Advances in Intelligent Systems and Computing 595

Jessie Chen *Editor*

Advances in Human Factors in Robots and Unmanned Systems

Proceedings of the AHFE 2017 International Conference on Human Factors in Robots and Unmanned Systems, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Advances in Intelligent Systems and Computing

Volume 595

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.

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

Jessie Chen
Editor

Advances in Human Factors in Robots and Unmanned Systems

Proceedings of the AHFE 2017 International Conference on Human Factors in Robots and Unmanned Systems, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Editor
Jessie Chen
U.S. Army Research Laboratory
Orlando, FL
USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-60383-4 ISBN 978-3-319-60384-1 (eBook)
DOI 10.1007/978-3-319-60384-1

Library of Congress Control Number: 2017943030

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Advances in Human Factors and Ergonomics 2017

AHFE 2017 Series Editors

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

AHFE
International
Applied
Human Factors
and Ergonomics
Conference
2017
www.ahfe.org

8th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences

Proceedings of the AHFE 2017 International Conference on Human Factors in Robots and Unmanned Systems, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Advances in Affective and Pleasurable Design	*WonJoon Chung and Cliff (Sungsoo) Shin*
Advances in Neuroergonomics and Cognitive Engineering	*Carryl Baldwin*
Advances in Design for Inclusion	*Giuseppe Di Bucchianico and Pete Kercher*
Advances in Ergonomics in Design	*Francisco Rebelo and Marcelo Soares*
Advances in Human Error, Reliability, Resilience, and Performance	*Ronald L. Boring*
Advances in Human Factors and Ergonomics in Healthcare and Medical Devices	*Vincent G. Duffy and Nancy Lightner*
Advances in Human Factors in Simulation and Modeling	*Daniel N. Cassenti*
Advances in Human Factors and System Interactions	*Isabel L. Nunes*
Advances in Human Factors in Cybersecurity	*Denise Nicholson*
Advances in Human Factors, Business Management and Leadership	*Jussi Kantola, Tibor Barath and Salman Nazir*
Advances in Human Factors in Robots and Unmanned Systems	*Jessie Chen*
Advances in Human Factors in Training, Education, and Learning Sciences	*Terence Andre*
Advances in Human Aspects of Transportation	*Neville A. Stanton*

(continued)

(continued)

Advances in Human Factors, Software, and Systems Engineering	*Tareq Z. Ahram and Waldemar Karwowski*
Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries	*Paul Fechtelkotter and Michael Legatt*
Advances in Human Factors, Sustainable Urban Planning and Infrastructure	*Jerzy Charytonowicz*
Advances in the Human Side of Service Engineering	*Louis E. Freund and Wojciech Cellary*
Advances in Physical Ergonomics and Human Factors	*Ravindra Goonetilleke and Waldemar Karwowski*
Advances in Human Factors in Sports, Injury Prevention and Outdoor Recreation	*Tareq Z. Ahram*
Advances in Safety Management and Human Factors	*Pedro Arezes*
Advances in Social & Occupational Ergonomics	*Richard Goossens*
Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future	*Stefan Trzcielinski*
Advances in Usability and User Experience	*Tareq Ahram and Christianne Falcão*
Advances in Human Factors in Wearable Technologies and Game Design	*Tareq Ahram and Christianne Falcão*
Advances in Communication of Design	*Amic G. Ho*
Advances in Cross-Cultural Decision Making	*Mark Hoffman*

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 are 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 the 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 require 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 multisensory 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.

A total of eight sections presented in this book:

I. Requirements and Groundwork for the Next Generation of Robotics Systems
II. Advanced Displays and Intuitive Interfaces
III. Advances in Unmanned Aerial Vehicles (UAV) and Drones Research
IV. Control Using Multimodal Input
V. Tactile Applications for Enhanced Performance in Naturalistic Settings
VI. Understanding the Key Drivers of Human-Machine Trust: Lessons Learned and Future Directions
VII. Human-Robot Teaming
VIII. Virtual Reality Technologies for Training Applications and C2 Of Unmanned Systems

Each section contains research paper that has been reviewed by members of the International Editorial Board. Our sincere thanks and appreciation to the board members as listed below:

July 2017 Jessie Chen

Contents

Advanced Unmanned Systems: Requirements and Groundwork for the Next Generation of Robotics Systems

Advanced Displays and Intuitive Interfaces

Advances in Unmanned Aerial Vehicles (UAV) and Drones Research

Control Using Multimodal Input

Tactile Applications for Enhanced Performance in Naturalistic Settings

Understanding the Key Drivers of Human-Machine Trust: Lessons Learned and Future Directions

Human-Robot Teaming

Virtual Reality Technologies for Training Applications and C2 of Unmanned Systems

Advanced Unmanned Systems: Requirements and Groundwork for the Next Generation of Robotics Systems

A Model for Temperament and Emotions on Robots

Lyle N. Long(✉)

Pennsylvania State University,
233 Hammond Building, University Park, PA, USA
lnl@psu.edu

Abstract. This paper describes a mathematical/computational model of emotions and temperaments (or personalities). This model has been implemented on cognitive mobile robots. Emotions such as fear, anger, sadness, happiness, disgust, surprise, and others can be modeled, and can vary due to reinforcers (e.g. rewards and punishments). The model incorporates exponential decay of the reinforcement effects, so without continual reinforcers the emotion will return to their steady-state values. It is shown that emotions and temperament are coupled through the theory. Most models of emotions do not relate emotions to temperament. The constants used in the model of emotions are related to the temperament of the robot. The main five temperaments discussed include Extrovert/Introvert, Neurotic/Rational, Conscientious/Careless, Agreeable/Disagreeable, and Open/Reticent. The emotion and temperament model developed here has been incorporated into SS-RICS, which is a cognitive architecture developed at the Army Research Laboratory and tested in both mobile robots and in simulators.

