. Stowers · P.A. Hancock
Department of Psychology, University of Central Florida, 4111 Pictor Ln.,
Orlando, FL 32816, USA
e-mail: keith.macarthur@knights.ucf.edu

K. Stowers
e-mail: kimberly_stowers@knights.ucf.edu

P.A. Hancock
e-mail: peter.hancock@ucf.edu

1.1 Trust in Robots

Trust can be defined in multiple ways. It is most easily defined as the willingness to be vulnerable to another, and has been defined as the belief that others (human or non-humans) will help in uncertain situations [2]. Trust is a critical factor in the acceptance of automated machines or robots among the general public, as it shapes the successful perpetuation and efficacy of collaborative relationships both in human-human relationships and in human-nonhuman relationships [3]. Furthermore, ongoing research indicates that human trust in automated machines may directly and indirectly affect human-machine interaction outcomes such as safety and performance [4]. However, trust is not the exclusive creator of successes or failures in such interactions. Instead, trust interacts with several additional factors as part of the larger human-machine system, making it necessary to examine the many factors that influence trust as a predictor of human-machine performance.

In Hancock et al. [5] identified such additional factors that influence trust in robots including: robot attributes such as proximity of the robot to the human, and human attributes such as degree of comfort with the robot. When considering an attribute such as comfort, it is also prudent to consider what factors affect such comfort. For example, do we feel more comfortable when we and those around us adhere to social norms?

Social norms are the rules of behavior considered acceptable by a group or society in a given context [6]. We argue that adherence to social norms is indeed a predictor of trust in human-robot relationships. Social norms are typically addressed in completely human relationships, but they may also provide useful guidance in human-robot relationships as well. While there are differences when comparing human-human and human-robot interaction, overtime these will diminish and become similarities as technological advances reduce the shortcomings in robotics [7]. Specifically, we aimed to address this gap by examining the influence of two factors deemed important in social norms: proximity and speed of movement. Research concerning the social norms and expectations of human-robot relationship informed our approach.

1.2 Proxemics

One important social norm is the use of interpersonal space, the scientific discipline that addresses this concept is proxemics, which evaluates the unconscious requirements of personal space and crowd density on behavior [8–10]. Hall's [8] seminal research on proxemics classified multiple nonverbal factors that contribute to acceptable interpersonal proximity (e.g., proximity, touching, and eye-contact). Proxemics can be applied to human-robot interaction; however, Hall's non-verbal factors will need to be evaluated based on the degree to which technology has advanced to provide for those factors (i.e., can your robot touch or have eyes?). The exploration of proxemics in human-robot interaction has only become systematic in the last decade [1, 11, 12].

Other areas of research have used proxemics experimentally to evaluate interpersonal space as it relates to their particular field. Virtual reality is one such area, evaluating the interpersonal distance of participants to human-avatars (human-controlled) and agent-avatars (computer-controlled) in a virtual environment [13, 14]. Although findings from the aforementioned research indicate that people respond differently to agents than humans, it creates a foundation for the use of proxemics to empirically evaluate interpersonal distance to a non-human agent.

1.3 Speed of Approach

A human walking down the road, when confronted with an obstacle, will alter their course and speed in avoidance [15]. This adjustment of speed and trajectory, two fundamental aspects of movement, originate as a safety precaution. The development of cities and metropolises have generated a cornucopia of walking spaces, leading to the inclusion of movement into human social norms.

If robotic entities are going to be successfully integrated into human environments they will be expected to conform to the social norms of movement. For example, robots are expected to modify their paths and speed to accommodate people as they approach [16]. High speeds of approach with low interpersonal distance may be tolerated by a human, if the robot signals the intention to pass [17]. However, research also indicates that humans prefer robots to move slower than 1 m per second or just under the average human walking speed [18].

1.4 Hypotheses

Proximity, interpersonal distance, rate of speed, and movement have an effect on human-human systems, and previous research has also evaluated these factors on human non-human systems. Our hypotheses were based on personal theories of human-robot interaction in regards to proximity and speed of approach.

- H₁: participants will report lower trust when the robotic entity moves quickly
- H₂: participants will report lower trust when the robotic entity's proximity is closer.

2 Method

We used a 'Wizard of Oz' paradigm in the present experiment [19]. Participants were under the belief that the robotic entity was fully autonomous and followed pre-programmed instructions whereas it was in fact controlled by a hidden operator.

2.1 Participants

One hundred and forty-eight students (55 males and 93 females), ranging in age from 18 to 31 years ($M = 18.82$, $SD = 1.62$), from the University of Central Florida's undergraduate population participated in the present study. Students were recruited via Sona, the university's anonymous recruitment service, which provides students the opportunity to participate in research. In an effort to mitigate any potential bias from the incentive to participate in research in exchange for course credit, Sona allows students to write an essay in substitution. The University of Central Florida's Institutional Review Board approved the use of human participants in this experiment.

All participants reported to have either normal or corrected-to-normal vision. No participants reported any hearing impairments. One participant reported having post-traumatic stress disorder, and was the only participant to report having military experience. Forty participants reported having interacted with a robot prior to the experiment, fifteen reported having built a robot before, and twenty reported having controlled a robot in the past. The majority of individuals who reported controlling a robot before reported using a videogame or remote control interface (Table 1).

2.2 Design

The present study utilized a 2 (speed) \times 2 (proximity) mixed factorial design, and block randomization to control for primacy effects. To assess trust in the physical presence of the robotic entity, we manipulated the proximity of the robot to the participant as well as the rate of speed at which the robot moved. Dependent measures of trust included scores on the Human Robot Trust Scale [20], the Trust in Automation Scale [21, 22], and the Interpersonal Trust Questionnaire [23]. Additionally, the Negative Attitude Towards Robots Scale was included to evaluate participants' predispositions toward robots [24]. Baseline scores for both the Human-Robot Trust Scale and Trust in Automation Scale were collected utilizing a photograph of the robotic entity prior to any interaction or observation.

Table 1 Participant demographic information

Age	$\bar{x} = 18.82$
Ethnicity	
Caucasian	n = 80
African-American	n = 23
Hispanic	n = 30
Asian	n = 8
Hawaiian/Pacific Isl.	n = 1
Other	n = 5

Fig. 1 Custom-designed and custom-built robot used in the present study

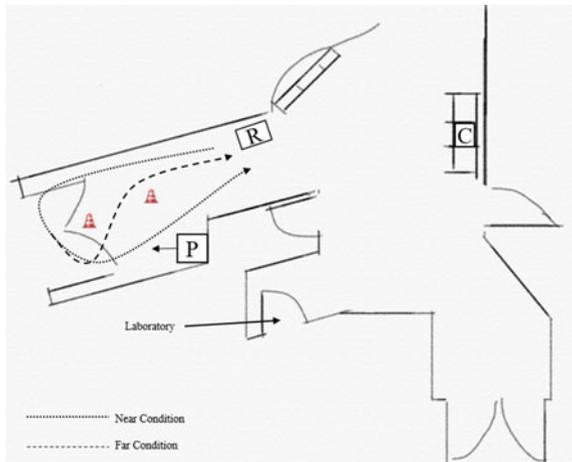


2.3 Apparatuses

A custom crafted robot-like entity was built using a Traxxas Stampede XL-5 series remote controlled truck and a foam-core constructed exterior, painted silver (see Fig. 1). A camera mount was affixed to the top of the base unit, and a dummy web camera attached to the top to give the illusion that the robot had the ability to “see” the environment.

A maintenance hallway, a public area on the second floor of the University of Central Florida’s Orlando campus psychology building, was used as the experimental space (see Fig. 2); allowing for an area sufficient to maneuver the robotic entity and allow a confederate (i.e., robot operator) to blend into the student population.

Fig. 2 Map of the experimental space, delineating the starting position of the robot, the confederate’s vantage point, and the proximal condition tracks



2.4 Procedure

Two researchers were needed to conduct this experiment. One researcher to guide and instruct the participant, and the other to surreptitiously control the robotic entity so as to afford it the necessary simulacrum of autonomy. When participants signed up for the experiment they were instructed to meet at the first floor entrance to the research laboratory. Participants were greeted by one of the researchers and subsequently escorted, via a stairwell within the laboratory, to the second floor. The second researcher, pretending to be a student studying, was pre-positioned on a bench in the hallway at location C in Fig. 2, prior to the participant's arrival. Participants were only in the hallway during the practice trial and experimental conditions; informed consent and administration of surveys were conducted within the laboratory. The informed consent document was given to the participant for their review, and both researcher and participant were required to sign it. Once consent was granted, the participant was instructed to complete the demographics portion of the survey which was comprised of the Interpersonal Trust Questionnaire [23], the Negative Attitude Towards Robot Scale [24], the Human-Robot Trust Scale [20], and the Trust in Automation Scale [21, 22]. Participants were then escorted into the experimental space, and seated in a classroom chair located at P in Fig. 2. Participants first observed a sample trial to acclimatize them to the robotic entity. Immediately after the sample trial, the participant observed the first experimental condition. After each of the experimental conditions participants were escorted back to the laboratory and completed the Human Robot Trust Scale and Trust in Automation Scale in reference to the trial they had just observed. Upon completing all of the experimental conditions, participants were thanked for their time and dismissed.

3 Results

Surveys were scored according to common practices and a mixed-design ANOVA using a within-subject factor of condition (quick-near, quick-far, slow-near, slow-far) was conducted. Analyses indicate proximity as a significant factor contributing to changes in trust scores [$F(2, 146) = 6.842$, $p < 0.01$, partial $\eta^2 = 0.086$]. These findings were observed in both the Human-Robot Trust Scale ($F = 13.212$, $p < 0.001$, partial $\eta^2 = 0.082$) and the Trust in Automation Scale ($F = 11.099$, $p = 0.001$, partial $\eta^2 = 0.07$). Our hypotheses were thus supported. A Bonferroni corrected pairwise comparison of proximity (near-far) presented a mean difference of -4.73 in the Human Robot Trust Scale, and a mean difference of -0.19 in the Trust in Automation Scale. This indicates participants rated higher trust scores in conditions where the robot was further from them, supporting our second hypothesis.

The speed of approach was also a significant factor contributing to changes in trust scores [$F(2, 146) = 2.885, p = 0.059, \text{partial } \eta^2 = 0.038$]. A Bonferroni corrected pairwise comparison of speed of approach (fast-slow) presented a mean difference of -0.702 in the Human Robot Trust Scale, and a mean difference of -0.12 in the Trust in Automation Scale. Support for the effects of speed on trust were found only in the Trust in Automation Scale ($F = 4.149, p = 0.043, \text{partial } \eta^2 = 0.027$) and not the Human Robot Trust Scale ($F = 0.572, p = 0.451, \text{partial } \eta^2 = 0.004$). This indicates participants rated higher trust scores on the Trust in Automation Scale in conditions where the robot moved slower, supporting our first hypothesis. There was no significant interaction between speed of approach and proximity [$F(2, 146) = 1.654, p = 0.195, \text{partial } \eta^2 = 0.022$].

The Negative Attitude Towards Robots Scale is comprised of three subscales measuring negative attitudes toward: situations of interaction with robots ($M = 15.32, SD = 3.98$), social influence of robots ($M = 15.18, SD = 3.6$), and emotions in interaction with robots ($M = 8.66, SD = 2.3$).

The Interpersonal Trust Questionnaire contains three subscales: Fear of disclosure ($M = 86.17, SD = 16.15$), Social coping ($M = 24.31, SD = 6.68$), and Social intimacy ($M = 13.76, SD = 3.58$).

4 Discussion

The present experiment was designed to evaluate the effects of a robotic entity's proximity and speed of approach on the levels of trust held by humans for robotic entities as measured by the Human Robot Trust Scale and the Trust in Automation Scale. Results support both hypotheses that proximity and speed of approach have a significant effect on trust levels. The difference in support of the two trust measures for the speed of approach factor may be related to the difference in the intended purposes of the two trust measures.

Our results add to the foundation of current literature in the area of proxemics and movement as they apply to field of non-human or human-robot interaction. Studies of this nature are critical for the further integration of robotic entities into the everyday lives of humans. Understanding how and why trust changes in response to a robotic entity's physical presence, will ultimately decide if and when robotic entities are widely accepted. Proximity and speed of approach are only two of many social norms that roboticists will have to consider while designing the physical structure of robotic entities and developing the software intelligence.

Limitations of our study and its experimental design include the following issues: (a) robot autonomy, (b) robot morphology and materials, and (c) participant pool. For the purpose of our research, we defined a robot as a machine that can perform the work of a human and work autonomously or controlled by pre-existing computer programming. Our use of the 'Wizard of Oz' paradigm precluded the robot-like entity in our experiment from meeting this definition of a robot which led to complications. During the protocol, students, staff, or faculty passing through the

hallway would stop and ask the confederate researcher about the experiment thus potentially divulging their researcher role to the participant sabotaging the illusion of robotic autonomy. In these situations, the participant continued the experiment to completion and would be politely asked if they become aware of the confederate's control of the robotic entity. Those participants who answered in the affirmative were excluded from the data set, as they were aware that the robot was not autonomous.

As mentioned previously, the robot-like entity was built from a Traxxas (Traxxas, Plano, TX, United States) Stampede XL-5 remote controlled monster truck with a maximum speed of 25 miles per hour. The exterior was constructed from foam-core poster board, hot glue, and silver spray paint. A Logitech C525 (Logitech International S.A., Romanel-sur-Morges, Switzerland) web camera was affixed to the top of the vertical extension to simulate a sensor by which the robot could detect the environment. Subjective testimony from colleagues had mixed responses regarding the physical appearance of our robot-like entity. Some Opinions were split as to how realistic a portrayal of a robot the device was. Future research may want to consider the construction of their robot and its conformity to participant expectations, as the morphology and material composition may be a confounding factor.

The range of participants spans from 18 to 31 years of age, however, the mean age was only 18.82 years indicating that the majority of participants were quite young. The lack of middle aged and senior participants reduces the generalizability of the findings to the general population. Such difficulty in generalizing these findings may delay the acceptance and integration of robotic entities into the daily lives of humans.

Future research in the field of human-robot interaction should consider proximity as a factor that contributes to trust levels; and speed of motion, should the robot have mobility, as a possible contributing factor. Further research needs to be conducted on a robotic entity's speed of approach to determine, to what degree, it effects trust levels.

Acknowledgments We would like to thank Gabriella M. Hancock and Theresa T. Kessler for their assistance in this article. The research reported in this document was performed in connection with Contract No. W911NF-10-2-0016 with the U.S. Army Research Laboratory, under UCF, P.A. Hancock, Principal Investigator. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory or the U.S. government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

References

1. Mumm, J. Mutlu, B.: Human-robot proxemics: physical and psychological distancing in human-robot interaction: In: 6th ACM/IEEE International Conference on Human-Robot Interaction, pp. 331–338 (2011)
2. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* **46**(1), 50–80 (2004)
3. Billings, D.R., Schaefer, K.E., Chen, J.Y., Kocsis, V., Barrera, M., Cook, J., Ferrer, M., Hancock, P.A.: Human-animal trust as an analog for human-robot trust: a review of current evidence (No. ARL-TR-5949). Army Research Laboratory (2012), <http://www.arl.army.mil/arlreports/2012/ARL-TR-5949.pdf>
4. Stowers, K., Oglesby, J., Leyva, K., Iwig, C., Shimono, M., Hughes, A., Salas, E.: A framework to guide the assessment of human-machine systems. *Hum. Factors*, submitted
5. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., de Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* **53**(5), 517–527 (2011)
6. Scott, J., Marshall, G.: *A Dictionary of Sociology*. Oxford University Press, USA (2009)
7. Krämer, N.C., von der Pütten, A., Eimler, S.: Human-agent and human-robot interaction theory: similarities to and differences from human-human interactions. *Hum-Comput. Interact.* **396**, 215–240 (2012)
8. Hall, E.T.: A system for the notation of proxemic behavior. *Am. Anthropol.* **65**(5), 1003–1026 (1963)
9. Russell, J.A., Ward, L.M.: Environmental psychology. *Annu. Rev. Psychol.* **33**(1), 651–689
10. Balgooyen, T.J.: A group exercise in personal space. *Small Group Behav.* **15**(4), 553–563 (1984)
11. Takayama, L., Pantofaru, C.: Influences on proxemic behaviors in human-robot interaction. In: 22nd IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 2009, pp. 5495–5502. St. Louis, MO (2009)
12. Walters, M.L., Dautenhahn, K., te Boekhorst, R., Koay, K.L., Woods, S., Nehaniv, C., Lee, D., Werry, I.: The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In: 14th International Workshop on Robots and Human Interactive Communication. RO-MAN, pp. 347–352 (2005)
13. Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M.: Interpersonal distance in immersive virtual environments. *Pers. Soc. Psychol. B.* **29**(7), 819–833 (2003)
14. Blascovich, J., Loomis, J., Beall, A.C., Swinith, C.L., Bailenson, H., Bailenson, J.N.: Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **13**(2), 103–124 (2002)
15. Olivier, A.H., Marin, A., Crétual, A., Pettré, J.: Minimal predicted distance: a common metric for collision avoidance during pairwise interactions between walkers. *Gait Posture* **36**(3), 399–404 (2012)
16. Wiltshire, T.J., Lobato, E.J., Wedell, A.V., Huang, W., Axelrod, B., Fiore, S.M.: Effects of Robot Gaze and Proxemic behavior on perceived social presence during a hallway navigation scenario. *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **57**(1), 1273–1277 (2013)
17. Pacchierotti, E., Christensen, H.I., Jensfelt, P.: Evaluations of distance for passage for a social robot. In: 15th Annual IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06), pp. 315–320. IEEE Press, New York (2006)
18. Butler, J.T., Agah, A.: Psychological effects of behavior patterns of a mobile personal robot. *Auton. Robot* **10**(2), 185–202 (2001)
19. Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications. *ACM T. Inform. Syst.* **2**(1), 26–41 (1984)
20. Schaefer, K.E.: Perception and measurement of human-robot trust. Doctoral dissertation. University of Central Florida Orlando, FL. http://etd.fcla.edu/CF/CFE0004931/Schaefer_Kristin_E_201308_PhD.pdf

21. Jian, J.Y., Bisantz, A.M., Drury, C.G., Llinas, J.: Foundations for an empirically determined scale of trust in automated systems (No. CMIF198). Air Force Research Laboratory (1998). <http://www.dtic.mil/get-tr-doc/pdf?AD=ADA395339>
22. Jian, J.Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* **4**(1), 53–71 (2000)
23. Forbes, A., Roger, D.: Stress, social support and fear of disclosure. *Brith. J. Health. Psych.* **4**(2), 165–179 (1999)
24. Nomura, T., Suzuki, T., Kanda, T., Kato, K.: Measurement of negative attitudes towards robots. *Interact. Stud.* **7**(3), 437–454 (2005)

Human Factors Issues for the Design of a Cobotic System

Théo Moulières-Seban, David Bitonneau, Jean-Marc Salotti, Jean-François Thibault and Bernard Claverie

Abstract We present a new approach for the design of cobotic systems. It is based on several steps with increasing complexity: Activity analysis, basic design, detailed design and realization. A particular attention is paid to human factors and human systems interactions. Different simulation levels are required to provide flexibility and adaptability.

Keywords Human factors · Cobotics · Ergonomics · Robotics · Cognitics · Workstation design

1 Introduction

Cobot is a neologism formed by the “collaborative” and “robot” terms. It has been used for the first time in 1999 by Peshkin and Colgate to conceptualize the direct interaction between a robot and a human on a dedicated workstation [1]. Its meaning evolved towards different definitions depending on the context of the application [2]. In the present study, a cobot is defined as a robot that has been designed and built to collaborate with humans. A workstation in which a robot and

T. Moulières-Seban · D. Bitonneau · J.-F. Thibault
Safran SAFRAN, 46 rue Camille Desmoulins, 92130 Issy-les-Moulineaux, France
e-mail: theo.moulieres-seban@u-bordeaux.fr

D. Bitonneau
e-mail: david.bitonneau@u-bordeaux.fr

J.-F. Thibault
e-mail: jean-francois.thibault@safran.fr

J.-M. Salotti (✉) · B. Claverie
IMS Laboratory, ENSC, 109 Av. Roul, 33400 Talence, France
e-mail: jean-marc.salotti@ensc.fr

B. Claverie
e-mail: bernard.claverie@ensc.fr

a human are collaborating is called a cobotic system. Cobotics is defined by the science and methods of designing, building, studying and evaluating cobotic systems.

A robot may have typical mechanical and hardware components for a possible collaboration with humans but if it is used in full autonomy, it is considered that it is not part of a cobotic system even if it can be called a cobot. Conversely, a standard industrial robot collaborating with an operator (by remote control for instance), is considered the part of a cobotic system.

This paper firstly presents a characterization of cobotic systems, then proposes a methodological approach to introduce cobotic systems on workstations. A use case of a cobotic workstation design at Safran illustrates this approach.

2 Cobotic Systems

Characterization of cobotic systems is very important for industry in order to understand the feasibility, the efficiency and the relevance of designing and implementing a new cobotic system for an industrial application.

A cobotic system includes a robot and a human collaborating in synergy to perform a task in the context of a workstation. In order to characterize a cobotic system, it is necessary to pay attention to the human operator, to the task, to human system interactions and to the robot. Several humans and several robots may be involved in a cobotic system but for the sake of simplicity, we will focus here on a simple cobotic system that involves a unique robot and a unique human operator.

2.1 Task Characterization

A task is defined by numerous variables [3]. The first one is the domain of application (industrial, domestic, medical, military, etc.). The proposed study is restricted to the industrial domain. Examples of tasks considered in this study are transporting, moving and carrying objects, assembling, surface processing, welding, cutting engraving, etc. The task can also be described by its variability and its necessary adaptation to new applications. Another important variable is the possible impact of a dysfunction or damage on the whole production process [4]. If there is an important risk of failure or a risk to human health, the use of a cobotic system might not be appropriate.

2.2 *Role of Operator*

In the past, only experts in robotics were using robots. Nowadays, more and more people are used to robots and it sometimes happens that newcomers have to interact with industrial robots without training. However, knowledge and know how greatly influence our perception and representation of robots, and our understanding of what they can do and what they cannot. It is of primary importance for the industry to design robots that anyone can easily work and interact with after very short training periods. The complexity of the interaction mainly depends on the role of the person at the workstation [5]:

- Operator: He pilots the robot (locally or remotely). The robot usually has a weak autonomy or even no autonomy at all.
- Coworker: He works with the robot on the same object.
- Supervisor: He provides instructions and checks the work of the robot.
- Bystander: He is present in the working zone of the robot without interaction. There is, however, a preliminary risk assessment to make sure that there is no risk with the current task.
- Maintenance operator: He checks and eventually updates mechanical parts, hardware or software components.
- Designer/programmer: Expert in robotics, he designs, builds or develops software tools and advanced behaviors for the robot.

An important characteristic of the human role concerns the decision process. It can be the result of a common planning, an order, a consensus between the cobot and the human, or an autonomous decision. Parasuraman and Sheridan propose 10 levels for the decision process, ranging from full assistance to no assistance at all [6].

2.3 *Human System Interactions*

The design of a cobotic system involves a clear understanding of the possible human robot interactions, both needs, both constraints and the type of robotic system [2].

The proximity between the operator and the robot is a crucial parameter for obvious security reasons. Ergonomic reasons must also be taken into account. The robot can be in contact with the operator (comanipulation for instance), nearby, or very far. Sometimes, the robot can be carried by the user (exoskeleton) or the user can be carried by the robot (robotic vehicle) [7]. Interactions may occur in real time with immediate feedback or be differed. In addition, the interaction can be brief, e.g., pushing a button, or continuous (comanipulation). Yanco and Drury propose to characterize the cobotic system by the type of interaction and the type of interface [4]. The sensor used for the interaction has an important impact on the abstraction of the message that is exchanged between the operator and the robot.

The operator can remotely interact with the robot by several means:

- Physically: button, joystick, mouse, handling a robot or end effector replica.
- Using touch-sensitive surfaces: screen or simple touch-sensitive surfaces.
- Visually (information for visual feedback): screen, glasses (virtual or augmented reality), by distance measurements.
- Using motion capture: eyetracking, fingertracking, arm motion tracking, or full body motion tracking.
- Soundly: voice recognition, alarm, oral communication.

In artificial intelligence, computer vision and speech recognition techniques allow high level interactions. However, in industrial applications, the complexity and robustness of these techniques are still considered not appropriate. Object recognition by humans is typically more efficient than computer vision techniques. For that reason, efficient cobotic systems are often made of a robotic manipulator that is directly operated by a person, who is in charge of the perception of the environment.

2.4 *Classification of Robots*

The traditional classification of robots is based on their morphology, which usually allows a visual and functional representation of their use:

- Robotic arm: Made of a serial kinematic chain.
- Parallel robot: Robot with ending components linked to the base by several independent kinematic chains.
- Cartesian robot: Robot with prismatic articulations in which axes are located according to Cartesian coordinates.
- Mobile robot: Unmanned vehicles.
- Exoskeleton: Robot worn by a human to improve its performance or mitigate his handicap.
- Hybrid robot: Combination of the above morphologies.

There are other classification methods [8, 9]. One of them is based on the “intelligence” level of the robot, as it is proposed by the American Robotic Industries Association and the JIRA (Japan Industrial Robot Association). The basic robot is an open loop command system and the most sophisticated is able to elaborate a complex planning process. Another classification has been proposed by Coiffet, see Table 1. It is an interesting approach that takes the environment and humans into account. However, there is no reference to the morphology of the robot.

Table 1 Robots classification

Entity	Human	Control system	Robot	Environment
Features affecting performances during the execution of task	Continuous action	Open loop	Fixed	Known
	Intermittent action	Regulation	Mobile	Partially known
	No action	Regulation and reflex		Unknown
		Regulation, reflex and decision		

2.5 Scheme to Describe a Cobotic System

We propose a characterization scheme. It is based on the information flow among the three components of a cobotic system: the environment of the workstation, the human and the robot. The generic scheme is presented in Fig. 1.

Interestingly, different cobotic systems have in general different schemes representing the information flow. Figures 2 and 3 are two representative examples of the differences among the cobotic systems:

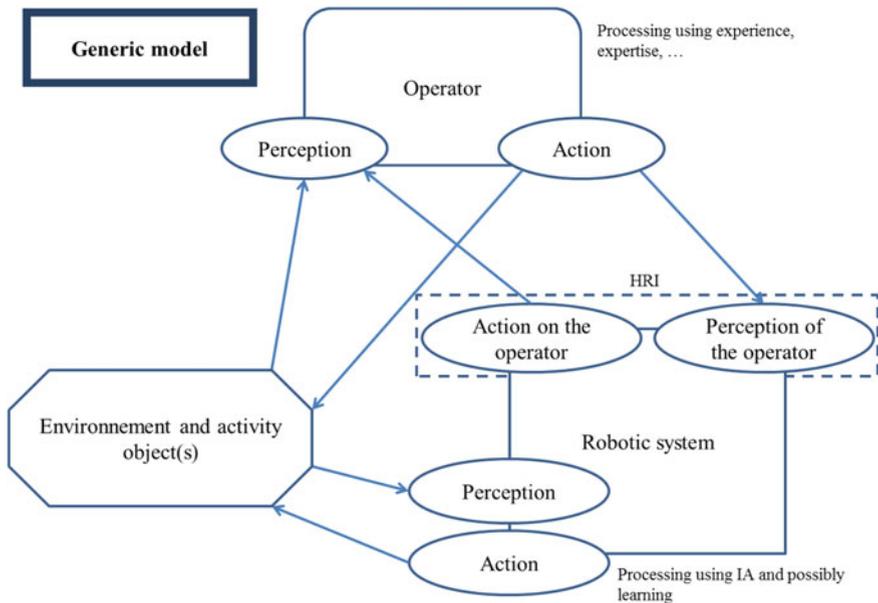


Fig. 1 Standard scheme of a cobotic system

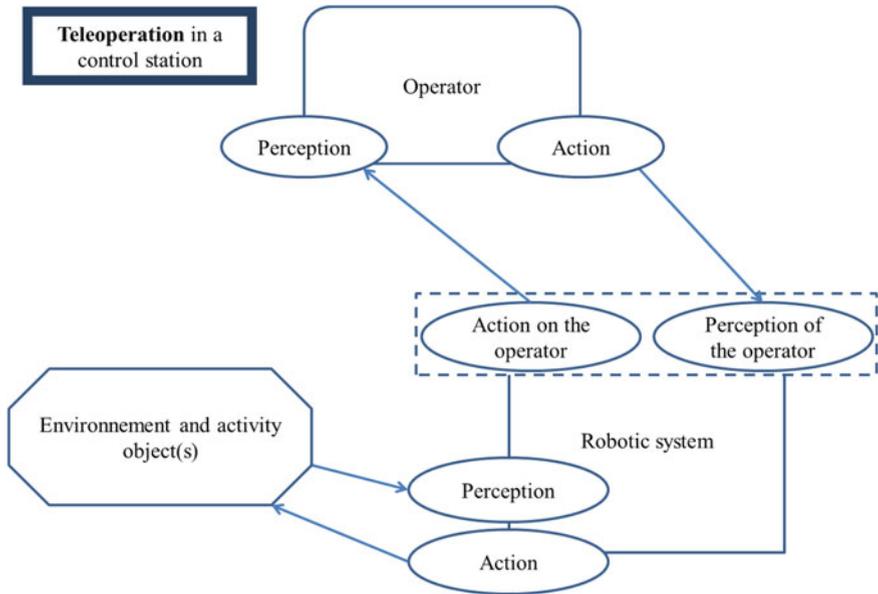


Fig. 2 Scheme of a teleoperating system

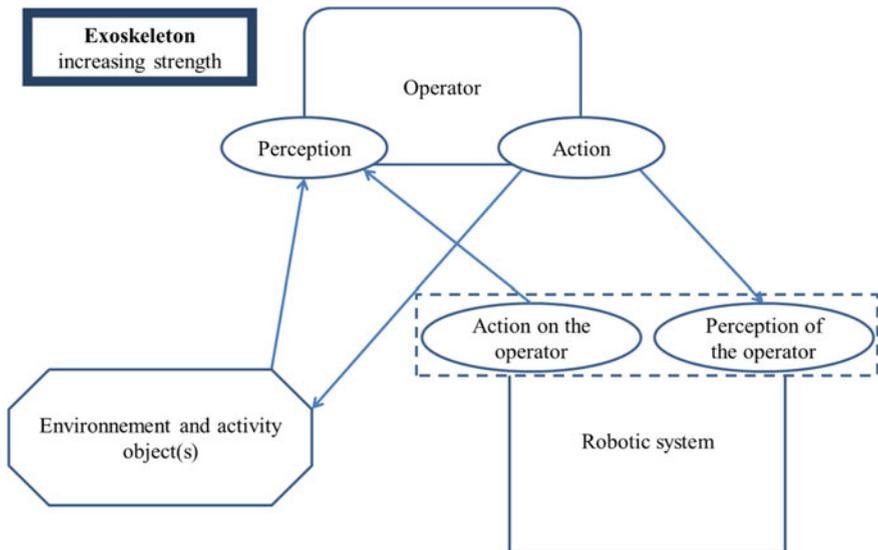


Fig. 3 Scheme of an exoskeleton system

- For a remotely controlled system, the flow of information between the environment of the task and the operator systematically goes through the robotic system, the operator does not interact directly with the environment;
- In the last example, an exoskeleton assists the operator without any interaction with the environment.

A given scheme describing an information flow does not always match with a unique type of cobotic system. Another important parameter is the abstraction level of information. If it is simple (data) or complex (object identification by vision for instance) the scheme might be the same but the role of each component might be completely different. A complementary idea is to use different types of links to provide the abstraction level.

3 Practical Case and Methodological Approach

A dedicated human centered design approach is proposed to determine the functional specifications of a cobotic system. The method is currently implemented within a Safran cobotic project (tank cleaning) with the collaboration of researchers from the Cognitics and Human Engineering team of the IMS laboratory and Ecole Nationale Supérieure de Cognitique. The previous cobotic systems' characterization is a valuable tool that enables a classification of all cobotic solutions in order to match the requirements of the workstation.

Let us consider an application. Nowadays, cleaning viscous and sticky chemical product off huge propellant tanks is carried out manually. An operator scrubs the tank using simple tools (kinds of spade). For a long time, operating at this workstation has been an issue because it is hard, tiresome and performed in a hazardous environment. As the task is long, complex and variable, a full automation is considered very difficult. The current objective is to design a cobotic system for that task. The idea is to minimize the presence of the operator at the station to reduce operational risks and improve working conditions to preserve operator's health.

3.1 *Task, Environment and Context Analysis*

The first step is the analysis of the current activity: the task, the environment and the context. A preliminary work is the study of other similar projects eventually with existing solutions.

The main work consists in interviewing the operators, their manager, anyone that is involved in the project. Then, it is crucial to observe the accomplishment of the task itself. The objective of this analysis is to explicit how the task is really performed, and the reason why it is performed this way to understand the workstation stakes. As many variables (concerning the product, the environment, the tools, the

communication among the operators, etc.) as possible have to be identified. In order to assist the operator with a cobotic system, it is important to identify his skills and experience, and the phases for which he has no or few expertise.

The output of this step is a document including the detailed functional specifications of the system. This document allows the first exchanges with experts in automation, the proposal of possible solutions (only the basic principles) and possible suppliers. If the technology readiness level (TRL) of the proposed solution is too low, the feasibility has to be checked by means of technical tests.

3.2 Example: Safran's Cobotic Project

The cleaning tank analysis led to the different products and their state, the different tank's dirtiness, and the different techniques of the operator to clean the tank. Two solution's principles were considered: robotic scrubbing and hydrogommage. The tests revealed that the hydrogommage, which had a low TRL, was a heavy going process, inappropriate to the shape of the tank. The robotic scrubbing has been chosen. Testing it permitted to decide its size and strength.

3.3 Basic Design

Here, the solution is to design and realize scenarios and mockups. At first, they cannot be accurate, but once implemented, they can be corrected and improved again and again, until they are validated by everyone involved in the project.

The mockup should present the best tradeoff between the time it takes to develop it and update possible solutions, and the distance to the industrial process. Virtual reality tools can be used [10]. Two benefits are expected with the mockup: it allows anticipating errors by testing design hypothesis, and it helps the operators, the designers and the decision-makers to share the same representation of the future workstation. The second point is important because it improves the acceptability of the system. An operator may indeed be afraid of losing his job.

The outputs of the basic design are the specifications for a prototype.

The mockup also allows increasing the TRL of the proposed systems and sub-systems to level 5 or 6.

3.4 Example: Safran's Cobotic Project

In the cleaning tank project, we realized an interactive mockup of a teleoperation system. We simulated a workstation with three camera views and a data screen. Along the half-automated cycle, we could test different ways to interact with the



Fig. 4 Picture and screenshot of the mockup realized for the cleaning tank project

simulated robot: a joystick, a haptic device, etc. (Fig. 4). Several operators tried the simulator and we are currently validating the mockup.