Keywords: Robotics · Emotions · Temperament · Affective

1 Introduction

Emotions and temperament are crucial for animal survival (including human). Emotions allow rapid behavior adjustments to changing circumstances. Also, group survival is enhanced by having a mixture of temperaments. Temperament (or personality traits) and emotions are not the same thing. Temperaments are traits that an individual animal possesses that are innate and typically fixed for that animal's life. Emotions vary continuously, sometimes on small time scales. In animals (including humans), temperament and emotion (and variations across groups) are as important to survival as cognition. They would make robots more effective also [1]. Emotions and temperament would also be useful in human-robot interactions.

There are five main types of temperament in humans and other animals, often called the Big Five [2]: Extrovert vs. Introvert, Neurotic vs. Rational, Conscientious vs. Careless, Agreeable vs. Disagreeable, and Open vs. Reticent. These are explained in more detail in Table 1, which shows definitions of the terms used to describe the Big Five. The first definition is from the Oxford dictionary [3] and the second is from the Merriam-Webster dictionary [4].

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_1

The literature on emotions is extensive and growing rapidly every year. A lot of these papers and books relate to humans, and we still do not understand the complexity of the human mind. In developing models of emotions and temperament, we should set our sights lower at first, and focus on robots or simpler animals (i.e. walk before we run). The model presented here has been implemented on cognitive mobile robots [5, 6], and the results were very encouraging. Instead of trying to use this model to predict human behavior, it should be viewed as a means of making a robot more effective and survivable by rapidly changing it's behavior in different situations. The particular emotions and temperaments discussed herein do relate to humans, but for robots we might use different emotions and temperaments, depending on the goals of the project. We are not trying to model human behavior or psychology here, although this model might eventually be useful for that.

When we design and build autonomous robots (for air, land, sea, or space) we do not generally think of behaviors varying across the group, but a heterogeneous mix of traits in a group will make the group more successful. In addition, unlike in biology

Table 1. Big five temperament definitions for humans from Oxford (1) and Merriam-Webster (2) dictionaries.

	Lower extreme	Upper extreme
1	Extrovert: (1) A person predominantly concerned with external things or objective considerations. (2) A friendly person who likes being with and talking to other people: an outgoing person	Introvert: (1) A person predominantly concerned with their own thoughts and feelings rather than with external things, (2) A shy person: a quiet person who does not find it easy to talk to other people
2	Neurotic: (1) Abnormally sensitive, obsessive, or tense and anxious. (2) Often or always fearful or worried about something: tending to worry in a way that is not healthy or reasonable	Rational: (1) Able to think clearly, sensibly, and logically, (2) having the ability to reason or think about things clearly
3	Conscientious: (1) Wishing to do what is right, especially to do one's work or duty well and thoroughly, (2) very careful about doing what you are supposed to do: concerned with doing something correctly	Careless: (1) Not giving sufficient attention or thought to avoiding harm or errors, (2) Not using care: not careful: done, made, or said without enough thought or attention
4	Agreeable: (1) Willing to agree to something, (2) Ready or willing to agree or consent	Disagreeable: (1) Unfriendly and bad-tempered, (2) Difficult to deal with: easily angered or annoyed
5	Open: (1) Frank and communicative; not given to deception or concealment, (2) Characterized by ready accessibility and usually generous attitude	Reticent: (1) Not revealing one's thoughts or feelings readily, (2) Inclined to be silent or uncommunicative in speech

where these traits are relatively fixed over the life of the organism, these could be varied in intelligent mobile robots. It might be extremely valuable if we could quickly change a group of robots from docile to aggressive, for example.

While people do not completely agree on a complete list of emotions, Damasio [7, 8] discusses six "universal" emotions: Fear, Anger, Sadness, Happiness, Disgust, and Surprise. Plutchik [9] discusses the same six emotions, but also includes "trust" and "anticipation." He also describes how there can be varying levels of each emotion in his emotion wheel. Ekman [10] describes 15 basic emotions. The six (or eight) emotions are common across animals [11] and cultures. It might not be essential to define the basic or primary emotions, since there is little agreement on this [12]. The method described herein, however, is not dependent on which emotions are used, they can be readily changed depending on what is of interest to the modeler or situation.

Damasio refers to emotions as automated programs for action that have been created through evolution. Emotions are related to reward, punishment, drives, perceptions, expectations, and motivations. There are typically negative and positive emotions, and they are tied to reinforcers (rewards, punishments, lack of reward, and lack of punishments), see [13–15]. Ortony et al. [12] discuss the activations and valences of emotions. They also discuss the difficulty in attaching words to emotions, and the futility of trying to list the basic emotions. Ortony et al. [12] also show a flowchart (their Figure 2.1) of how emotions might be structured. The figure shows three ways rewards and punishments can be triggered: events, actions of agents (self and other), and objects. For example, a tragic *event* might be a reinforcement that would tend to make you sad. A person (*agent*) could say something to you to make you sad, happy, etc. And thirdly, you might see an *object* (e.g. a tiger) that would cause your fear to increase.

While many investigators have studied affective computing in robots, there are very few studies which quantitatively define temperament and emotions or incorporate them into mobile robots. And the ones that do exist, do not properly distinguish temperament from emotions (e.g. [16, 17]). Gray [13] discusses the connection between emotions and cognition. The model used here has been incorporated into a cognitive robot [18]. Groups of mobile robots with a mix of personality and emotions will be more effective and have increased mission success.

An interesting anecdote relates to the well-known robot soccer competition. One of the researchers remarked that the robots play in the same manner at the start of the game as at the end, whereas a human would play very differently in the last few minutes of the game, especially if they were losing. The model presented here could allow the robots to change their behavior depending on the circumstances. Another example is group behavior. In nature there are many examples of groups (ants, fish, rats, humans, etc.) that are very effective, and the groups usually include a wide variety of personality types.

Emotions are basically state variables, and the robot could behave differently depending on which emotion it is experiencing. This is a very effective approach for rapidly changing a robots behavior, just as it is in animals. The temperaments, as described below, are fixed characteristics (although, unlike in animals, we could vary them in time). This model is also very well suited to coupling to cognitive architectures such as SOAR, ACT-R, or SS-RICS.