3.5 Detailed Design

The detailed design is the logic next step after the basic design: the principle is to design the solution using a prototype (temporary version of the final system). A first imprecise design is done, then tested. After that, corrections and improvements can be done and tested again, etc., until everyone involved in the project validates it. Thus, the operators can try their future way of performing their task, and help improving it. Simulating these interactions is decisive when designing a cobotic system.

The TRL increases to level 7 or 8.

The outputs of this step are the technical specifications for the final system, coming to a solicitation of offers from suppliers.

3.6 Example: Safran’s Cobotic Project

When the base design is finished, we plan to realize a prototype at supplier’s test facility and experiment with it. The results of the experiments will lead us to the technical specifications of the cobotic system.

3.7 Production and Adjustment

Once the suppliers have answered the solicitation, at this step, the best one is selected according to several criteria (quality of the solution, cost, experience, etc.), and realizes the cobotic system, followed up by the project team. Then, the supplier and possibly the concerned department of the enterprise, install, adjust and

Table 2 Overview of the methodological approach

I. Activity analysis
1. Observations—Interviews—Debriefings
2. Functional specifications
3. Solution principle
4. If low TRL, technical test, if failure go to
II. Basic design
1. Exchange with experts
2. Design and realization of the mockup
3. Test with the mockup, if corrections or enhancement needed go to 2
4. Validation and specifications for a prototype
III. Detailed design
1. Choice of a supplier response for the prototype
2. Design and realization of the prototype
3. Test the prototype, if corrections or enhancement needed go to 2
4. Validation and technical specifications for the final cobotic system
IV. Realization, setup, validation and putting into service

validate it before it is put in service. The project team has to make sure that the cobotic workstation is adapted to all the operators and the production cycle, and that the operators are trained enough to work with the system (Table 2).

4 Conclusion

The proposed methodological approach has to be carried out and achieved with the current use case and then tested on other use cases to validate, correct and complete it.

This approach is global and can be adapted to most cobotic situations. A specific emphasis is placed on the analysis of human/robot interactions. Different elements have to be considered depending on the exact interaction scheme. Several experts from three different disciplines are involved in the project: ergonomics for the analysis of the workstation the variability of the tasks, cognitive engineering to design the human robot interactions and robotics for the robot itself. This multi-disciplinary aspect is a source of wealth for the project.

References

1. Peshkin, M., Colgate, J.: Cobots. *Ind. Robot: Int. J.* **26**(5), 335–341 (1999)
2. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a survey. *Found. Trends® Hum.-Comput. Interact.* **1**(3), 203–275 (2007)
3. Barcellini, F., Van Belleghem, L., Daniellou, F.: Design projects as opportunities for the development of activities. In: Falzon, P. (ed.) *Constructive Ergonomics*, pp. 150–163. CRC Press, Boca Raton, FL, USA (2014)

4. Yanco, H.A., Drury, J.: Classifying human-robot interaction: an updated taxonomy (2004)
5. Parasuraman, R., Sheridan, T.B.: A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cyber. Part A: Syst. Hum.* **30**(3) (2000)
6. Scholtz, J.: Theory and evaluation of human robot interactions. In *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)* (2003)
7. Walther, S., Guhl, T.: Classification of physical human-robot interaction scenarios to identify relevant requirements. *Conf. ISR Robotik* (2014)
8. Alami, R., Chatila, S., Fleury, M., Ghallab, M., Ingrand, F.: An architecture for autonomy. *Int. J. Robot. Res.* **17**, 315–337 (1998)
9. Angeles, J., Park, F.C.: Performance evaluation and design criteria. In: Siciliano, B. (ed.) *Springer Handbook of Robotics*, pp. 229–244. Springer, New York (2007)
10. Maurice, P.: Virtual ergonomics for the design of collaborative robots (Thèse de doctorat). Université Pierre et Marie Curie, Paris IV (2015)

A Natural Interaction Interface for UAVs Using Intuitive Gesture Recognition

Meghan Chandarana, Anna Trujillo, Kenji Shimada
and B. Danette Allen

Abstract The popularity of unmanned aerial vehicles (UAVs) is increasing as technological advancements boost their favorability for a broad range of applications. One application is science data collection. In fields like earth and atmospheric science, researchers are seeking to use UAVs to augment their current portfolio of platforms and increase their accessibility to geographic areas of interest. By increasing the number of data collection platforms, UAVs will significantly improve system robustness and allow for more sophisticated studies. Scientists would like the ability to deploy an available fleet of UAVs to traverse a desired flight path and collect sensor data without needing to understand the complex low-level controls required to describe and coordinate such a mission. A natural interaction interface for a Ground Control System (GCS) using gesture recognition is developed to allow non-expert users (e.g., scientists) to define a complex flight path for a UAV using intuitive hand gesture inputs from the constructed gesture library. The GCS calculates the combined trajectory on-line, verifies the trajectory with the user, and sends it to the UAV controller to be flown.

Keyword Natural interaction · Gesture · Trajectory · Flight path · UAV · Non-expert user

M. Chandarana (✉) · K. Shimada
Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: mchandar@cmu.edu

K. Shimada
e-mail: shimada@cmu.edu

A. Trujillo · B. Danette Allen
NASA Langley Research Center, Hampton, VA, USA
e-mail: a.c.trujillo@nasa.gov

B. Danette Allen
e-mail: danette.allen@nasa.gov

1 Introduction

Rapid technological advancements to unmanned aerial systems foster the use of these systems for a plethora of applications [1] including but not limited to search and rescue [2], package delivery [3], disaster relief [4], and reconnaissance [5]. Many of these tasks are repetitive or dangerous for a human operator [6]. Historically, unmanned aerial vehicles (UAVs) are piloted remotely using radio remotes [7], smartphones [8], or ground control systems. This is done at the control surface level and requires a deep understanding of the extensive workings of the internal controller (much like a pilot) and the various algorithms used for system autonomy. Recently, advances in autonomy have enabled UAVs to fly with pre-programmed mission objectives, required trajectories, etc.

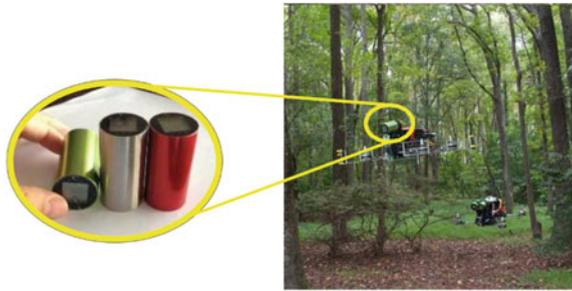
As their manufacturing costs decrease so does the cost of entry, thereby making UAVs more accessible to the masses. UAVs have also gained traction due to more robust wireless communication networks, smaller and more powerful processors, and embedded sensors [9]. Although these systems have the potential to perform a myriad of intricate tasks whose number increases with the complexity of the technology, very few people possess the required skills and expert knowledge to control these systems precisely.

1.1 *Atmospheric Science Mission*

Researchers are leveraging these autonomous vehicles to transition from traditional data collection methods to more autonomous collection methods in fields such as environmental science. Currently, environmental scientists use satellites, air balloons, and manned aircrafts to acquire atmospheric data. Although many studies require the analysis of correlative data, traditional collection methods typically only deploy individual, uncoordinated sensors, thereby making it difficult to gather the required sensor data. The current methods are also costly in terms of time and equipment [11]. Many systems such as remotely operated aerial vehicles require expert users with extensive training for operation. In addition, the possibility of sensors being lost or damaged is high for unpredictable methods like aerosondes and balloons.

Atmospheric scientists are interested in using UAVs to collect the desired data to fill the data gaps seen in current methods while giving scientists the capability to perform more elaborate missions with more direct control [12]. The portability of many sensors (Fig. 1) engenders each UAV to carry its own—potentially unique—sensor payload, thereby generating a larger density of accumulated in situ data. Each enhanced—high density—data set would provide the necessary supplemental information for conducting more sophisticated environmental studies.

Fig. 1 Example ozone sensor used on a UAV for collecting atmospheric data [10]



Ideally, scientists would like to deploy multiple UAVs to fly desired flight paths (while simultaneously taking sensor data) without being obligated to understand the underlying algorithms required to coordinate and control the UAVs. This mission management centered role requires the implementation of a Ground Control System (GCS) interface that can accurately map seemingly simplistic operator inputs into fully defined UAV maneuvers. Figure 2 depicts an example mixed-reality environment where the operator is able to manage the mission by naturally interacting with a fleet of vehicles whose geo-containment volume has been defined by the operator.

Systems which simulate human-human communication increase their accessibility to non-expert users [13]. These systems utilize natural interaction-based

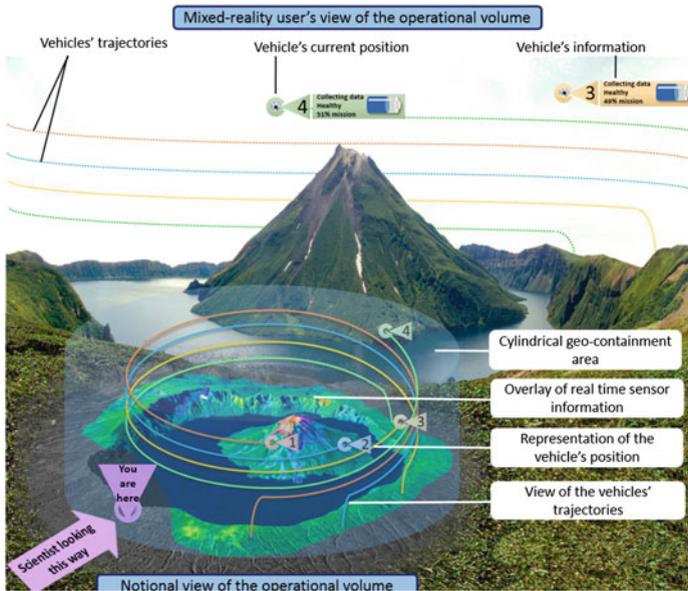


Fig. 2 Artist's representation of the future vision for an outdoor, mixed-reality environment in the context of an atmospheric science mission with multiple UAVs

interfaces. Previous researchers explored the operators' informational needs [14] and how human operators can collaborate with autonomous agents using speech and gestures [15].

The remainder of this paper will outline a natural interaction interface for a GCS in the context of an atmospheric science mission and is outlined as follows. Section 2 describes gesture-based human-robot interaction. Section 3 explains the developed natural interaction interface and the system modules. Section 4 examines the results. Section 5 draws some conclusions and discusses future work.

2 Gesture-Based Human-Robot Interaction

Using a more “natural” input is intuitive and increases system efficiency [16, 17]. However, natural interaction systems are often difficult to develop as some interpretation is required to computationally recognize their meaning. In the context of generating trajectory segments for an atmospheric science mission, gestures are an effective input method as humans naturally use gestures to communicate shapes and patterns amongst themselves [18, 19].

In the past, researchers have used a variety of gesture input methods in order to simulate interaction between an operator and a computer. Many of these input methods required the operator to hold or wear the sensor [20–22]. These input devices tended to restrict the operator's hand movements, thereby reducing the natural feel of the interface. Over time, researchers developed systems where sensors were no longer mounted or attached to the operator. Some of these systems focused on using full body gestures to control mobile robots [23], while others used static gestures to program a robot's pose by demonstration [24, 25] or even encode complex movements [26]. Neither type of system focused on using unmounted sensors to map simple, dynamic hand gestures to complex robot movements.

The Leap Motion™ controller (Fig. 3) is used to track the 3D gesture inputs of the operator. This controller uses infrared-based optics to produce gesture and position tracking with sub-millimeter accuracy. Each individual bone's position within all ten fingers of the operator are estimated (Fig. 4). User's fingertips are detected at up to 200 frames per second at 0.01 mm accuracy. Three infrared cameras work together to provide 8 ft³ volume of interactive space [27].

Fig. 3 Leap Motion™ controller

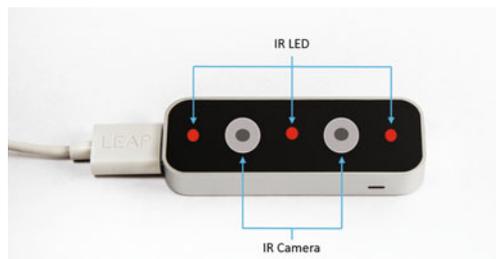
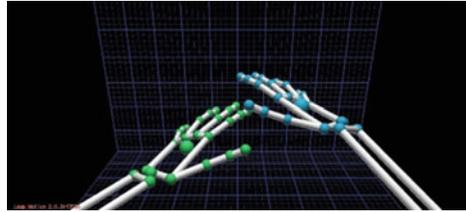


Fig. 4 Visual representation of the Leap Motion™ controller's estimate of the operator's finger locations



3 Ground Control System Framework

A natural interaction interface for a GCS using gesture recognition was developed to allow non-expert operators to quickly and easily build a desired flight path for an autonomous UAV by defining trajectory segments. The developed system addresses several key requirements.

- i. The gesture inputs shall be intuitive and make the operator feel as if he/she were naturally interacting with another person.
- ii. The interface shall be simple to walk through so as to reduce the required training time and increase the usability for non-experts.
- iii. There shall be ample user feedback for decision making.
- iv. The system shall abstract the operator away from the complexity of the low-level control operations, which are performed autonomously, and allow them instead to have a high-level mission management role.

The system's interface is broken down into several key modules (Fig. 5):

- A. volume definition module
- B. gesture module
- C. trajectory generation module

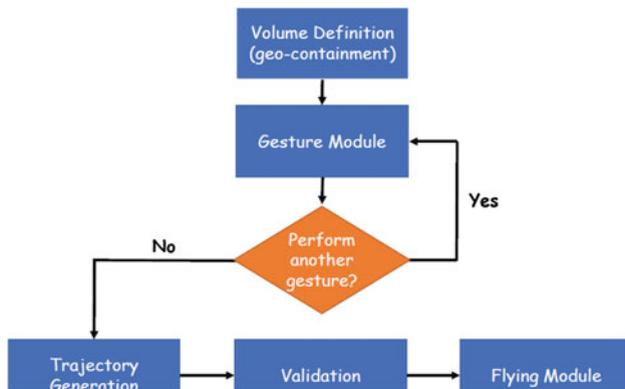


Fig. 5 Diagram of the interface's process flow between modules

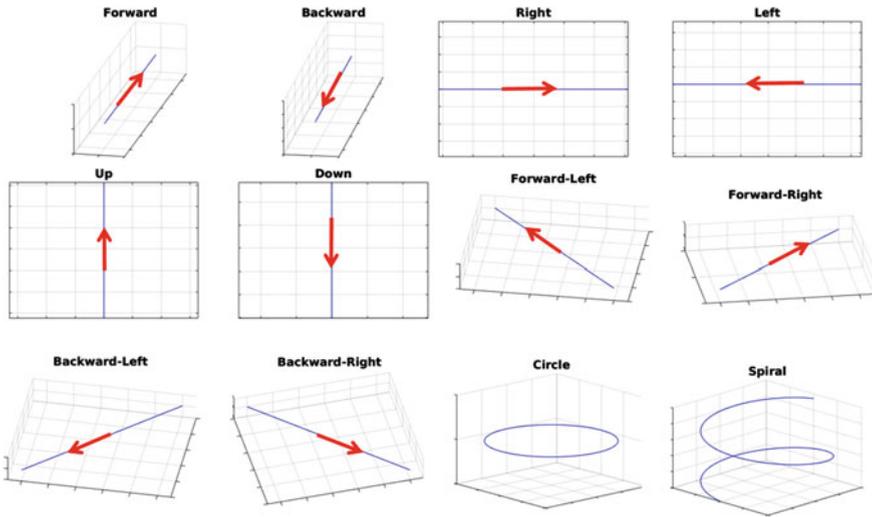


Fig. 6 Library of gesture inputs

D. validation module

E. flight module

The first module asks the operator to define the size of the desired operational volume. The operator can then build the flight path by gesturing one of twelve trajectory segments currently defined in the intuitive gesture library (Fig. 6). The interface allows the operator to build up a path with any number of segments. Once the operator has fully defined the flight path, the system combines the trajectory segments with first order continuity before taking the operator through a validation step. This combined trajectory is then sent to the vehicle controller and the operator may then send takeoff and land commands. Throughout each step, the interface provides the operator with feedback displays so as to increase the efficiency of decision making, as well as, provide a logical progression through the system.

3.1 Volume Definition Module

The operator is able to define a desired operational volume for the UAV using the volume definition module. This operational volume can demarcate the operational area, as well as, a geo-containment zone and may be set as a hard boundary for the vehicle's position in the onboard controller. As a safety measure, if the vehicle were to leave the geo-containment area, the vehicle controller can immediately command the vehicle to land. Once the user creates the geo-containment volume, all defined trajectories are restricted to this volume. Currently, this is done by manually defining the lengths and orientations of the segments such that they are guaranteed

to stay within the boundary. This can be extended to automatically scale and orient the segments by using the size and shape of the boundary as parameters.

In order to aid the operator in defining a volume, the interface displays multiple 2D viewpoints of the working environment so as to accurately display the full 3D shape of the volume. These viewpoints are planar diagrams of the total available working volume which, when put together give the operator a 3D understanding. A representation of the desired operational volume (to scale) in an indoor setting is overlaid on the viewpoints (Fig. 7). These viewpoints allow the operator to visualize the entire volume as shown in Fig. 7 (top). Figure 2 shows an example volume overlay in an outdoor setting. The operator can use an intuitive pinching gesture to expand or compress the volume to the desired size. This example uses a cylindrical volume represented in Imperial units. The units used to define the operational volume are tailorable.

3.2 Gesture Module

By using the defined gesture library (Fig. 6) the operator is able to employ a variety of trajectory segments. The gesture module is responsible for characterizing the raw sensor data from the Leap Motion™ controller into desired trajectory segment shapes.

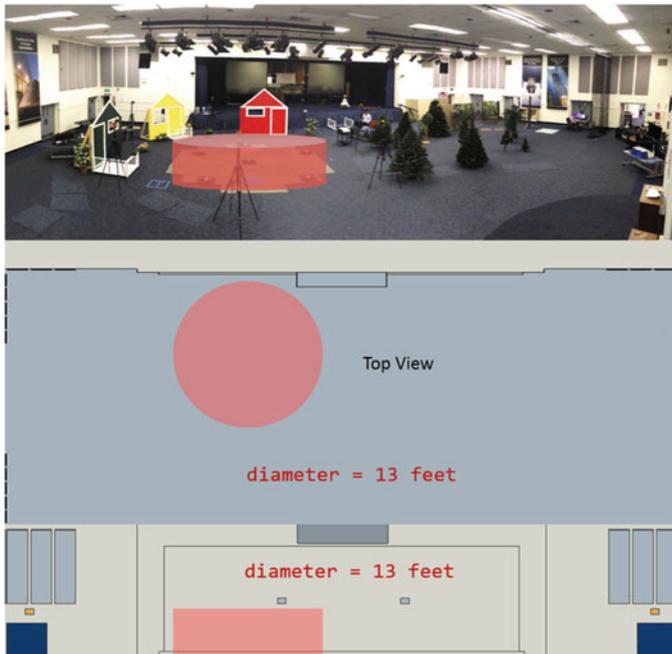


Fig. 7 Top view (middle) and front view (bottom) representations shown in the interface for use in volume definition

Gesture Library. Based on feedback from subject matter experts, the gesture library was developed to define flight paths that would be of interest to scientists for collecting various sensor data. The library is composed of twelve gestures ranging from simple linear gestures to more complex diagonal and curved gestures: forward, backward, right, left, up, down, forward-left, forward-right, backward-left, backward-right, circle, and spiral.

The intuitive nature of the gesture library lends itself to applying several of the gestures in other aspects of the interface, in addition to path definition. This is done by breaking the flow of the interface into several modes. After building a trajectory segment, the Message Mode gives the operator the ability to use the right and left gesture inputs to navigate a menu questioning if they have taught all desired trajectory segments. Additionally, the up and down gesture inputs are adopted for sending takeoff and land commands to the UAV controller in the Flight Mode, which occurs after the operator has built all their desired segments.

3D Gesture Characterization. The operator's gesture input is characterized as one of the twelve gestures in the defined library using a Support Vector Machine (SVM) model. As opposed to a simplistic threshold separation, the SVM model provides a more robust classification by reducing biases associated with any one user.

Data from eleven subjects was used to train the SVM model using a linear kernel [28]. Each subject provided ten labeled samples per gesture in the library. Features were extracted from each pre-processed input sample. The features used for training included the hand direction movement throughout the gesture, as well as, the eigenvalues of the raw hand position data.

3.3 Trajectory Generation Module

$$B(t) = \sum_{i=0}^n \binom{n}{i} (1-t)^{n-i} t^i p_i. \quad (1)$$

Once the operator has defined all the desired trajectory segments, the trajectory generation module translates each segment into a set of fifth order Bézier splines [29] represented by Eq. 1. The system then automatically combines the spline sets into a complete path by placing the Bézier splines end-to-end in order of their definition. Transitions between splines are smoothed with first order continuity so that any sudden direction and velocity changes are eliminated. This produces a continuous, flyable path for the UAV. Each combined trajectory includes the necessary time, position, and velocity data required by the UAV controller. Using a data distribution service (DDS) middleware, the trajectory information can be transmitted on-line to the UAV controller [30].

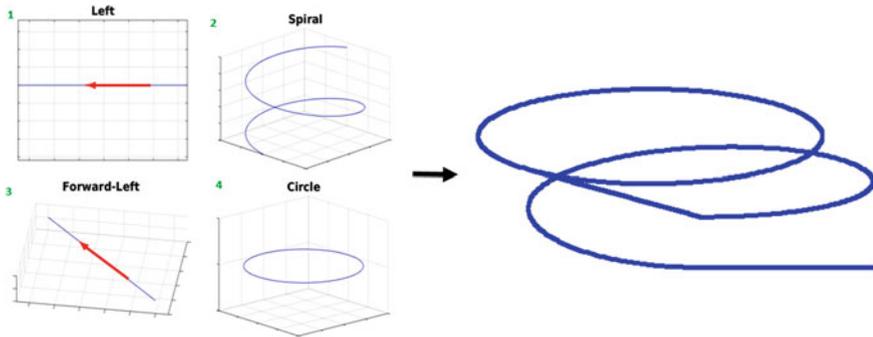


Fig. 8 Sample rendered 3D combined trajectory built by the operator (*right*) with input gestures (*left*)

3.4 Validation Module

Prior to transferring the combined trajectory data taught by the operator to the UAV controller, the system displays a 3D representation of the flight path. The rendered flight path is presented to the operator so that they can validate and approve the proposed path of the UAV before flight. Figure 8 depicts an example of a rendered flight path where the operator taught the following trajectory segments (in order): left, spiral, forward-left, and circle. This flight path simulates a UAV collecting data as it spirals up to a desired altitude, moves to a new region of interest, and circles around a target while collecting more data.

3.5 Flight Module

After the UAV controller receives the flight data, the operator is able to initiate and end path following using the features of the flight module. In the current setup, Up and Down gestures are translated into takeoff and land commands respectively. The commands are sent directly to the UAV controller using the DDS middleware. Once a takeoff command is sent and executed, the UAV autonomously begins traversing the newly defined flight path using the transferred data. Immediately upon completing its defined path, the UAV hovers in place until the operator sends a land command.

4 Results

A straightforward system setup (Fig. 9 left) exudes the sense of effortless execution to the operator. The system requires just two components: (1) a flat surface to place the Leap Motion™ Controller and (2) a display with the interface loaded and a



Fig. 9 System setup for operator use (*left*) and view of UAV operational volume from operator's perspective (*right*)

USB port to connect the Leap Motion™ Controller. In the context of an atmospheric science mission, the operator is situated such that they have a frame of reference for building the trajectory segments for the UAV while maintaining a safe separation (Fig. 9 right). The operator uses the gesture library to intuitively walk through the steps required to build desired trajectory segments. There is no limit placed on the number of segments the operator can teach.

Each time the operator uses the natural interaction interface, the system calculates the combined trajectory and discretizes every segment's set of Bézier splines [6] for visualization in the validation module on-line. The system is able to accurately transmit the complete trajectory data to the UAV controller such that the UAV flies a recognizable implementation of the path generated by the operator. Figure 10 depicts a resulting sample flight path generated by employing the developed gesture-based interface. For this example, the operator used the developed interface to generate a trajectory for a single UAV with the following directions (in order): left, forward and spiral. The position of the UAV as it traverses the complete trajectory was overlaid on one image for easier understanding. A light blue line is used to trace the path of the UAV. Although not depicted, the operator was also able to perform an Up gesture to send a takeoff command to the vehicle and a Down gesture to land the vehicle.

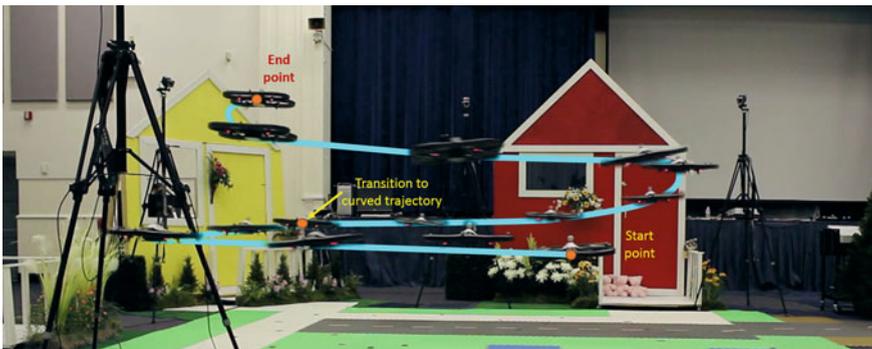


Fig. 10 Sample flight path after trajectory data was sent to the controller and a takeoff command

5 Conclusion and Future Work

A GCS for autonomous UAV data collections was developed. The system utilizes a gesture-based natural interaction interface. The intuitive gesture library allows non-expert operators to quickly and easily define a complete UAV flight path without needing to understand the low-level controls. There are several planned extensions that can be made to this work to augment its efficacy and applicability.

1. The interface can be extended to include the definition of additional geometric constraints if necessary (e.g., clockwise vs counterclockwise direction on a circle).
2. The gesture library can be extended (e.g., spiral forward).
3. Gesture segmentation: A user may wish to define a complete square shape instead of teaching each segment one-by-one.
4. Extending the system to include real-time mission supervision and trajectory modification.
5. Perform user studies to fully validate the methodology.

Acknowledgments The authors would like to thank Javier Puig-Navarro and Syed Mehdi (NASA LaRC interns, University of Illinois at Urbana-Champaign) for their help in building the trajectory module, Gil Montague (NASA LaRC intern, Baldwin Wallace University) and Ben Kelley (NASA LaRC) for their expertise in integrating the DDS middleware, Dr. Loc Tran (NASA LaRC) for his insight on feature extraction, and the entire NASA LaRC Autonomy Incubator team for their invaluable support and feedback.

References

1. Jenkins, D., Bijan, B.: The economic impact of unmanned aircraft systems integration in the United States. Technical Report, AUVSI (2013)
2. Wald, M.L.: Domestic drones stir imaginations and concerns. *New York Times* (2013)
3. Barr, A.: Amazon testing delivery by drone, CEO Bezos Says. *USA Today*. <http://www.usatoday.com/story/tech/2013/12/01/amazon-bezos-dronedelivery/3799021/>
4. Saggiani, G., Teodorani, B.: Rotary wing UAV potential applications: an analytical study through a matrix method. *Aircraft Eng. Aerosp. Technol.* **76**(1), 6–14 (2004)
5. Office of the Secretary of Defense.: Unmanned aircraft systems roadmap 2005–2030. Washington, DC (2005)
6. Schoenwald, D.A.: AUVs: in space, air, water, and on the ground. *IEEE Control Syst. Mag.* **20**(6), 15–18 (2000)
7. Chen, H., Wang, X., Li, Y.: A Survey of autonomous control for UAV. In: *IEEE International Conference on Artificial Intelligence and Computational Intelligence*, pp. 267–271. Shanghai (2009)
8. Hsu, J.: MIT researcher develops iPhone app to easily control swarms of aerial drones. In: *Popular Science*. <http://www.popsci.com/technology/article/2010-04/video-usingsmartphonescontrol-aerial-drones>
9. Zelinski, S., Koo, T.J., Sastry, S.: Hybrid system design for formations of autonomous vehicles. In: *IEEE Conference on Decision and Control* (2003)

10. Miniature Air Quality Monitoring System, <http://www.altechusa.com/odor.html>
11. Wegener, S., Schoenung, et al.: UAV autonomous operations for airborne science missions. In: AIAA 3rd Unmanned Unlimited Technical Conference, Chicago (2004)
12. Wegener, S., Schoenung, S.: Lessons learned from NASA UAV science demonstration program missions. In: AIAA 2nd Unmanned Unlimited Systems, Technologies and Operations Conference, San Diego (2003)
13. Perzanowski, D., Schultz, A.C., et al.: Building a multimodal human-robot interaction. In: IEEE Intelligent Systems, pp. 16–21 (2001)
14. Trujillo, A.C., Fan, H., et al.: Operator informational needs for multiple autonomous small vehicles. In: 6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, Las Vegas, pp. 5556–5563 (2015)
15. Trujillo, A.C., Cross, C., et al.: Collaborating with autonomous agents. In: 15th AIAA Aviation Technology, Integration and Operations Conference, Dallas (2015)
16. Reitsema, J., Chun, W., et al.: Team-centered virtual interaction presence for adjustable autonomy. In: AIAA Space, Long Beach (2005)
17. Wachs, J. P., K“olsch, M., Stern, H., Edan, Y.: Vision-based hand-gesture applications. *Commun. ACM* **54**(2), 60–71 (2011)
18. McCafferty, S.: Space for cognition: gesture and second language learning. *Int. J. Appl. Linguist.* **14**(1), 148–165 (2004)
19. Torrance, M.: Natural communication with robots. M.S. Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge (1994)
20. Iba, S., Weghe, M.V., et al.: An architecture for gesture based control of mobile robots. In: IEEE International Conference on Intelligent Robots and Systems, pp. 851–857 (1999)
21. Neto, P., Pires, J.N.: High-level programming for industrial robotics: using gestures, speech and force control. In: IEEE International Conference on Robotics and Automation, Kobe (2009)
22. Yeo, Z.: GestureBots: intuitive control for small robots. In: CHI 28th Annual ACM Conference on Human Factors in Computing Systems, Atlanta (2010)
23. Waldherr, S., Romero, R., Thrun, S.: A gesture based interface for human-robot interaction. *Auton. Robots* **9**, 151–173 (2000)
24. Becker, M., Kefalea, E., et al.: GripSee: a gesture-controlled robot for object perception and manipulation. *Auton. Robots* **6**, 203–221 (1999)
25. Lambrecht, J., Kleinsorge, M., Kruger, J.: Markerless gesture-based motion control and programming of industrial robots. In: 16th IEEE International Conference on Emerging Technologies and Factory Automation, Toulouse (2011)
26. Raheja, J.L., et al.: Real-time robotic hand control using hand gestures. In: IEEE 2nd International Conference on Machine Learning and Computing, Washington, DC (2010)
27. Bassily, D., et al.: Intuitive and adaptive robotic arm manipulation using the leap motion controller. In: 8th German Conference on Robotics, Munich (2014)
28. King, D.E.: Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* **10**, 1755–1758 (2009)
29. Choe, R., et al.: Trajectory generation using spatial pythagorean hodograph Bézier curves. In: AIAA Guidance, Navigation and Control Conference, Kissimmee (2015)
30. DDS Standard, <https://www.rti.com/products/dds/omg-ddsstandard.html>

Part VII
Optimizing Human-Systems Performance
Through System Design

An Analysis of Displays for Probabilistic Robotic Mission Verification Results

Matthew O'Brien and Ronald Arkin

Abstract An approach for the verification of autonomous behavior-based robotic missions has been developed in a collaborative effort between Fordham University and Georgia Tech. This paper addresses the step after verification, how to present this information to users. The verification of robotic missions is inherently probabilistic, opening the possibility of misinterpretation by operators. A human study was performed to test three different displays (numeric, graphic, and symbolic) for summarizing the verification results. The displays varied by format and specificity. Participants made decisions about high-risk robotic missions using a prototype interface. Consistent with previous work, the type of display had no effect. The displays did not reduce the time participants took compared to a control group with no summary, but did improve the accuracy of their decisions. Participants showed a strong preference for more specific data, heavily using the full verification results. Based on these results, a different display paradigm is suggested.

Keywords Display of uncertainty · Decision support system · Formal verification · Behavior-based robotics

1 Introduction

Robotics has the potential to be a key technology for combating weapons of mass destruction [1]. This domain presents new challenges for autonomous robotic systems. In these types of missions, failure is not an option. Human operators must be confident in the success of a robotic system before the technologies can be applied. To address these problems our research, conducted for the Defense Threat

M. O'Brien (✉) · R. Arkin
School of Interactive Computing, Georgia Tech, Atlanta, GA 30332, USA
e-mail: mjobrien@gatech.edu

R. Arkin
e-mail: arkin@gatech.edu

Reduction Agency (DTRA), has successfully developed the methods and software to perform robotic mission verification [2].

While robotic mission verification is similar to traditional software verification, there are several additional complications. The real world is continuous, and both robotic sensors and actuators are noisy. The robotic controller is only one piece, and the result of a mission is also determined by the physical robot and its interaction with the environment, and modeling of both will always be imperfect. This means any verification is fundamentally probabilistic.

This presents a new challenge. People do not use all the available data or systematic methods when assessing probabilistic data. Instead, heuristics are applied to simplify the analysis, which can lead to systematic errors and bias [3]. The methods of displaying the information must ensure an operator can easily and accurately interpret the data. This paper explores methods to achieve this goal.

2 Related Work

Research on the presentation of probabilistic data and uncertainty has shown that participant's decisions in various tasks are not significantly affected by the format the data is presented in (graphical, numerical, or verbal) [4]. Though numeric statements offer more precision and consistency than linguistic phrases, it's hypothesized that people treat all probabilities in a vague manner, utilizing membership functions [5].

These results were extended in [6] where both display format and specificity level were varied. Display formats included linguistic, numeric, and multiple graphical icons. Specificity level was the size of the range of probabilities represented by a single icon or expression. Results agreed with previous research, showing that display format had no significant effects. However, specificity did have significant effects on performance in a simulated stock purchasing task.

This work expands on these results in two ways related to the application of robotic mission verification. First, participants in this study have access to more information than a single measure of probability. Success in a robotic mission is tied to multiple criteria, such as time to completion or allowable distance from a goal location, whose values may have some variability. The full verification results have probabilities of achieving each criterion independently over a range of values. This information is important to operators, and will affect their decisions, so it must be included. Secondly, the context of the tasks is significantly different. Participants were asked to make decisions on high-risk missions, where lives are (hypothetically) at risk. These types of risks/costs are difficult to quantify and participants may resort to different methods of reaching a decision.