2 Emotion and Temperament Model

The survival of animals (including humans) depends on emotions and temperament. Robots with emotions and temperaments will also allow better interactions with humans. If the robot could understand the humans emotional state, and the robot behavior could change accordingly, that would be very interesting. Likewise, it would be useful if the human could sense the robot's emotional state. In addition, teams of animals (including humans) are more effective when the groups have a mix of temperaments. This has been shown true for robots [19], cockroaches [20], fish [21], ants [22], spiders [23], humans [24, 25], sheep [26], and other animals. Also, Eskridge and Schlupp [19] state:

> *The combination of different personalities within a group and the associated roles assumed by different members have been found to improve the overall success of the group (Couzin et al., 2005; Dyer et al., 2009; Modlmeier and Foitzik, 2011; Modlmeier et al., 2012). Studies have shown that these personality differences can be stable and maintained over time (Dal et al., 2004; Oosten et al., 2010).”* [27–32].

The model used herein builds upon the model for happiness by Rutledge et al. [33]:

$$Happiness(t) = w_o + \sum_{j=1}^{t} \gamma^{(i=j)} (w_1\, CR_j + w_2\, EV_j + w_3\, RPE) \tag{1}$$

where Happiness ranges from 0 to 100 and:

CR = Certain Reward (e.g. 10)
EV = Expected Value (e.g. 10)
RPE = Reward Prediction Error (e.g. 10)
w_o = Steady state value of happiness (e.g. 50)
w_1 = Magnitude of change (e.g. 0.52)
w_2 = Magnitude of change (e.g. 0.35)
w_3 = Magnitude of change (e.g. 0.80)
g = Rate of decay (e.g. 0.72)

The model has exponential decay given by γ. If the subject chose a certain reward (CR), then EV and RPE were zero. If the subject chose a gamble, then CR was zero. They also state:

> *“Conscious emotional feelings, such as momentary happiness, are core to the ebb and flow of human mental experience. Our computational model suggests momentary happiness is a state that reflects not how well things are going but instead whether things are going better than expected.”*

The model discussed here is guided by the above model, but modified to fit the task of modeling multiple emotions and temperament. The model is:

$$Emotion(t)_i = w_{o_i} + \sum_{j=1}^{t} \gamma_i^{(t-j)} \left(w_{1_i} R_{ij}^{+} - w_{2_i} R_{ij}^{-} \right) \tag{2}$$

where there are eight emotions (Fear, Anger, Sadness, Happiness, Disgust, Surprise, Anticipation, and Trust), each denoted by the subscript *i*, and there is exponential decay of rewards also. It would be quite easy to remove or add more emotions. The emotions change with time. The current time is given by *t*, and the subscript *j* denotes a previous time instance. At the present time, 45 previous steps are being used, but fewer would probably be fine too. R^{+} and R^{-} denote positive and negative reinforcers, respectively; and they have weights associated with them. These *R* terms are completely general, and could include reward prediction errors.

So the emotional state is a vector of length eight, but could be smaller or larger to accommodate different sets of emotions. Each emotion varies from 0 to 100, which is similar to the variations shown in Plutchik's [9] color wheel, which might not be the best analogy for emotions. One could use the maximum current emotional value, and conclude that that is *the* current emotion, but it might be better to consider all the emotions. The winner take all approach was used in Breazeal and Brooks [34].

Figure 1 shows the data flow of this approach. Events, agents, or objects in the environment generate reinforcers, which are used in the emotion model. The model outputs the value of each emotion, and it also computes which emotion has the highest value.

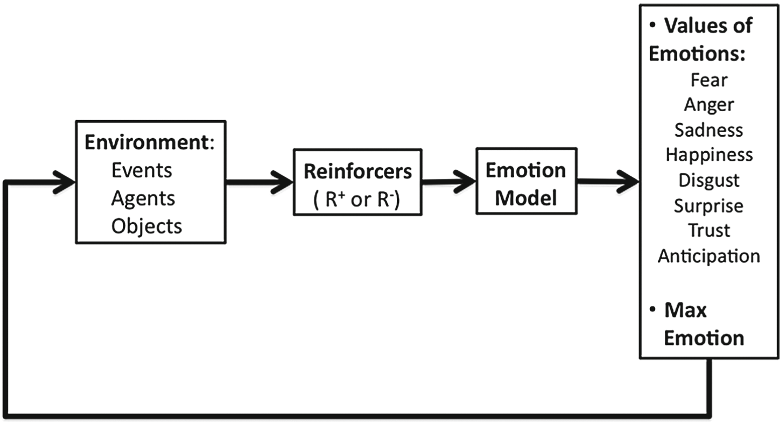


Fig. 1. Flowchart showing data flow.

In order to illustrate how the emotions vary with rewards (R^{+}) and punishments (R^{-}), a simulation was performed for just one emotion with a fairly complicated reward/punishment input. Figure 2 shows the input time history and the resulting emotion values. The left figure shows the time histories of the positive and negative reinforcers. The figure on the right shows the emotion value as a result of those. The model described here has been programmed in C++ (and the code is available from the author).

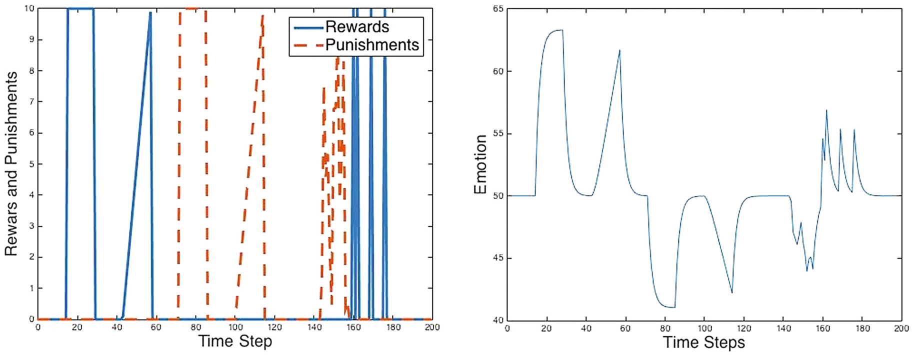


Fig. 2. Input time history to emotion model with resulting emotion levels

The positive and negative reinforcers can be due to numerous things in the environment. Ortony et al. [12] discuss emotions being tied to events, agents, and objects. All of these could be incorporated into the above model. For example, the robot might hear a loud explosion (an event), it might see an animal, human, or other robot (agents), or it might see an object that triggers an emotion. The implementer of the model needs to convert these into quantitative positive or negative reinforcers for one or more of the emotions. These reinforcers could be used to effect more than one emotion also.

Mood is also an important topic, and is related to emotions and temperament. One could model mood as a longer-term variation of the steady-state values (i.e. W_0). The model would produce this effect if there was a sustained low-level reinforcer for a longer period of time. This would cause the emotion to increase or decrease and remain at that level while the reinforcer was active, effectively varying the steady-state value for that period of time. An example of this is shown in Fig. 3, where a small (R^+ = 1) positive reinforcer was applied during time steps 10 to 170.

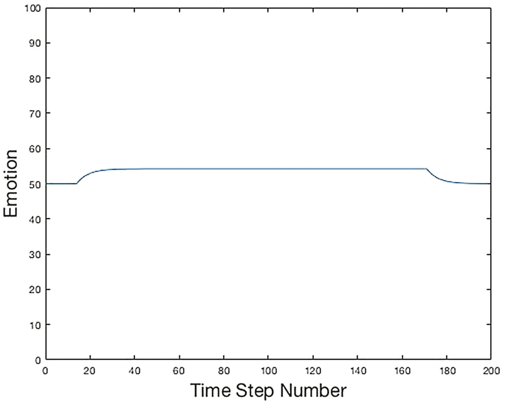


Fig. 3. Modeling mood using a small reinforcer over a long period of time.

This model is well suited to coupling to a cognitive architecture or other rule-based system. Using IF-THEN rules one can manage the values of the reinforcers. For example, if the robot sees something dangerous (e.g. a gun), the R^+ for fear at that time step could be set to a non-zero value. So there would be a set of rules needed to activate the positive and negate reinforcers. There would also need to be rules to handle the emotions. For example, one could determine which emotion has the highest value and then have some rules to handle it, e.g. if robot is afraid then run and hide.

One of the very interesting and crucial aspects of this model is that it also models temperament (i.e. personalities) through the coefficients. The temperament matrix T_{ij} can be defined as:

$$T_{ij} = \begin{bmatrix} w0_1 & w1_1 & w2_1 & \gamma_1 \\ w0_2 & w1_2 & w2_2 & \gamma_2 \\ w0_3 & w1_3 & w2_3 & \gamma_3 \\ w0_4 & w1_4 & w2_4 & \gamma_4 \\ w0_5 & w1_5 & w2_5 & \gamma_5 \\ w0_6 & w1_6 & w2_6 & \gamma_6 \\ w0_7 & w1_7 & w2_7 & \gamma_7 \\ w0_8 & w1_8 & w2_8 & \gamma_8 \end{bmatrix} \tag{3}$$

Each row of this matrix is associated with a particular emotion. The first column represents the steady state value of each emotion (0 to 100). The second and third columns are the weighting factors (0 to 1) on the rewards and punishments, respectively. The fourth column represents the decay rate (0 to 1), where the smaller the value the larger the decay rate. These 32 values can be used to define a person or robots personality. Few, if any, previous works have delimited the difference between modeling emotions and modeling temperament. It should be possible to model (using the above matrix) the big five temperaments presented earlier, but that is a very long-term goal.

The simplest form of the temperament matrix would have all the rows the same (all the emotions would then vary in the same way, but to different inputs), i.e.

$$T_{ij} = \begin{bmatrix} 50 & 0.52 & 0.35 & 0.61 \\ 50 & 0.52 & 0.35 & 0.61 \\ 50 & 0.52 & 0.35 & 0.61 \\ 50 & 0.52 & 0.35 & 0.61 \\ 50 & 0.52 & 0.35 & 0.61 \\ 50 & 0.52 & 0.35 & 0.61 \\ 50 & 0.52 & 0.35 & 0.61 \\ 50 & 0.52 & 0.35 & 0.61 \end{bmatrix} \tag{4}$$

where for simplicity representative values from Rutledge et al. [33] have been used. This matrix is probably not a good idea as the model might continually (and quickly) jump from one emotion to another (some people behave that way!), and the different emotions and reinforcers will require different coefficients. For example, “surprise” might decay very quickly (i.e. small γ), while “trust” might decay very slowly.

The rows of this matrix relate to: fear, anger, sadness, happiness, disgust, surprise, anticipation, and trust; respectively. All 32 values in the temperament matrix can be varied however.

In humans the big five temperaments [2], as mentioned earlier, are: Extrovert vs. Introvert, Neurotic vs. Rational, Conscientious vs. Careless, Agreeable vs. Disagreeable, and Open vs. Reticent. Characterizing these via the 32 adjustable constants corresponding to eight emotions is not trivial. The constants represent mean levels for each emotion, how fast perturbations decay, how strongly do they react to reinforcers, etc. Experiments would be required in order to compute the values of this temperament matrix for the various types of personalities.

The above model has been implemented on cognitive robots and with the ARL's SS-RICS software. It worked extremely well, as discussed in [18]. It was implemented on the robot shown in Fig. 4. As it roamed the building it could recognize objects such as guns, food, etc. These objects would act as reinforcers to the emotion engine. Its behavior would then change depending on its emotional state.

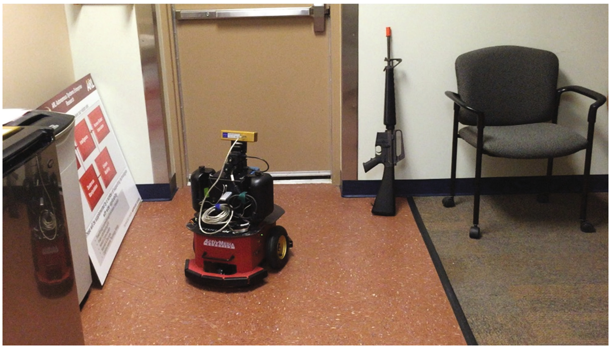


Fig. 4. Robot used for emotion and temperament model.