3 VIPARS—The Verification Tool

VIPARS, or *Verification in Process Algebra for Robot Schemas*, is a robot mission verification tool [7] designed for use with *MissionLab*, a graphical programming environment for behavior-based robots [8]. Informally, VIPARS determines how likely a robot mission is to succeed. In formal terms it takes as input a behavior-based robotic controller (software), models of the robot hardware and environment, and performance criteria. With this information, VIPARS can calculate and return a probability of success. All of these components are descriptions of the physical system except for performance criteria, which define what a successful mission is. The two most fundamental criteria, and those used for this study, are time (how long a robot may take to achieve its goal) and space (how far from a goal a robot may be).

VIPARS achieves verification by defining the state of the system as a set of random variables. Flow functions, created from the robot’s behaviors and the environmental models, describe how these random variables map from one time step to the next. This allows VIPARS to avoid the state-space explosion, caused by the continuous dynamics and noisy sensing/actuation of the real world, that plague traditional verification techniques such as model checking [9]. This paper will not go deeper into the description of VIPARS. For a thorough description of the verification process please see [2].

Though VIPARS can produce a single probability for specific performance criteria, a more complete understanding of a mission can be gathered from observing how the probability changes over a range of criteria. See Fig. 1 for the

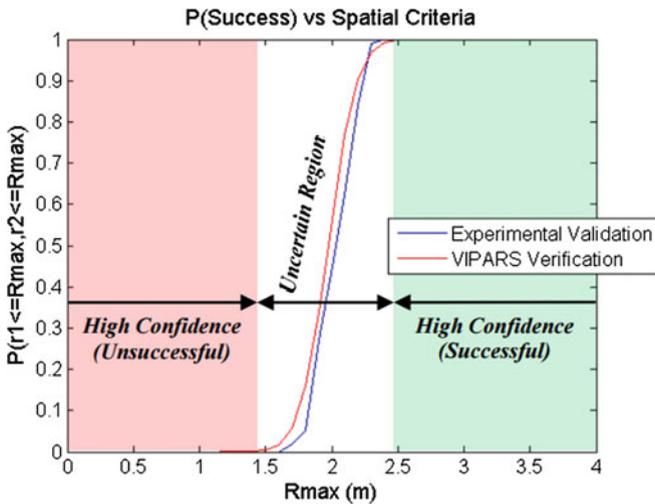


Fig. 1 Example verification results compared to empirical validation from real robots. Rmax is the spatial criterion, or the max distance from the goal location that still counts as a successful mission. The y-axis is the probability that both robots, r1 and r2, meet their spatial criterion

results of a multi-robot mission verification from [7]. The red curve is the VIPARS verification result, while the blue curve is experimental data gathered from real executions of the robot mission for validation purposes. Results can be broken down into three regions. In the Unsuccessful region, the performance criterion is so strict that success is impossible, i.e., the precision or speed demands exceed the capabilities of the system. In the Successful region, the criteria are easy enough to guarantee success (ignoring unmodeled possibilities). Both of these regions are high-confidence, where it is certain the actual mission probability will match the verification results. In between lies the uncertain region, where uncertainty is introduced in two ways. First, the results are between 0 and 100 %, so mission success is uncertain even with a perfect verification. Second, in this region small errors or simplifications in modeling can create moderate differences between the predicted and actual probability of success. Thus, the results of the verification itself are low-confidence. With a basic understanding of VIPARS and the data it produces, the experiment described in this paper and the displays used can be discussed.

4 Experimental Design

4.1 Task

Participants were asked to make decisions on whether to execute high risk autonomous robotic missions based on situational information and the verification results. Participants were presented scenarios appropriate for a mobile robot mission. They were given access to the VIPARS graphical interface via a laptop under the assumption the robot (hardware) and controller (software) had been decided and are fixed. The participants reviewed information on a scenario (robot's task, risk factors, time or spatial constraints) and then executed the VIPARS verification. Using the information VIPARS provided, the participant made a decision on whether to execute the robot mission or defer to a human team, and rated their confidence in both the mission and their decision. Scenarios included some limited information about the performance and risk for human teams.

Each participant was presented five total scenarios. The scenarios were divided into two categories. Certain scenarios were made to have a clear correct decision, with probabilities of success either being 0 % or 100 %, and high confidence in the verification. There were three certain scenarios, two successful and one unsuccessful. The uncertain scenarios had probabilities of success at 30 and 70 %, as well as low confidence in the verification.

4.2 Independent Variable—The Displays

Participants were divided into four conditions. Every condition had the low-level display available, which showed the full verification results. For three conditions (A-C), subjects were presented a variant of a high-level display and could switch to the low-level display at will, while in the control condition (D) subjects could only view the low-level display.

The Low-Level Display. The low-level display provides the full probabilistic information given by VIPARS. This display is based on the validation graphs discussed in Sect. 3. The graph is a cumulative distribution function (CDF) for the probability of achieving a performance criterion over a range of values. Figure 2 shows an example of the graph. It is augmented in several ways to aid a user. Areas with 0 % or 100 % success probabilities make up the “high-confidence success” or “high-confidence failure” regions. These are colored for rapid identification. The threshold a user has selected for the specific criteria is marked with a dashed line. In addition, the scales of the presented graph are selected relative to this threshold; from zero to twice the value. This presentation is limited to one criterion at a time, so a user must manually switch which criteria they are viewing.

The High-Level Displays. The high-level display summarizes the verification results in two ways. First, only the probability at the selected criteria value is used. This means information on the effect of changes in mission criteria is lost. Second, the results of all criteria are combined to give a total probability of success. This removes mental calculations from the user, but hides potential causes of failure. Three display types were chosen that vary with respect to type and specificity.

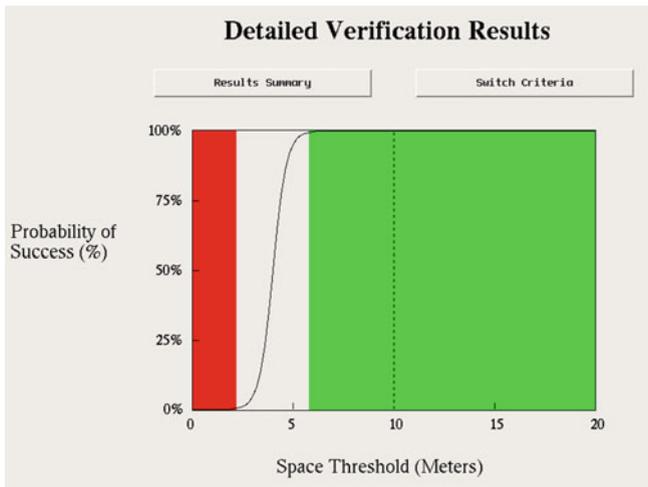


Fig. 2 Example low-level display for a spatial criterion set at 10 m

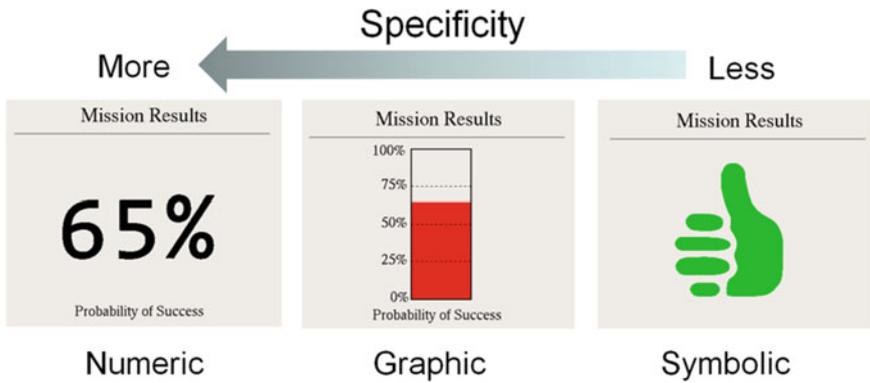


Fig. 3 The high-level displays. The numeric display was used for condition A, the graphic for B, and symbolic for C. Condition D was the control group

At the high end of the specificity scale is a simple numeric display of the final mission probability, which can be considered the most basic approach. A less precise means of displaying a percentage is graphically, using a bar. A bar was selected because reading position along a common scale has been shown to be the most accurate task for extracting quantitative information from a graphical representation [10] and it is commonly used in decision support systems (e.g. [11, 12]). At the lowest level of specificity, a symbolic system only presenting three options (success, failure, and uncertain) could be used for the high-level display. This scheme takes advantage of the current predictions of VIPARS which typically have low confidence in probabilities between 0 and 100 %. In this symbolic system a green thumbs up represents success, a red thumbs down represents failure, and a question mark represents uncertain results. Figure 3 presents the options along a scale of specificity.

4.3 Dependent Variables

For each scenario five dependent variables were recorded, shown in Table 1. The first three variables were automatically recorded by the software, while the last two are selected by the participant. Mission and decision confidence were presented as Likert

Table 1 The five dependent variables recorded in the study

User-decision	Binary choice on whether to execute the robotic mission
Time-to-decision	The time between VIPARS execution and final decision
Time-on-raw-data	Time spent viewing the low-level display
Mission-confidence	Confidence the robotic mission would be successful if ran
Decision-confidence	Confidence the user’s decision (to execute or not) is correct

scales with values ranging from 1 to 9. For participants in the control group, *time-to-decision* is equal to *time-on-raw-data*, as they can only view the low-level display.

4.4 Hypotheses

Based on the related work discussed previously, three hypotheses were formed. This section covers the hypotheses and their predictions on the dependent variables.

- Hypothesis 1: Displays summarizing VIPARS results can improve the comprehension accuracy and speed of users over the direct display of VIPARS output.

Hypothesis 1 predicts that *time-to-decision* and *time-on-raw-data* will be reduced when using high level displays versus the control case, and that the accuracy of user-decision will increase for certain scenarios. If participants in the control cases achieve perfect accuracy (i.e. always select correct decision) for certain scenarios, then it will be assumed that perfect accuracy on the high-level displays validates this hypothesis.

- Hypothesis 2: Various representations of the VIPARS output will provide similar understanding of the mission probability.

Hypothesis 2 predicts that between the high level displays, user-decision will not vary significantly for uncertain scenarios.

- Hypothesis 3: More precise representations of probability will bias operators towards interpreting higher certainty in the result.

Hypothesis 3 predicts that *decision-confidence* will increase as the specificity of the high-level display increases.

Finally, additional analysis is performed to look for effects that do not have explicit hypotheses. For example, if one particular display has a higher *time-on-raw-data* on the average, it may indicate that users find the representation inadequate for decision making.

4.5 Execution Details

A total of 45 participants were tested. Participants were screened for color blindness with a shortened version of the Ishihara colorblind test, two failed and were excluded. In addition, two participants performed the tasks incorrectly,¹ their data

¹Participants used prior situational information for new scenarios.

was also excluded. The results include 41 participants, 23 male and 18 female, with an average age of 24.3 (range from 18 to 54).

Each participant first went through a tutorial session that introduced the VIPARS system and allowed them to try an example scenario. Afterwards, they were given information on one scenario at a time by the proctor. The proctor was nearby and available for questions, but not able to view the computer or participant's choices. Sessions were video recorded, and the time taken for questions and answers during the test was removed from the measurements of *time-to-decision* and *time-on-raw-data*.

5 Results

5.1 Hypothesis 1

The first hypothesis made two predictions. First, that users would make faster decisions when provided with the high-level displays, lowering *time-to-decision* and *time-on-raw-data*. The second was that the accuracy of user's decisions on certain scenarios would be improved when using high-level summaries. First we examine the data on *time-to-decision* and *time-on-raw-data*.

Both *time-to-decision* and *time-on-raw-data* were analyzed using one-way ANOVA over the four conditions. For *time-to-decision*, or the total time a user took, there was no significant difference between display types when all scenarios were averaged together ($P = 0.688$). Scenarios were also tested independently, and showed no significant differences. Figure 4 shows the *time-to-decision* for each display type. In contrast, there was a statistically significant difference between

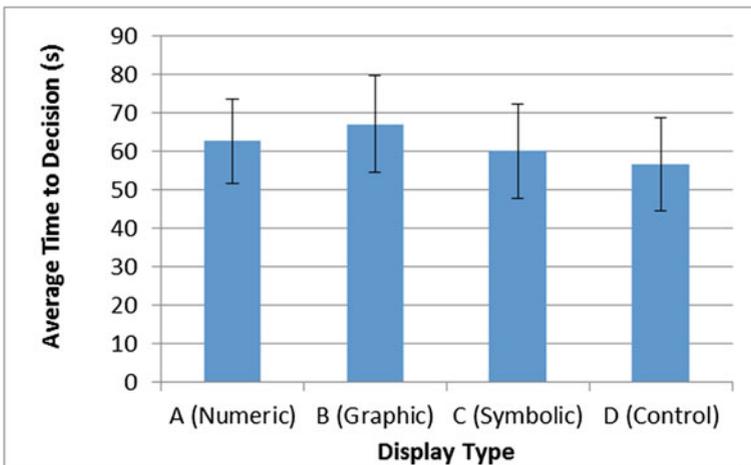


Fig. 4 Average *time-to-decision* per display type for all scenarios plotted with 95 % confidence intervals

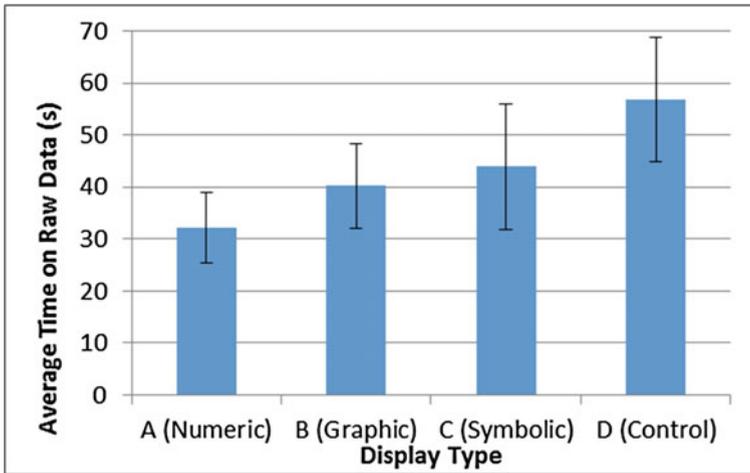


Fig. 5 Average *time-on-raw-data* per display type for all scenarios plotted with 95 % confidence intervals

Table 2 All decisions for the certain-scenarios, sorted by condition

		Display type			
		A	B	C	D
Decision	Correct	28	33	32	22
	Incorrect	2	0	1	5

time-on-raw-data ($P = 0.012$). Post hoc tests using Games-Homell showed only display A had a significant reduction ($\alpha = 0.05$) compared to the control (means = 32.22, 56.84, SD = 24.47, 41.27) (Fig. 5).

For decision accuracy, uncertain scenarios were excluded as no correct decision could be assumed. This left the three certain scenarios. For each user and display, a correct decision was an “execute” for missions with 100 % probability of success, and a “do not execute” for missions with a 0 % probability of success. The table of decisions for each display is shown in Table 2. The reader can see that the control case D has a larger number of incorrect decisions. As the table is sparsely populated, Fisher’s exact test was used to test for statistical significance. A significant difference between display types was found ($P = 0.026$). Thus hypothesis one is partially confirmed; the accuracy of users improved with the high-level displays, but their times to decisions were not reduced.

5.2 Hypothesis 2

The second hypothesis predicted that between the high level displays, the understanding of mission probability and thus *user-decision*, would not vary significantly

Table 3 *User-decisions* for the first uncertain scenario, total probability of success = 70 %

		Display type			
		A	B	C	D
Decision	Execute	6	8	8	7
	Don't	4	3	3	2

Table 4 *User-decisions* for the second uncertain scenario, total probability of success = 30 %

		Display type			
		A	B	C	D
Decision	Execute	5	5	6	7
	Don't	5	6	5	2

for uncertain scenarios. As these scenarios had different probabilities of success (70 and 30 %) they will be analyzed separately. The decisions for each scenario are broken down in Tables 3 and 4, and a Fisher's exact test reported the difference between display types was not statistically significant. ($P = 0.906, 0.526$ for scenarios one and two, respectively). Thus hypothesis two is confirmed.

5.3 Hypothesis 3

The final hypothesis predicted that more precise representations of probability will bias operators towards interpreting higher certainty in the results, thus *decision-confidence* will increase as the precision of the high-level display increases. This hypothesis needs to be tested per scenario, as different risks and probabilities with each scenario should affect the confidence of the user. Performing an ANOVA for *decision-confidence* versus display type showed no significant differences between displays. Table 5 and Fig. 6 display the P values and average values for each scenario. Thus hypothesis three is rejected.

6 Discussion

This section will cover the key results from this study, and how they have impacted the design of the VIPARS interface.