An example output from the robots emotion engine is shown in Fig. 5 from [5]. As the robot perceives different objects it's emotions vary significantly.

For the robot studies we used a few different temperament matrices, and showed how the robot behaved differently depending on its "personality." But we need a better approach to determining these coefficients, and we also need to show how the model varies with the input temperament.

We can rewrite the above equation as:

$$Emotion(t)_i = w_{o_i} + w_{1_i} \sum_{j=1}^{t} \gamma_i^{(t-j)} R_{ij}^{+} + w_{2_i} \sum_{j=1}^{t} \gamma_i^{(t-j)} R_{ij}^{-} \qquad (5)$$

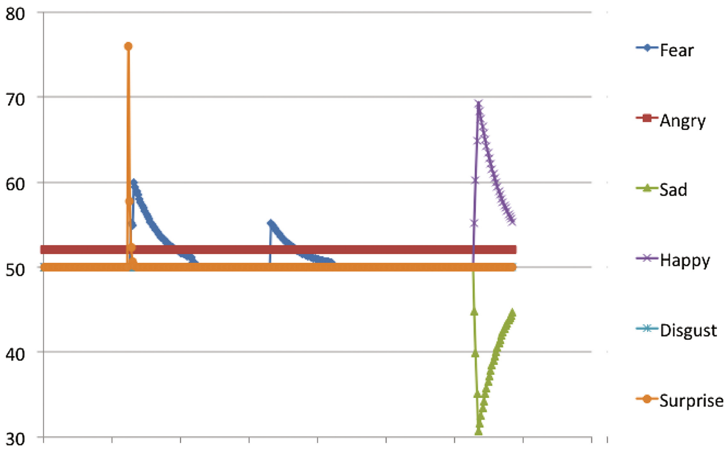


Fig. 5. Emotion (six) variations of cognitive robot [5].

So given some measurements (at several time steps) of the emotional state, some decay factors (γ) and the reinforcers (R+ and R^{-}), we could solve this to find the weights (w_o, w_1, and w_2) using a least squares approach. To make this clearer, lets rewrite this again as:

$$Emotion(t)_i = w_{o_i} + w_{1_i} A_i(t) + w_{2_i} B_i(t) \tag{6}$$

where

$$A(t)_i = \sum_{j=1}^{t} \gamma_i^{(t-j)} R_{ij}^{+} \quad B_i(t) = \sum_{j=1}^{t} \gamma_i^{(t-j)} R_{ij}^{-} \tag{7}$$

We could determine the coefficients for each emotion by performing some experiments on animals (or robots), where we measure the emotion and estimate the reinforcers. For example, for happiness, H(t), we'd have:

$$\begin{bmatrix} H(t_1) \\ H(t_2) \\ H(t_3) \\ H(t_4) \\ H(t_5) \\ H(t_6) \end{bmatrix} = \begin{bmatrix} 1 & A(t_1) & B(t_1) \\ 1 & A(t_2) & B(t_2) \\ 1 & A(t_3) & B(t_3) \\ 1 & A(t_4) & B(t_4) \\ 1 & A(t_5) & B(t_5) \\ 1 & A(t_6) & B(t_6) \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_3 \end{bmatrix} \tag{8}$$

The *H* vector contains the values of happiness at different times. If we rewrite this as:

$$\begin{gathered} \vec{H} = [M]\vec{W} \\ \text{or} \\ [M]\vec{W} = \vec{H} \end{gathered} \tag{9}$$

Using a least-squares approach this can be rewritten as

$$\begin{aligned}&[M]^T[M]\,\vec{W} = [M]^T\vec{H}\\&\text{or}\\&\vec{W} = ([M]^T[M])^{-1}[M]^T\vec{H}\end{aligned} \tag{10}$$

Thus, given experimental data for the reinforcers and the emotional responses, we could determine the appropriate values for *W*. That is we can quantify the personalities.

3 Conclusions

This paper has presented a mathematical and computational model for emotions and temperament. It is very flexible, and can be adapted to the needs of both biologists and roboticists. It is very well suited to coupling to a cognitive architecture (e.g. SOAR, ACT-R, or SS-RICS). Also presented is a least-squares approach for determining the proper coefficients for the model, given experimental data. The C++ computer code is available from the author.

References

1. LeDoux, J.E.: Emotion circuits in the brain. Ann. Rev. Neurosci. **23**, 155–184 (2000)
2. Digman, J.M.: Personality structure: emergence of the five-factor model. Ann. Rev. Psychol. **41**, 417–440 (1990)
3. Oxford Dictionary. www.oxforddictionaries.com
4. Merriam-Webster Dictionary. www.merriam-webster.com
5. Long, L.N., Kelley, T.D.: A review of consciousness and the possibility of conscious robots. J. Aerosp. Inf. Syst. **7**(2), 68–84 (2010)
6. Long, L.N.: Modeling emotion and temperament on cognitive mobile robots. In: 22nd Annual ACT-R Workshop, Carnegie Mellon University, Pittsburgh, PA, 17–19 July 2015
7. Damásio, A.R.: Descartes's Error: Emotion, Reason and the Human Brain. Putman, NY (1994)
8. Damasio, A.R.: Self Comes to Mind: Constructing the Conscious Brain. Pantheon Press, New York (2010)
9. Plutchik, R.: The nature of emotions. Am. Sci. **89**, 344–350 (2001)
10. Ekman, P.: Basic Emotions, Chapter 3, Handbook of Cognition and Cognition. Wiley, Chichester (1999). doi:10.1002/0470013494
11. Braithwaite, V.A., Huntingford, F.A., van den Bos, R.: Variation in emotion and cognition among fishes. J. Agric. Environ. Ethics **26**, 7–23 (2013)
12. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge Press, Cambridge (1988)
13. Gray, J.A.: Brain systems that mediate both emotion and cognition. Cogn. Emot. **4**(3), 269–288 (1990)
14. Rolls, E.T.: A theory of emotion, and its application to understanding the neural basis of emotion. Cogn. Emot. **4**, 161–190 (1990)