Table 5 ANOVA results for *decision-confidence* versus display type for each scenario

Scenario	1	2	3	4	5
P value	0.56	0.48	0.09	0.49	0.52

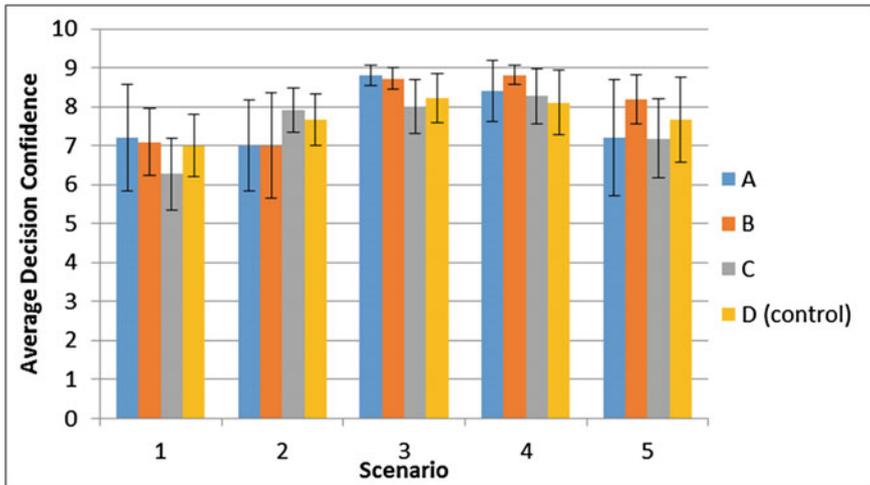


Fig. 6 Average decision-confidence per display type, per scenario, and plotted with 95 % confidence intervals

1. Users wanted the most detail possible

Almost all users in high-level display conditions heavily utilized the low-level display as well. The author predicted the high-level displays would decrease the time a user needs, but the opposite was true. Users in the control condition had the lowest *time-to-decision*, though it was not statistically significant. The reason is obvious from test data, users spent time reviewing both levels of displays when they were available. As seen in Sect. 5.1, only condition group A (having the most specific high-level display) had a statistically significant reduction in *time-on-raw-data* compared the control group. This is consistent with previous work which found preferences for higher specificity [6].

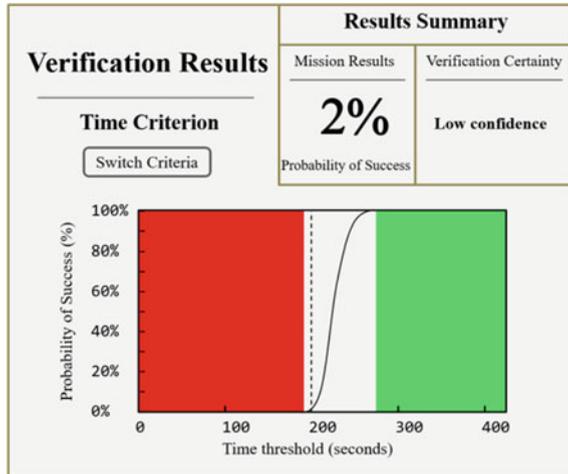
2. The type of high-level display had almost no effect

There was no significant difference between the numeric, symbolic, or graphical displays except for *time-on-raw-data*. This is consistent with the previous work [4, 6] that showed display format has little impact, but in this experiment specificity was also varied. Does this disagree with previous results that showed specificity had a significant effect? The authors do not believe so. In this experiment, users had access to a more specific information source in the low-level display. As most participants heavily utilized the low-level display, it seems likely that the variance in specificity at the high-level was overshadowed by the information from the low-level display.

3. The high-level displays helped reduce errors

Due to either misinterpreting the low-level graphs, or improperly combining the results of multiple criteria, more mistakes were made in the control group. While in a more realistic setting users would have additional training (reducing the likelihood of errors), the actual situations may be more complex and include several extra criteria (increasing the likelihood of errors).

Fig. 7 New example display design that combines the high-level summary with the complete low-level results



Initial designs for the VIPARS user interface, and the prototype display for this experiment, utilized a layered system, where a user is presented with a high-level summary first, and would only view low-level detailed information if necessary. These results indicate that while a summary of results is useful, it likely should not be the primary focus. Instead, the complete verification results should be the primary output, with automatic summaries displayed alongside as a mental check for users. See Fig. 7 for an example design. The choice of display format for the summary is not critical, as no option showed superior performance, however results suggest users may prefer the numerical display due to its greater precision.

7 Conclusion

This paper has presented research on the display of uncertainty towards robotic mission verification. A human study on the display of probabilistic data for robot missions was performed. Three high-level summaries were chosen to present the results of a mission verification software toolkit. Surprisingly, the high-level summaries did not affect the time a user took, or their confidence with their decision. Instead, participants preferred to utilize the low-level detailed results. The control group, without access to the summarized data, made more mistakes. This implies some value in the high-level displays for the purpose of ensuring a user has accurately interpreted the verification results. The outcomes of this study have improved the design paradigm of the VIPARS interface; helping to ensure users will be able to quickly and accurately interpret the probabilistic information.

Acknowledgments This research is supported by the United States Defense Threat Reduction Agency, Basic Research Award #HDTRA1-11-1-0038.

References

1. Doesburg, J.C., Steiger, G.E.: The evolution of chemical, biological, radiological, and nuclear defense and the contributions of army research and development. NBC Report, the United States Army Nuclear and Chemical Agency (2004)
2. Lyons, D.M., Arkin, R.C., Jiang, S., Liu, T.M., Nirmal, P.: Performance verification for behavior-based robot missions. *IEEE Trans. Rob.* **31**(3) (2015)
3. Tversky, A., Kahneman, D.: Judgment under uncertainty: heuristics and biases. *Science* **185** (4157), 1124–1131 (1974)
4. Budescu, D.V., Weinberg, S., Wallsten, T.S.: Decisions based on numerically and verbally expressed uncertainties. *J. Exp. Psychol. Hum. Percept. Perform.* **14**(2), 281 (1988)
5. Wallsten, T.S., Budescu, D.V.: A review of human linguistic probability processing: general principles and empirical evidence. *Knowledge Eng. Rev.* **10**(01), 43–62 (1995)
6. Bisantz, A.M., Marsiglio, S.S., Munch, J.: Displaying uncertainty: investigating the effects of display format and specificity. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **47**(4), 777–796 (2005)
7. Lyons, D.M., Arkin, R.C., Jiang, S., Harrington, D., Liu, T.M.: Verifying and validating multirobot missions. *IEEE/RSJ Int. Conf. Intelligent Robots Syst. (IROS)* (2014)
8. MacKenzie, D.C., Arkin, R.C., Cameron, J.M.: Multiagent mission specification and execution. In: *Robot Colonies*, pp. 29–52. Springer, US (1997)
9. Jhala R., Majumdar R.: Software model checking. *ACM Comput. Surv.* **41**(4) 21–53 (2009)
10. Cleveland, W.S., McGill, R.: Graphical perception: theory, experimentation, and application to the development of graphical methods. *J. Am. Statistical Association* **79**(387), 531–554 (1984)
11. Daradkeh, M., Churcher, C., McKinnon, A.: Supporting informed decision-making under uncertainty and risk through interactive visualisation. In: *Proceedings of the Fourteenth Australasian User Interface Conference*, vol. 139, pp. 23–32. Australian Computer Society, Inc. (2013)
12. Masalonis, A., Mulgund, S., Song, L., Wanke, C., Zobell, S.: Using probabilistic demand predictions for traffic flow management decision support. In: *Proceedings of the 2004 AIAA Guidance, Navigation, and Control Conference*. American Institute of Aeronautics and Astronautics (2004)

A Neurophysiological Assessment of Multi-robot Control During NASA's Pavilion Lake Research Project

John G. Blich

Abstract A number of previous studies have explored the value of an external or “3rd person” view in the realm of video gaming and augmented reality. Few studies, however, actually utilize a mobile robot to provide that viewpoint, and fewer still do so in dynamic, unstructured environments. This study examined the cognitive state of robot operators performing complex survey and sample collection tasks in support of a time sensitive, high profile science expedition. A solo robot control paradigm was compared with a dual condition in which an alternate (surrogate) perspective was provided via voice commands to a second robot employed as a highly autonomous teammate. Subjective and neurophysiological measurements indicate an increased level of situational awareness was achieved in the dual condition along with a reduction decision oriented task engagement. These results are discussed in the context of mitigation potential for cognitive overload and automation induced complacency in complex and unstructured task environments.

Keywords Human robot interaction • Cognitive state • Situational awareness • Workload • Decision making • Robot assisted rescue

1 Surrogate Perspective(s) for Robot Control

The integration of an omniscient or “3rd Person” perspective into the video game and augmented reality industry has inspired a volume of research that is directly relevant to multi-robot control. Salamin and colleagues, for example, report that many video gamers prefer to use a third person perspective for moving an avatar through an artificial environment, but usually switch back to a 1st person viewpoint from within the avatar itself to perform “thin” tasks where dexterous manipulation

J.G. Blich (✉)

AFRL 711th HPW/RHC, Wright Patterson AFB, Ohio, USA

e-mail: john.blich@us.af.mil

© Springer International Publishing Switzerland 2017

P. Savage-Knepshild and J. Chen (eds.), *Advances in Human Factors in Robots*

and Unmanned Systems, Advances in Intelligent Systems and Computing 499,

DOI 10.1007/978-3-319-41959-6_34

is required [1]. Similar preferences have been reported in teleoperation of mobile robots [2, 3] and the use of augmented reality for remote grasping and manipulation [4, 5].

Despite an abundance of research published on Human Robot Interaction (HRI) there appears to be a dearth of literature documenting the use of a robotic team mate to provide this 3rd “person” perspective. Note that since an external offset of this nature can be provided by any number of non-human devices, the term “surrogate perspective” is used in lieu of the “3rd “(or 4th) person label. In turbulent conditions where sediment, dust or other obscurants interfere with the primary robot’s sensors, a mobile offset or “surrogate” perspective can potentially save the mission from total failure by finding a gap in the obscurant cloud to exploit and guide the primary platform and/or its manipulator(s) from a different angle. This approach has the potential to enable and expand upon an otherwise ineffective object identification or retrieval task. Multiple perspectives also present the potential for stereo-sensing approaches that offer far more capability in handling dynamic targets and environments than mono-sensing.

Given these considerations, this investigation reflects a seedling investment in the SPAARC, or Surrogate Perspective for Adaptive and Agile Robot Control concept. Its goal was to determine quantitative measures of surrogate perspective value (or burden as the case may be) derived from teaming a highly autonomous robotic teammate with a human tele-operator in a challenging survey and manipulation task. This particular effort explored the SPAARC concept in an underwater context with Remotely Operated Vehicles (ROVs) in order to slow down human robot interactions for close examination in a relatively viscous operating realm while exploiting the inherent risks and stressors associated with NASA’s Pavilion Lake Research Project (PLRP).

2 Research Environment and Expectations

This research was conducted in an underwater obstacle course set up in the diving pool at the Wright State University Natatorium in Dayton OH and aboard a 24’ research vessel on Canada’s Pavilion Lake located in British Columbia between the towns of Lillooet and Cache Creek. This unique glacial feature has been investigated by NASA and the Canadian Space Agency (CSA) for more than a decade, primarily because of its role in hosting the development of microbialite colonies that are of direct interest as an extremophile species analog for exploration missions of Mars and other heavenly bodies. For more details on this prominent research endeavor see Lim et al. 2011 [6].

Unlike other Mars analog simulations and research endeavors that involve mock sample collection and “analysis” of placeholder objects, the Pavilion Lake Research Project (PLRP) involves a challenging operational profile conducted against a backdrop of “real” scientific activity. The ambitious research carried out in this project creates a situation where the many scientists, astronauts, and NASA

engineers involved in it depend upon actual technical analysis results obtained for published research and professional stature. As such, the need for delicate handling of samples collected and minimal disturbance of the environment creates a daunting challenge for robot operators attempting to support the project with Remotely Operated Vehicles (ROVs).

The self-imposed stress for robot operators trying to impress world renown scientists, astronauts and engineers with their dexterity in handling delicate tasks of this nature is of direct interest to the military commensurate with minimally invasive demands for special reconnaissance and covert sample collection activity in denied areas. In that context, this investigation offered an unprecedented opportunity for the Air Force Research Laboratory to examine the cognitive state and mental workload experienced by mobile robot operators performing complex tasks under stressful conditions in and unstructured environment far beyond what is typically available in a laboratory setting.

The simulation vignette used in this research was a modified version of the robot assisted sample collection mission enacted by Blicht and colleagues during the 2001 Mars Flashline project [7]. This simulation involved a robot assisted sample collection EVA (Extra Vehicular Activity) in which rovers were first used to survey an area of interest and collect (i.e. pick up) readily available samples in the immediate vicinity of the target. It is important to note that the term “readily available” in this context refers to sample collection that did not involve highly sensitive procedures requiring human dexterity or tactile intuition such as gently prying protrusions loose from rock strata, or lightly chipping away samples with a geologist’s pick. Human exploration parties then proceeded into the same area by manned rovers (simulated by ATVs) with limited environmental resources to focus their intuitive sample collection and in situ analysis efforts on specific items of interest established by examination of data obtained during the robotic pre-cursor missions.

A 24’ research vessel (modified pontoon boat) was used to simulate a local habitat on Mars (or in orbit) from which rovers (simulated by submersible ROVs) were dispatched commensurate with mission objective stated above. An Earth based command and control center was simulated by NASA KSC’s Mobile Mission Control Center (MMCC) located on land along the northeastern shoreline of Pavilion Lake. Data transfer delays and bandwidth limitations replicating typical positional accuracy challenges associated with space exploration operations of this magnitude were simulated by a sophisticated software package and communications network connecting the MMCC with the forward research base (pontoon boat) and shuttle craft (simulated by other boats) carrying manned EVA teams (divers) between various exploration sites. The reader is directed once again to Lim et al. 2011 for details of this setup [6].

Returning now to relevant HRI literature, an abundance of research on cognitive demand associated with multi-robot control calls into question which way the cost benefit scale might tip if surrogate perspective(s) for a delicate sample collection operation were implemented in the form of mobile robots. The majority of this literature, however, endorses the role of adaptive automation as a powerful ally in

mitigating the mental workload and attentional demands associated with multi-robot control [8, 9]. With that in mind, the second robot intermittently injected into this rescue scenario was simulated at a high level of autonomy and artificial intelligence sufficient to respond effectively to voice commands.

Given the naturalistic interface between human and robotic teammate implemented above it was expected that ROV operators attempting challenging sample collection tasks in dynamic, unstructured environments would perform better when a surrogate perspective from that teammate was offered in a DUAL condition compared to singular robot trials made in a SOLO condition. In addition to this primary (H1) hypothesis, it was expected that operators in the DUAL condition would also report increased situational awareness (H2) and a reduction in mental demand (H3) compared to the SOLO condition. It was also expected that neurophysiological measures would echo the self-reported increase in awareness (H4) and cognitive workload (H5). Given the inherent difficulties associated with unpredictable and unstructured field environments, undesirable drops in performance (H6) and situational awareness (H7) were expected during operations conducted on Pavilion Lake along with an increase in workload (H8) compared to pool measurements—all of which would be mitigated in the DUAL condition as evidence of surrogate perspective value.

3 Method

3.1 Participants

Four male military personnel from the 88th Security Forces Squadron at Wright Patterson Air Force base were selected from a field of 13 other volunteers to participate in this research based on their performance during a 4-h assessment and selection process described below. As active duty military service members, these participants ranged in age from 21 to 30 years of age, had normal hearing and normal or fully corrected eyesight, and received no monetary incentive beyond standard travel reimbursements.

3.2 Apparatus

The mobile robot manipulators used for this experiment were Video Ray submersibles equipped with a simple two Degree Of Freedom (DOF) gripper and an Outland 1000 ROV fitted with 3-jawed claw for small object retrieval. Details and performance specifications for each of these platforms are available at each company's website. The Outland ROV was also fitted with a Tracklink Transponder that communicated its relative position underwater to a monitor aboard the research



Fig. 1 Video Ray Pro-III Submersible and B-Alert wireless EEG system worn by ROV operators supporting NASA's Pavilion Lake Research Project (shown during pool and field phases on the *left* and *right* respectively)

vessel. Neurophysiological data was collected with a wireless brainwave (or EEG for ElectroEncephaloGraphy) monitoring device called the X-10 B-Alert system manufactured by Advanced Brain Monitoring Inc. This data was recorded on ruggedized Getac V110 laptops. Video data of participant control activity was captured on 2ea QSEE color cameras mounted on each ROV control console as indicated in Fig. 1.

3.3 Measures

This investigation endeavored to examine the influence of a secondary robot's surrogate perspective in terms of the cognitive state—a representation of how involved the operator was with the task environment, and mental workload which exemplifies cerebral intensity. Given the non-invasive real time advantages of neurophysiological measurement promoted by Parasuraman and Wilson [10] as well as recent findings by Matthews and colleagues that measures of cognitive load don't necessarily converge [11], a combination of subjective and EEG based methods were used to examine these aspects of cognition.

Neurophysiological data was collected using the B-Alert system described above. This headset acquires 9 channels of EEG collected along the scalp at standard locations: Fz, Cz, POz, F3, F4, C3, C4, P3, and P4. The proprietary acquisition software used in this process includes artifact decontamination algorithms for eye blink, muscle movement, and environmental/electrical interference such as spikes

and saturations. The software also contains a cognitive state and workload classification package that compares EEG collected in field conditions with baseline data acquired while participants perform a series of computer-based vigilance tasks lasting from 15 to 30 min under nominal conditions in a controlled setting.

The first baseline task requires the participant to remain vigilant while choosing between different symbols presented on a laptop display. The second task requires a simple keyed response to a single stimulus (a red circle appearing on the screen) without any choice or decision-making process involved. The third task requires the participant to respond in a similar fashion to the second task, but with an auditory stimulus presented while their eyes were closed.