15. Rolls, E.T.: Emotion and Decision Making Explained. Oxford University Press, Oxford (2013)
16. Barteneva, D., Lau, N., Reis, L.P.: A computational study on emotions and temperament in multi-agent systems. In: Proceedings of the AISB 2007: Artificial and Ambient Intelligence, Newcastle, GB (2007)
17. Canamero, L.: Emotion understanding from the perspective of autonomous robots research. Neural Netw. **18**, 445–455 (2005)
18. Long, L.N., Kelley, T.D., Avery, E.S.: An emotion and temperament model for cognitive mobile robots. In: 24th Conference on Behavior Representation in Modeling and Simulation (BRIMS), March 31–April 3, 2015, Washington, DC
19. Eskridge, B.E., Schlupp, I.: Effects of personality distribution on collective behavior. In: ALIFE 14: 14th International Conference on the Synthesis and Simulation of Living Systems (2014)
20. Planas-Sitja, I., Deneubourg, J.-L., Gibon, C., Sempo, G.: Group personality during collective decision-making: a multi-level approach. Proc. R. Soc. B **282**, p. 20142515
21. Mittelbach, G.G., Ballew, N.G., Kjelvik, M.K.: Fish behavioral types and their ecological consequences. Can. J. Fish. Aquat. Sci. **71**, 927–944 (2014)
22. Pinter-Wollman, N.: Personality in social insects: How does worker personality determine colony personality? Curr. Zool. **58**(4), 579–587 (2012)
23. Pruitt, J.N., Keiser, C.N.: The personality types of key catalytic individuals shape colonies' collective behaviour and success. Anim. Behav. **93**, 87–95 (2014)
24. Pieterse, V., Kourie, D.G., Sonnekus, I.P.: Software engineering team diversity and performance. In: Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT 2006), pp. 180–186. Republic of South Africa (2006)
25. Moynihan, L.M., Peterson, R.S.: The Role of Personality in Group Processes, Chapter 12, Personality and Organizations, pp. 312–345. Psych. Press (2004)
26. Michelena, P., Sibbald, A.M., Erhard, H.W., McLeod, J.E.: Effects of group size and personality on social foraging: the distribution of sheep across patches. Behav. Ecol. **20**(1), 145–152 (2009)
27. Couzin, I.D., Krause, J., Franks, N.R., Levin, S.A.: Effective leadership and decision-making in animal groups on the move. Nature **433**(7025), 513–516 (2005)
28. Dyer, J.R., Croft, D.P., Morrell, L.J., Krause, J.: Shoal composition determines foraging success in the guppy. Behav. Ecol. **20**(1), 165–171 (2009)
29. Modlmeier, A.P., Foitzik, S.: Productivity increases with variation in aggression among group members in Temnothorax ants. Behav. Ecol. **22**(5), 1026–1032 (2011)
30. Modlmeier, A.P., Liebmann, J.E., Foitzik, S.: Diverse societies are more productive: a lesson from ants. Proc. R. Soc. B: Biol. Sci. **279**(1736), 2142–2150 (2012)
31. Dall, S.R., Houston, A.I., McNamara, J.M.: The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecol. Lett. **7**(8), 734–739 (2004)
32. Oosten, J.E., Magnhagen, C., Hemelrijk, C.K.: Boldness by habituation and social interactions: a model. Behav. Ecol. Sociobiol. **64**(5), 793–802 (2010)
33. Rutledge, R.B., Skandalia, N., Dayan, P., Dolana, R.J.: A computational and neural model of momentary subjective well-being. Proc. Natl. Acad. Sci. **111**(33), 12252–12257 (2014)
34. Breazeal, C., Brooks, R.: Robot emotion: a functional perspective. In: Fellous, J.-M., Arbib, M.A. (eds.) Who Needs Emotions, pp. 271–310. Oxford University Press, Oxford (2003)

A Cybernetic Approach to Characterization of Complex Sensory Environments: Implications for Human Robot Interaction

Kelly Dickerson(✉), Jeremy Gaston, and Kelvin S. Oie

US Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA
kelly.dickerson5.civ@mail.mil

Abstract. Humans are increasingly interacting and collaborating with robotic and intelligent agents. How to make these interactions as effective as possible remains, however, an open question. Here, we argue that consistent understandings of the environment on the part of the human and agent are critical for their interaction and basing these understandings on only the objective features of sensory inputs may be inadequate. To that end, the current paper presents a novel approach to more integrated characterizations of the sensory environment that encompass objective and subjective features of sensory inputs. We propose that an approach to signal and behavioral estimation consistent with the control and communication theoretic perspective of Cybernetics could inform human robot interaction (HRI) applications. Specifically, we offer a potential path forward for quantifying similarity in stimulus events that can lead to consistent understandings of the environment, which when applied to HRI can enhance human-agent communication in HRI applications.

Keywords: Cybernetics · Human robot interaction · Stimulus classification · Information theory

1 Introduction

In recent years, the US Army has seen an increased integration of highly skilled Soldiers with advanced technologies, with a significant emphasis in future interactions with robotic and other intelligent agents. To support these efforts, we have adopted the conceptual framework of Cybernetics. This transdisciplinary approach, popularized most notably by [1], is the "scientific study of control and communication in the animal and machine." Cybernetics, therefore, encompasses the disciplines of control theory and communication (a.k.a., information) theory, as a general approach to understanding closed-loop, feedback systems [2, 3].

Feedback can be defined generally as the information that is generated by a system's control actions and the resulting interactions with the external environment, which when sensed by the system can then be utilized in planning and executing future actions. In a complex system, such as the human brain, feedback operates at multiple levels; for example, feedback influences internal affective or physiological states, or the

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_2

interactions among single or small clusters of neurons. At the cognitive level, feedback influences intention and shapes action “closing” the loop that moves from intention to action, to sensing the outcome of action, to the comparison of an outcome to the original intention, and provides information to make appropriate adjustments in further action. This closed loop circularity is essential to “cybernetic” systems as it enables adaptation under complex and dynamic conditions.