EEG classification for this experiment was conducted post hoc via comparisons with these three baseline profiles and a large EEG database of typical human sleep and distraction patterns. Classifications are presented as second by second probabilities that each epoch (or second) of data matched the participant's Hi (or decision based) engagement state established in the first (3-choice) baseline task, a Lo (or awareness based) engagement state observed during the second (eyes open) baseline task, or a distracted state determined by the third (eyes closed) baseline task. EEG workload classification is conducted in a similar fashion by comparing elements from each participant's 3-choice baseline task with a large data base of typical human performance on forward and backward span tasks. See Berka et al. 2004 for more detail on this classification process [12].

Subjective measures of cognitive state were collected using a popular situational awareness survey validated by Adams and colleagues on behalf of the U.S. Navy called the China Lake Situational Awareness (CLSA) scale [13]. This survey essentially asks the participant to rate their awareness of what was going on around them during recent activity based on criteria categorized on a 5-point scale from 1 (Very Good) to 5 (Very Poor). Subjective measurement of workload was measured in a similar fashion using the first (Mental Demand) factor of the NASA Task Load Index or TLX—a popular instrument in the human factors field for more than two decades [14]. Both of these instruments were administered to participants immediately upon completion of each ROV training/evaluation trial and field session. Scores were then normalized to a common 100-point system following a higher-as-better paradigm for consolidated graphing with other metrics.

3.4 Design and Procedure

Upon completion of informed consent documentation, participants were administered a standard spatial ability battery [15] and spatial orientation survey [16]. Evaluations were conducted commensurate with the findings of Blich and colleagues that these instruments were effective in predicting robot teleoperation effectiveness for semi-autonomous mobile robots engaged in a similar mission set during the NASA Haughton Mars Project in 2001 [7]. Participants were then outfitted with a B-Alert device and asked to perform the baseline tasks described

above before performing 8 training/evaluation trials in an underwater obstacles course set up in a local pool (see Fig. 1.). Each evaluation trial required the participant to “fly” their ROV through the underwater obstacle course, retrieve a target object (4" cylindrical buoy with a weight attached by fishing line) sitting on the bottom at the opposite corner of the pool. They were then required to return through the obstacle course and drop the target inside a bucket located at the pool corner near the start point.

Every participant completed eight trials alternating between SOLO and DUAL conditions. These trials were also counterbalanced in sequence between participants to avoid experiential bias. The surrogate ROV operator in the DUAL condition was required to simulate a highly autonomous and intelligent robot in support of the primary operator’s target retrieval/sample collection tasks. As such, the surrogate operator was not allowed to speak to the primary operator at any time, and was only allowed to activate the controls on his ROV in the DUAL condition when the primary operator spoke instructions from a set of simple scripted commands such as descend, ascend, turn right/left, forward/backward, etc. Due to a critical equipment malfunction experienced early in the assessment and selection process at the pool, evaluation trials under the DUAL condition had to be conducted with a hand held pan/tilt video camera walked along the side of the pool to simulate the surrogate ROV maneuvering back and forth and/or holding station above a sample collection site.

The top four performers from these evaluation sessions were chosen for deployment to Canada in support of NASA’s sample collection and scientific analysis efforts at Pavilion Lake. Participants were fitted with neurophysiological instruments and administered baseline tasks before embarking upon the simulated habitat research boat for a 30-min ride to whatever portion of the lake was chosen for survey/study by the science analysis team in the MMCC. Upon arrival at the survey site, each operator received instructions from the IV (Inside Vehicle) supervisor responsible for the particular EVA mission at hand. ROV operators (or “pilots”) were asked to fly over and around obstacles taking extra care to avoid tether entanglement with a variety of underwater hazards as well as diver (astronaut) umbilical cables that were required to simulate typical EVA radio communications from underwater. ROVs were occasionally used to retrieve survey markers as well, but the vast majority of biological samples of interest to the science team were deemed far too fragile and friable to handle with mechanical grippers.

4 Results

4.1 *Assessment and Selection (Pool) Phase*

Unfortunately a combination of equipment malfunctions, umbilical entanglements, and synchronicity challenges made it impractical to evaluate operator performance on a rigorously quantitative (trial time) basis. It was possible however, to account

Table 1 Engagement, awareness and workload statistics for pool phase

	Solo		Dual		Cohen's		
	M	SD	M	SD	$t(15)$	p	d
EEG Hi Eng.	0.536	0.150	0.513	0.142	2.298	0.036	0.159
EEG Lo Eng.	0.302	0.043	0.321	0.046	1.989	0.065	0.425
CLSA Score	75	18	81	17	1.865	0.082	0.354
TLX MD	49	18	50	17	1.101	0.288	-0.080
EEG Wkld.	0.658	0.064	0.657	0.059	0.770	0.453	0.012

(Significant values shown underlined in bold type with a relaxed alpha of 0.10)

for these anomalies in the EEG and subjective instrument data (albeit to a slightly lesser degree with the latter due to recency bias potential). The EEG data was averaged across all epochs in each simulation session and analyzed alongside subjective survey data as indicated in Table 1.

Given that all data was collected on the same participants across temporal phases with and without a robotic teammate, initial analysis was performed using a paired t-test with a relaxed alpha set at 10 to accommodate the inherent variance involved with data collected under the anomalous “noisy” circumstances described above. Considering cognitive state first, participants recorded a moderately higher probability of engagement in the “Hi” or “choice” based metric for the SOLO condition than when a teammate with a surrogate perspective was present in the DUAL condition. The data collected also indicate a slightly higher probability that ROV operators were engaged in the “Lo” or “awareness” cognitive state with a small effect size. This trend was followed in the subjective situational awareness data as well, with higher CLSA scores recorded in the DUAL condition, again with a moderate effect size. No significant differences in subjective or EEG based workload was observed between groups.

4.2 Field Phase at Pavilion Lake

The biological research objectives that dominate NASA's Pavilion Lake Research Project typically create a dynamic and unstructured environment for ROV operators providing robotic survey, sample collection, and astronaut (diver) assistance. Although it was hoped that a sufficient number of dives pursued at Pavilion Lake would be similar enough in scope, environment, and task conditions to perform SOLO v. DUAL comparisons of operator cognitive state and workload, that turned out not to be the case. Enough similar lake dives were conducted in the SOLO condition, however, to enable an isolated comparison between the relatively well organized pool environment and the dynamic, unstructured NASA operating environment in Canada. Given the independent nature of the data collected under these conditions, analysis was conducted via the MANOVA process with statistics

Table 2 Engagement, awareness and workload statistics SOLO condition in lake and pool phases

	Lake		Pool		<i>Partial</i>		
	M	SD	M	SD	<i>F</i> (1)	<i>p</i>	<i>eta</i> ²
EEG Hi Eng.	0.454	0.246	0.536	0.150	0.848	0.369	0.043
EEG Lo Eng.	0.308	0.128	0.302	0.043	0.029	0.866	0.002
CLSA Score	<u>50</u>	<u>31</u>	<u>75</u>	<u>18</u>	<u>5.170</u>	<u>0.035</u>	<u>0.214</u>
TLX MD	<u>68</u>	<u>15</u>	<u>49</u>	<u>18</u>	<u>8.849</u>	<u>0.008</u>	<u>0.318</u>
EEG Wkld.	<u>0.781</u>	<u>0.126</u>	<u>0.658</u>	<u>0.064</u>	<u>4.740</u>	<u>0.042</u>	<u>0.200</u>

(Significant values shown underlined in bold type with alpha of 0.05)

listed in Table 2. While no significant differences in engagement data were noted, a substantial decrease in situational awareness and prominent increase across both workload metrics was observed in the transition from pool to lake environment with moderately large effect sizes throughout.

5 Discussion, Conclusions, and Future Work

Although none of the data collected in either phase provided any support for the primary performance improvement hypotheses (−H1, −H6), the pool data provided substantial evidence that a desirable increase in situational awareness (+H2, +H4) is possible with implementation of the SPAARC concept. While a reduction in the EEG Hi Engagement suggests that a lower probability of choice-based neurological activity was evident in the DUAL condition, the lack of any appreciable difference in subjective workload (−H3) or EEG based workload (−H5) precludes any direct support for the claim that a surrogate perspective can reduce cognitive load. It is interesting, however, that a prominent increase in all three decision/workload oriented metrics was evident in the transition to the dynamic and unstructured Pavilion Lake environment. This suggests the possibility that an early learning plateau reached during repetitive trials administered in the assessment phase may have prevented such an advantage to be revealed in the pool setting. Unfortunately, the lack of a crisp performance comparison (−H7) provides no direct support for that notion beyond anecdotal observation. The prominent drop in situational awareness recorded in the transition to the lake condition, however, provides optimism that more support for SPAARC is possible if a SOLO/DUAL comparison can be conducted in unstructured and unpredictable environments.

Taken across both phases, the neurophysiological and subjective data support approximately half of the expected results, thereby providing a modicum of evidence for SPAARC concept value. That said, it must be pointed out that statistical support for this claim is weak in terms of sample size, and a relaxed alpha invoked during analysis of pool data to accommodate equipment malfunctions and tether drag issues mentioned above. Despite this acknowledged increase in likelihood of committing a Type I error during this initial investigation, the relatively large effect

sizes observed suggest that replication would be quite worthwhile in a manner consistent with Wickens' Common Sense Statistics [17].

Assuming that the SPAARC concept is eventually validated and/or expanded beyond this effort to include air-ground collaborative systems, the implications can be profound for multi-robot control. Not only do these findings provide ample room for optimism regarding resilient and adaptive human robot team performance in stressful, unstructured environments, but they also show promise for the pursuit of enhanced trust and confidence between humans and machines championed by Hancock and colleagues [18]. It should also be noted that although real time cognitive state classifications were not used in this study, the B-Alert system is capable of providing them in near real time (with a 1–5 s delay depending on signal characteristics and display options). Detection of brain state changes in this manner presents a powerful mitigation option in that intervention strategies can be cued and implemented *before* performance suffers from cognitive overload or *before* complacency becomes evident in operator behavior.

Acknowledgments This work was sponsored by the Warfighter Interface Division (RHC) of the 711th Human Performance Wing at the Air Force Research Laboratory. The author would like to express the utmost in profound gratitude to Dr. Darlene Lim of NASA/JSC and her amazing team for providing 711HPW/RHC with such a magnificent research opportunity. Boundless appreciation is also extended to Arnis Mangolds, Mark Micire, James Christensen, Justin Estepp, Ethan Blackford, Maggie Bowers, Jeffrey Bolles, and Samantha Klosterman for their extra efforts and technical prowess in handling complex data collection and participant management issues under daunting conditions and a compressed time line.

References

1. Salamin, P., Thalmann, D., Vexo, F.: The benefits of third-person perspective in virtual and augmented reality? In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology. ACM, (2006)
2. Milgram, P., Rastogi, A., Grodski, J.J.: Telerobotic control using augmented reality. In: Proceedings of 4th IEEE International Workshop on Robot and Human Communication. RO-MAN'95 TOKYO. IEEE (1995)
3. Okura, F., et al.: Teleoperation of mobile robots by generating augmented free-viewpoint images. In: International Conference on Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ. 2013. IEEE
4. Leeper, A.E., et al.: Strategies for human-in-the-loop robotic grasping. In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction. ACM (2012)
5. Hashimoto, S., et al., Touchme: an augmented reality based remote robot manipulation. In: Proceedings of ICAT2011 21st International Conference on Artificial Reality and Telexistence (2011)
6. Lim, D.S., et al.: A historical overview of the Pavilion Lake research project—analogue science and exploration in an underwater environment. *Geol. Soc. Am. Spec. Pap.* **483**, 85–115 (2011)
7. Blich, J.G., et al.: Correlations of spatial orientation with simulation based robot operator training. In: 4th International Conference on Applied Human Factors and Ergonomics (AHFE). San Francisco CA, 2012

8. Parasuraman, R., Hancock, P.: Mitigating the adverse effects of workload, stress, and fatigue with adaptive automation. *Performance Under Stress* (Epub), p. 45, 2012
9. Parasuraman, R., Cosenzo, K.A., De Visser, E.: Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. *Mil. Psychol.* **21**(2), 270 (2009)
10. Parasuraman, R., Wilson, G.F.: Putting the brain to work: neuroergonomics past, present, and future. (Cover story). *Hum. Factors* **50**(3), 468–474 (2008)
11. Matthews, G., et al.: The psychometrics of mental workload multiple measures are sensitive but divergent. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **57**(1), 125–143 (2015)
12. Berka, C., et al.: Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *Int. J. Hum.-Comput. Interact.* **17**(2), 151–170 (2004)
13. Adams, S., Kane, R., Bates, R.: Validation of the China Lake situational awareness scale with 3D SART and S-CAT. China Lake, CA: Naval Air Warfare Center Weapons Division (452330D), 1998
14. Hart, S.: NASA-task load index (NASA-TLX); 20 years later. *Ann. Meet. Hum. Factors Ergon. Soc.* **50**(9), 904–908 (2006)
15. Newton, P., Bristoll, H.: Psychometric success spatial ability practice test 1. pp. 1–12, (2010)
16. Kozhevnikov, M., Hegarty, M.: A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cogn.* **29**(5), 745–756 (2001)
17. Wickens, C.D.: Statistics. *Ergon. Des. Q. Hum. Factors Appl.* **6**(4), 18–22 (1998)
18. Hancock, P.A., et al.: A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **53**(5), 517–527 (2011)

A Method for Neighborhood Gesture Learning Based on Resistance Distance

Paul M. Yanik, Anthony L. Threatt, Jessica Merino, Joe Manganelli,
Johnell O. Brooks, Keith E. Green and Ian D. Walker

Abstract Multimodal forms of human-robot interaction (HRI) including non-verbal forms promise easily adopted and intuitive use models for assistive devices. The research described in this paper targets an assistive robotic appliance which learns a user's gestures for activities performed in a healthcare or *aging in place* setting. The proposed approach uses the Growing Neural Gas (GNG) algorithm in combination with the Q-Learning paradigm of reinforcement learning to shape robotic motions over time. Neighborhoods of nodes in the GNG network are combined to collectively leverage past learning by the group. Connections between nodes are assigned weights based on frequency of use which can be viewed as measures of electrical resistance. In this way, the GNG network may be traversed based on distances computed in the same manner as resistance in an electrical circuit. It is shown that this distance metric provides faster

P.M. Yanik (✉)

Department of Engineering and Technology, Western Carolina University,
Cullowhee, NC, USA

e-mail: pyanik@wcu.edu

A.L. Threatt · J. Manganelli · K.E. Green

School of Architecture, Clemson University, Clemson, SC, USA

e-mail: anthont@clemson.edu

J. Manganelli

e-mail: manganelli@clemson.edu

K.E. Green

e-mail: kegreen@clemson.edu

J.O. Brooks

Department of Psychology, Clemson University, Clemson, SC, USA

e-mail: jobrook@clemson.edu

J. Merino · I.D. Walker

Department of Electrical and Computer Engineering, Clemson University,
Clemson, SC, USA

e-mail: jmerino@clemson.edu

I.D. Walker

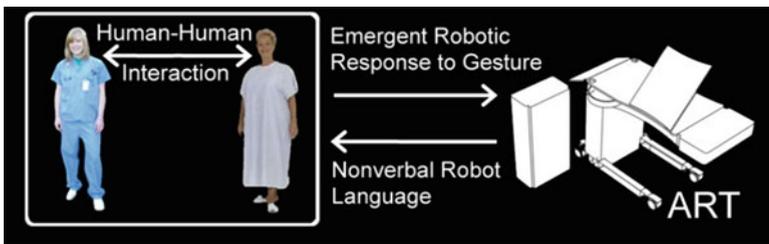
e-mail: kegreen@clemson.edu

convergence of the algorithm when compared to shortest path neighborhood learning.

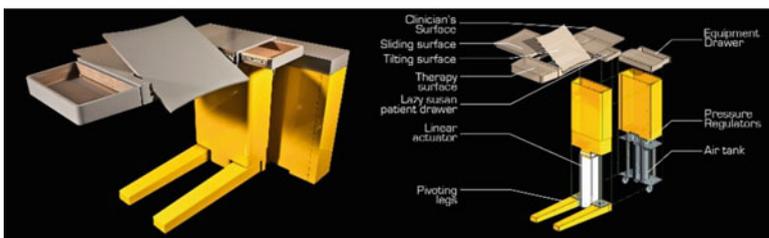
Keywords Machine learning · Gesture recognition · Human-Robot interaction · Resistance distance

1 Introduction

Multimodal human-robot interaction (HRI) including non-verbal forms comprises a very active area of research with relevance in ubiquitous computing and assistive robotics. In particular, assistive devices are seen as critical to meeting the needs of unskilled, impaired, or aging individuals whose abilities are changing over time. Communication modalities such as gesture and eye gaze offer natural, intuitive, and easily adoptable use models. The work described in this paper is motivated by presence of few viable systems in this problem space. The Assistive Robotic Table (ART) project begun at Clemson University is an effort which seeks to develop a class of intelligent devices that are well-integrated into the built environment and that supports users whose abilities may be evolving with age. ART is a robotic over-the-bed table as may be found in a healthcare setting (Fig. 1).



(a)



(b)

Fig. 1 **a** The non-verbal communication loop between the Assistive Robotic Table and the user. ART adapts to the desires of the user through repeated interaction via gesture. **b** A recent project artifact. ART has three degrees of freedom including a vertical lifting column, a *horizontal* sliding table top, and a tilting work surface [1]

Typically, designers of such systems implement detection schemes for interpersonal communication messages by comparing them to one of a known gallery of prescribed message forms. However, in order to accommodate the changing abilities of the candidate user population, a communication interface must be capable of learning new *unchoreographed* message types while recognizing those already learned as the user's ability to perform them evolves. This paper extends past work by the authors in the area of gesture learning and recognition as a means of non-verbal HRI.