However, understanding the mechanisms underlying the incorporation of feedback into higher-level human behaviors, such as decision-making, is more conceptually challenging than such an idealized loop might suggest. Human sensation and perception is often incomplete, inaccurate, and ambiguous, and it is not always possible to extract all of the information needed to support behavior directly from immediate sensing of the environment. This can create conditions of undesirable uncertainty about one’s relationship with the environment. In the face of this uncertainty, humans likely weigh the input from multiple modalities, and can use the combined or integrated perceptual result to guide action. The information that underlies this weighting in a highly complex real world interaction remains an open question. The answer to this question, of course, can depend strongly on the situational context, and humans are generally quite flexible in their putative information processing strategies. When a human must communicate their intentions to an agent, the information encoded in flexible combinations of modalities, stimulus features, and assumptions or expectations that typically serve human-to-human communications well are not available for decoding by the agent. This lack of effective communication can lead to significant misunderstandings between human users and the systems they rely on for successful task performance.

Here, we suggest, that in human-agent teaming situations, for effective communication between an agent and their human counterpart it is likely critical that their understandings of the environment be consistent with each other. For example, humans and robots, depending on their relative size, may understand the same physical objective differently: For a micro-autonomous system, a shoebox would be a significant obstacle, whereas for their human counterpart, it would not. However, for both the human and robot, the shoebox may pose a threat, because something could be contained within it that neither of their visual sensors would be able to detect. This is just a single example of the difficulties human-agent teams could face while navigating a complex and dynamic environment. However, creating consistent representations of the everyday environment is no small feat given the differences in sensing and perceptual architectures between humans and the myriad potential artificial agents with which they might team. The representations that typically underlie such understandings in robotic and other intelligent agent applications still mainly reflect low-level, quantitative aspects of their physical sensory inputs. By contrast, as argued above, understanding the environment for humans does not just comprise mappings of the immediate physical domain. Human representations, instead, typically reflect the integration of the more objective information based on current sensory inputs with more subjective information that strongly depends on assumptions or expectations derived from previous experiences in other contexts, which is difficult, if not impossible, to directly measure.

The subjective attributes of human perceptual experiences conceived of here as cognitive features that vary across individuals, but with values bounded by prior information from experience within relevant behavioral contexts. For example, the subjective experience of a cognitive feature, such as the perceived pleasantness of the sound of a car's engine, may be low for someone unfamiliar with the loud and intermittently impulsive mechanical noise. However, when that engine sound has become associated with the returning home of a spouse or parent at the end of the day, those previous experiences might positively influence the perceived pleasantness of the input. In turn, this influence may also change subjective experiences across a variety of contexts. The example provided here is meant to illustrate that the influences of subjective experience on the representations that are fundamental to understanding in complex environments are likely pervasive. Still, the challenging task of quantifying subjective stimulus attributes is an area of research that has been somewhat lacking, with even less information available on how one would apply or translate this research in the context of enhancing human-robot interactions (HRI).

1.1 Objectives of This Paper

This paper will briefly discuss the relative strengths and weaknesses of human and systems approaches to stimulus classification from visual, auditory, and multimodal inputs. We will highlight the relatively limited research that compares human and agent performance on common tasks and discuss how these comparisons can support the development of better models of interaction between humans and agents. We then describe our work on characterizing the sensory environment, which has thus far focused largely on auditory environment quantification. This research on human performance in the context of subjective stimulus attributes is discussed with an eye towards using such an approach to improve models of multisensory integration for HRI applications.

2 Comparing Human and Machine Classification Approaches

One critical skill required for successful navigation of the environment is the ability to detect, localize, and recognize objects and boundaries. Using vision, humans perform this task seemingly effortlessly, while object localization and recognition in machine vision is resource-intensive and cannot yet match human performance in all conditions. For example, in a direct comparison of a robotic machine vision algorithm and human classification, [4] found that humans were 1.7% more accurate. This difference in accuracy may seem small, however, the source of accuracy differences was revealing. Specifically, when the algorithm made classification errors, it was found to be due to image features that humans typically have no difficulty processing, such as view point and color invariance [4, 5]. The algorithm also had difficulty classifying images that were graphic or symbolic representations of real objects (i.e. drawing of a coffee cup, or an image of a stuffed bear). Despite these limitations, machine vision is improving rapidly and there are emerging examples where Deep Learning approaches have exceeded the best human performance for specific image data sets. For example, [6] showed machine image classification for the ImageNet 2012 classification dataset that exceeded the best-reported human performance by more than 5%.

In audition, there are also examples of machine algorithms classifying environmental sound events; however, in almost every case the stimulus set is restricted to a homogeneous class of events, such as the detection and localization of gunfire events, [7, 8]. Under these conditions human and algorithm classification performance was comparable. However, in the majority of real world environments, the input event distribution is much more varied than that used in these studies. This is a pervasive issue for current machine language solutions: in application domains where the behavioral space intrinsically involves greater variety, machine language approaches cannot yet match human performance. For example, in speech recognition, the domain where the majority of auditory machine learning applications have been developed [9], humans typically do better than implemented machine approaches in terms of accuracy of recognized speech [10, 11]. However, as with machine vision, machine speech recognition systems are rapidly improving; a very recent application of machine learning approaches by Microsoft has yielded parity with humans in transcribing speech from the NIST 2000 speech test set [12].

Multimodal machine learning classifiers have also had demonstrated success, but with limited comparison to human performance. Importantly, in multimodal classification, the addition of redundant, as well as potentially unique, information from another modality is one obvious way to improve machine classification. For example, adding audio to visual information can improve classification by increasing bandwidth, accuracy, and decreasing processing time by using converging evidence to support classification of environmental objects under difficult or ambiguous conditions. Examples include the audio-visual classification of speech [13] and audio-visual and textual sentiment analysis [14], where in both cases the additional sensory cues led to better classification performance.

These direct comparisons between human and machine performance are useful in assessing advances in machine performance and potentially identifying where further advancements may occur. However, another approach to understanding the differences between human and machine performance is to examine their capabilities in the context of collaboration. One robust example of human-agent collaboration is the Human-Autonomous Image Labeler (HAIL) developed by the US Army Research Laboratory. The performance of the HAIL system depends critically on both human and computer vision systems [15]. Specifically, it takes advantage of the capability for rapid, but sometimes inaccurate classification of tens of thousands of images by computer vision agents, and couples that with the capability for very accurate, but much slower classification by human agents. The outcome is a very accurate and fast classification of a large set of images [15, 32] that, instead of highlighting the limitations of human and machine performance, takes advantage of the respective strengths of human and autonomous agents to increase the performance of the "system" as a whole.