The approach and experimentation described herein extend previous work by the authors [1] to incorporate resistance distance to the learning of three candidate gestures: *come closer*, *go away*, and *stop*. Gesture forms are selected from the American Sign Language dictionary for repeatability across the pool of participant volunteers. However, as previously stated, no specific choreography is required. It is shown that this metric yields favorable results over a simple a shortest path distance calculation using Floyd's algorithm [2]. This gesture-based interface construct is seen as the basis of a future task-learning paradigm which recalls sequences of configurations to assist the user through daily activities and to track changes in behavior so as to prevent injury and to inform care planning.

This paper is organized as follows. Sect. 2 discusses related past research in the area of machine learning and its influence on the present research. Section 3 discusses key algorithms used for experimentation, the data collection fixture used to capture gestures, and the simulation platform used to test the approach. Section 4 discusses experimentation and results. Conclusions are presented in Sect. 5.

2 Related Work

Human gesture in the form of hand and arm gesticulation accounts for some 90 % of gesture-based communication [3]. As such, the use of gesture as the foundation of a robot command interface offers the prospect of a natural communication modality between assistive devices and their users.

In order for an assistive device to make use of gesture, it must possess a vocabulary of recognized gestures and map them to specific robotic configurations or motion sequences. This mapping between a sensed event and a motor actuation (termed *sensorimotor* mapping) has been used successfully in previous research. Even common gestures, however, are typically performed with high variability across the user population, or even by an individual user. Methods for making crisp classifications in such widely varying sensory input have frequently made use of neural network based approaches. Ritter et al. [4] and Martinetz [5] made use of Kohonen's Self-Organizing Feature Map (SOFM) to discretize the input space into receptive fields surrounding neurons (*reference nodes*). Each reference node was mapped to a distinct motor actuation.

Because of the nature of the SOFM to preserve topological relationships between input variations, the network learns faster than conventional neural networks.

This owes to the idea that similar inputs should tend to produce similar outputs. Hence, topological neighbors in an SOFM which encode similar sensor input patterns can be collectively made to learn desired output actuations as a group. Walter and Schulten made use of this concept applying the similar Neural Gas (NG) algorithm with *softened* Gaussian receptive fields for smoother learning [6]. These approaches, however, presuppose a simple and specific actuation for each mapping.

In dealing with gestures, the desired actions of the robotic agent may not be known a priori. Hence, a learning phase for the actions themselves must be undertaken. Reinforcement Learning (RL) techniques are also applied to the problem of robot control and are applicable to these situations where the desired action is unknown. In particular, the RL framework of *Q-Learning* [7] is often used for its ease of implementation, and because the system design is not required to maintain a model of the environment. With Q-Learning, an agent in a particular state s may choose an action a from of a set of available actions. Following each operating trial, state-action pairs which resulted in desirable outcomes are assigned a reward. Rewards accumulate (as the *Q value*) over multiple trials so that an operating *policy* develops to maximize total reward. State-action pairs with the highest Q values are expected to yield the most successful outcomes.

Touzet [8] combined the features of Kohonen's SOFM with Q-Learning in the *Q-Kohon* approach. By preserving the topological relationships of input data space, *neighborhoods* of states may leverage past learning by members of the group so as to speed convergence to an appropriate input-output mapping across varied input patterns. When a gesture sample falls within the receptive field of a given reference node, all nodes within the same neighborhood may be examined to find actions with the highest Q value. The approach used in this paper adapts Q-Kohon using the Growing Neural Gas (GNG) algorithm [9]. GNG is adopted for its capability of adding and subtracting nodes (as gestures may be added or subtracted to an agent's learning base) and to track a nonstationary distribution of input gesture forms.

Neighborhoods of reference nodes may be formed in various ways. Typically, Euclidean distance between input samples is used (nearest neighbor approach). However, the connectivity of the GNG network and its ability to allow connections to *age* with lack of use provides another avenue for determining neighbors. In this research, two graph-based algorithms are compared for their ability to locate neighbors whose state-action mapping is most worth emulating: Floyd's shortest path algorithm [2], and Klein and Randic's [10] resistance distance algorithm. These are described in detail in Sect. 3. Graph-based approaches show promise in their ability to consider complex network topology in the search for high-reward state-action pairs.

3 Method

This section describes the laboratory fixture used to collect gesture samples, the system constructed to process samples, and the learning algorithms employed.

3.1 Sensing

Gesture samples were collected using the Microsoft Kinect RGB-D sensor [11]. The RGB-D sensor type was selected for its ability to collect gesture data non-intrusively (preserving user privacy) and to work in darkness. Gestures were performed by volunteer participants who completed informed consent forms in accordance with an IRB-approved study procedure. The Kinect was used to capture 3D coordinates of each participant’s left hand over a 5-s interval at 30 Hz for a total of 150 data points per gesture performed.

3.2 Gestures

The gestures performed included *come closer*, *go away*, and *stop*. These were selected for their relevance to common commands which might be useful for any mobile assistive device. Motion models for these gestures were taken from the American Sign Language video dictionary [12]. Although no specific choreography of gestures was required, the ASL dictionary provided a template for participants to perform the gestures in a roughly consistent manner. Goal configurations of the robotic agent are given by Table 1. These configurations represent the three degrees of freedom of the ART appliance. For simulation purposes, they were visualized as final configurations in the Robot Operating System (ROS) Turtlesim environment and are scaled accordingly. Turtlesim is a facility within ROS for developing simple robot control programs with the same three degrees of freedom: translation in 2D space (x , y) and angle of approach θ .

3.3 Data Representation (Feature Extraction)

Motion features based on maxima of acceleration known as *Dynamic Instants* (DI) as described in were used to generate compact feature vectors for each gesture sample. DIs have been shown to be view-invariant. Gesture data was first smoothed using a discrete Gaussian kernel:

$$G = [1, 4, 6, 4, 1]/16. \tag{1}$$

Table 1 Goal configurations for gestures

Gesture type	Goal configuration (x , y , θ)
Come closer	3.95 m, 3.95 m, 45°
Go away	3.95 m, -3.95 m, 315°
Stop	-3.95 m, -3.95 m, 225°

A moving average (evolution time) of 7 was also used as an added smoothing step. A feature vector consisting of the (x, y, z) coordinates of five DIs of greatest magnitude (scaled to $[0,1]$), plus their temporal frame numbers, were constructed to represent each gesture sample (a 20×1 feature vector).

3.4 The Growing Neural Gas (GNG) Algorithm

The basic GNG algorithm was first proposed by Fritzke [9]. The GNG network is a variant of the SOFM with the added capability to add or remove reference nodes based on their frequency of activation with a given input distribution. Our adaptation of GNG also includes a configuration field for each node in the network node list. This field is interpreted as the desired configuration for the robot in response to a given input gesture. In this way, our implementation avoids the task of labeling gestures in favor of a mapping that develops over time in response to user preference. The assignment of user preference is described below.

3.5 Q-Learning

In a reinforcement learning framework, an agent attempts to learn an optimal policy of mapping its current states to optimal actions that will yield the greatest rewards through a chain of state-action pairs. In our implementation of the Q-Learning paradigm, a simple *good/bad* indicator is used as reward with numerical values $+1$ or -1 respectively.

A third reward value of 0 is also used when a given state is fully trained and requires no more updates. The reward is added as a linear step in physical configuration space of the robot. Thus, accumulated reward is interpreted as a vector into the configuration space from a global origin. Longer vectors are indicative of greater reward as assigned by the user. Vectors for which the agent has moved farther from the goal than they had in the past receive a reward of -1 , and thus, an exploration (randomized) step is applied. The reward function can be stated as:

$$Q(s, a)_{t+1} \leftarrow r(\text{stepLength}) + \max_a Q_t(s', a). \quad (2)$$

where r is reward, stepLength is 0.1 m and Q is the expected reward for a given state s and action a .

3.6 Floyd's Shortest Path Algorithm

Application of Floyd's algorithm [2] returns a matrix of shortest paths between all vertex pairs ($D_{n \times n}$) in an undirected graph. In our implementation, vertices are reference nodes in the GNG network, and the age parameter for each connection between nodes is used as the edge length. The algorithm is given by the pseudocode below.

```

1: Initialize  $D_{n \times n}$  with all  $d_{ij} = 1$ .
2: for  $k = 1$  to  $n$  do
3:   for  $i = 1$  to  $n$  do
4:     for  $j = 1$  to  $n$  do
5:        $d_{ij} = \min\{d_{ik} + d_{kj} ; d_{ij}\}$ 
6:     end for
7:   end for
8: end for

```

3.7 Resistance Distance

The resistance distance algorithm [10] returns a matrix of path lengths between all pairs of vertices in an undirected graph. Distances are computed as the electrical resistance between points where edge lengths are interpreted as resistances. As with Floyd's algorithm, vertices are reference nodes in the GNG network, and the age parameter for each connection is the resistance. The resistance distance (Ω) is computed as

$$\Omega_{ij} = \begin{cases} L_{ii}^+ + L_{jj}^+ - 2L_{ij}^+, & i \neq j \\ \infty, & i = j \end{cases} \quad (3)$$

where L^+ is the Moore-Penrose generalized inverse of the graph Laplacian L

$$L = K - Av \quad (4)$$

Av is the admittance matrix

$$Av_{ij} = \begin{cases} \frac{1}{r_{ij}}, & i \neq j \\ 0, & i = j \end{cases} \quad (5)$$

where r is the edge length between vertices. K is the degree matrix

$$K = \text{diag} \left(\sum_{z=1}^n \frac{1}{r_{xz}} \right) \quad (6)$$

3.8 *Reward Generation*

Gestures are submitted to a simulated robotic agent. The GNG algorithm is executed and an action is selected from a neighborhood of candidate actions as previously described. The candidate whose action is selected is that of the nearest node having a Q value greater than the reference node for the sample. A simulated user observes the action and issues a reward of +1 (the agent moved closer to the desired goal), -1 (the agent moved farther from the desired goal), or 0 (the agent has reached the goal and the node is fully *trained*). Based on the reward, the appropriate action vector in the GNG reference node is updated accordingly. This process is repeated for all gesture samples according to the experimentation procedure described below.

3.9 *Data Collection, Experimentation and Results*

Five volunteer participants each performed 50 samples for each of the three candidate gestures (*come closer*, *go away*, and *stop*). This yielded a data set of $5 \times 3 \times 50 = 750$ total samples. The data set was randomly divided into two groups. The first group consisting of 300 samples (100 of each gesture type) was used to pretrain the GNG network to the topology of the input data space. No action learning occurred during this pretraining phase.

A second group consisting of the remaining 450 samples (150 of each gesture type) was applied to the system. All 450 samples were applied as an *epoch*. For each sample, a candidate action was selected and performed, award was generated by the simulated user, and the action vector for the reference node associated with each input pattern was updated. The error for each sample was computed as the difference between the updated configuration of the simulated agent and the known goal for the gesture type. Following each epoch, the average error and cumulative error for each gesture type were also calculated. In this way, 250 epochs were conducted. The process was repeated using both Floyd's algorithm and the resistance distance method.

4 **Results and Discussion**

Results for simulations using Floyd's algorithm and the resistance distance metric are shown in Fig. 2. Although both methods are seen to converge to a low error between the desired goal configuration and the robotic agent's actions, Floyd's algorithm yields slowest convergence for the *go* and *stop* gestures. Resistance distance yields slowest convergence for the *come* gesture. Fewer perturbations (i.e. greater stability of results) in later epochs are seen for resistance distance. Speed and stability of convergence are seen as a key metric given the application with human users who may become fatigued or otherwise burdened by the need to

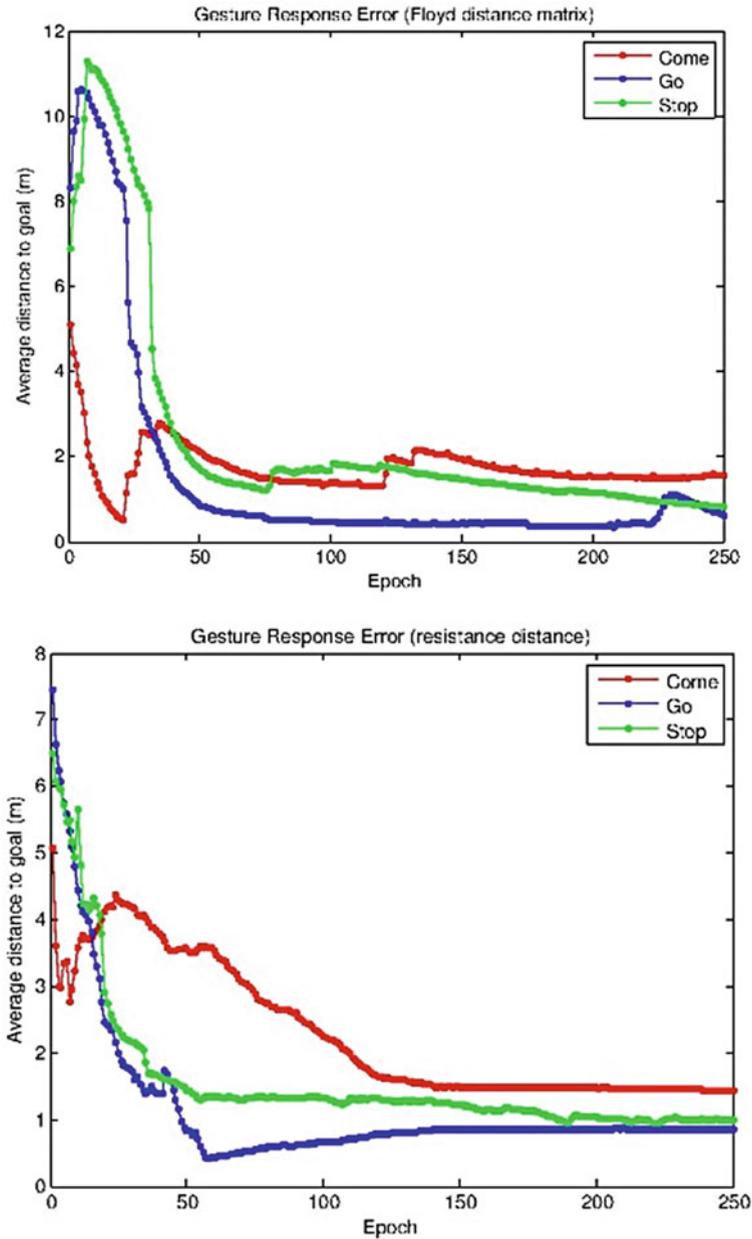


Fig. 2 Error curves for Floyd’s algorithm (top) and resistance distance (bottom)

provide extensive supervisory feedback to the system. Further, resistance distance produces the smallest cumulative error over the training session: 1253.6 m for resistance distance; 1423.3 m for Floyd’s algorithm (a 13.5 % disparity).

It is noted that, for well-separated data, network distance metrics outperform Euclidean distance metrics. However, given the wide variation in human performance of gesture (i.e. poor separation), past research [9] has shown that network distance metrics may underperform Euclidean metrics. This work shows general promise for neighborhood learning and suggests the need for improved edge-weighting schemes.

5 Conclusions

A method for neighborhood gesture learning in a graph-based data topology has been introduced. A novel extension of the GNG algorithm with Q-Learning makes use of the concept of resistance distance to locate neighboring nodes for which past rewards are most promising. Results show the network distance metric of resistance distance offers a significant improvement over Floyd's algorithm when using the same edge weighting method. Future work will concentrate on development of improved edge-weighting schemes so as to circumvent the overall slowness in convergence due to poor separation in the input data set.

Acknowledgments This research was supported by the U.S. National Science Foundation under award IIS-SHB-116075.

References

1. Yanik, P., Merino, J., Threatt, A.L., Manganeli, J., Brooks, J.O., Green, K.E., Walker, I.D.: A gesture learning interface for simulated robot path shaping with a human teacher. *IEEE Trans. Hum.-Mach. Syst.* **44**(1), 44–54 (2014)
2. Mitra, S., Acharya, T.: Gesture recognition: a survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **37**(3), 311–324 (2007)
3. Martinetz, T., Ritter, H., Shulton, K.: Three-dimensional neural net for learning visiomotor coordination of a robot arm. *IEEE Trans. Neural Netw.* **1**(1), 131–136 (1990)
4. Walter, J., Shulton, K.: Implementation of self-organizing feature maps for modeling and control of robotic manipulators. *J. Intell. Robot. Syst.* **36**(1), 86–96 (1993)
5. Sutton, R.S., Barto, A.G.: *Reinforcement Learning: An Introduction*. Cambridge University Press (1998)
6. Touzet, C.F.: Neural reinforcement learning for behavior synthesis. *Robot. Auton. Syst.* **22**(3), 251–281 (1997)
7. Fritzke, B.: A growing neural gas learns topologies. *Adv. Neural Inf. Process. Syst.* **7**(7), 625–632 (1995)
8. Klein, D., Randic, M.: Resistance distance. *J. Math. Chem.* **12**, 81–95 (1993)
9. Microsoft Kinect Website, <http://www.xbox.com/en-US/xbox-360/accessories/kinect>
10. ASL Pro Website, <http://www.aslpro.com/cgi-bin/aslpro/aslpro.cgi>
11. ROS Website, <http://www.ros.org>
12. Rao, C., Yilmaz, A., Shah, M.: View-invariant representation and recognition of actions. *Int. J. Comput. Vis.* **50**(2), 203–226 (2002)