2.1 The Problem of Similarity

Classification performance by intelligent agents, humans, or human agent teams is, in many cases, negatively impacted when the to-be-classified content is highly similar to background or distractor information [16]. Although some image algorithms excel at classifying highly similar images, such as those that could be part of the same fine-grained (local level) category, many developed algorithms tend not to be robust to suboptimal viewing conditions and could still produce significant classification errors. Multimodal cuing could aid classification by using auditory information to disambiguate highly similar visual inputs and support discrimination; for example, two dogs of different breeds likely have distinctive barks.

Indeed, for humans, it is well-known that visual task performance is often augmented by the presence of auditory information. This multisensory enhancement effect [see 19] is possible due, to the fact that many objects in the environment are only fully described by the combination of distinct auditory and visual features. This suggests that processes for audiovisual integration can capitalize on informational redundancy to reduce uncertainty in perceptual estimates, enhancing the resultant representation of the world and making it both more coherent and more robust [17, 18]. More generally, human perceptual systems combine and integrate information from their multiple different sensory modalities, which reduces the variance and, generally, increases the reliability of perceptual estimates that support the higher cognitive functions, including decision making.

There is clearly value in adapting these strategies for HRI applications, yet it remains unclear whether multisensory perception based solely on current sensory inputs provide adequate information for complex decision making. Indeed, [4, 33] (see also [18] for review) have shown that some of the efficient and robust sensory combinations and integrations underlying humans' higher-level perceptual capabilities rely on representations of prior experiences, and that these subjective attributes may not be easily translated from human to non-human systems. For example, in social interactions using text-based communication (i.e., IM), the presence of punctuation can influence perceived sincerity of a comment in younger users who are used to crafting text in the absence of traditional punctuation [20]. Similarly, in reading, the same words can convey different senses of urgency depending on the contextual framing provided by story narration. A reader can perceive the activity of a character as urgent if the narrator uses language that suggests fast movements, but perceived urgency is limited when the narrator uses language to suggest slower movements or does not describe the rate of activity [21].

3 Characterizing the Sensory Environment

As discussed above, humans use information from multiple modalities to reduce uncertainty in perceptual estimates, which supports efficient decision-making. Multisensory integration is often biased based on previous experiences with a given object in a particular context [19]. This bias can manifest in two opposing ways, as a performance decrement or as a performance enhancement, depending on how the information available is combined or integrated with prior information from previous experience in the representation of the current situation. However, as alluded to above, sometimes complex and high dimensional, experience-based factors can be difficult to define and are resistant to direct measurement.

In the human multisensory perception literature, Bayesian approaches have emerged as an important tool in understanding how multimodal sensory cues can be integrated.

There are a number of examples where Bayesian maximum likelihood estimation (MLE) models predict multisensory integration [9, 20, 22, 23]. Maximum likelihood estimation models of multisensory integration [22] (also known as Bayesian inference models) maximize the *maximum a posteriori* (MAP) estimate associated with a particular response by dynamically updating the maximum likelihood functions associated with the sensory cues. Across trials, the resultant MAP predictions are weighted sums of the unimodal sensory inputs, where the weight reflects the relative cue uncertainty gathered from previous trials. However, the Bayesian priors are theoretical values that estimate the magnitude of the impact of previous experiences with a given set of multimodal cues and these estimates are not necessarily based on the real distribution of experiences. It is not clear how well this approach would extend to dynamically changing multimodal cues, or even if this approach could scale up to real world audio visual events. A prerequisite to evaluate the possibility of quantifying experience-based factors using a Bayesian approach would be a better understanding of the physical and perceived signal qualities of stimuli in the environment. [24] found that informational and contextual factors affect listeners' ability to identify environmental sounds. For example, they found that the presence of a competing background, particularly a background with overlapping information reduced sound identification accuracy. Similar effects have emerged in our own work, [25] found that identification accuracy was better when sounds had a clear originating event ("concrete") than when the link between the sound generating object and the sound produced was less obvious ("abstract").

The results of [24, 25] suggest that subjective and contextual factors convey important task relevant information. To better understand the content of environmental sounds, and the interaction between these subjective factors and object stimulus parameters researchers have applied stimulus classification techniques to environmental sound perception. These techniques offer a method for addressing possible limitations in the way Bayesian priors are estimated for real stimulus events. [26] used listener defined similarity scores, as well as objective measures of spectral and temporal features of sounds, to create a classification space for a large set of common environmental sounds. Dickerson et al. [27] extended this approach by using prior subjective ratings of stimulus similarity to characterize human listener performance on a variety of different behavioral tasks. It is possible that data of this nature could be used to improve Bayesian estimation techniques and could provide consistent and meaningful feedback about the environment to both human and non-human intelligent agents. In several related experiments, we have further extended our understandings of and approaches to quantifying the relationship between informational and contextual effects using the construct of similarity. [28] found that, for change discrimination, similarity in perceived loudness influenced the likelihood of noticing that an element within a scene had changed. This effect of similarity also manifested in the identity relationships among the 25 signals in their stimulus set, with a linear relationship found between the likelihood of change discrimination and the overall similarity (defined via user rating and a multidimensional scaling (MDS) analysis) of the sounds in the stimulus set. This relationship between identifiability, similarity, and perceptual performance is not particularly compelling on its own; the well-established informational masking literature would likely predict some of these effects [29]. However, this becomes more compelling when we examine the robustness of these trends across changes in methods and paradigms. [27] found that similarity among sounds in a scene affected both change discrimination and change localization performance, where increasing similarity decreased accuracy. [30] found that this effect extended to performance on cued-recall tasks, as well. It was further revealed that a complex interaction exists between user ratings of identifiability and similarity and later memory performance: Sound sources group together based on identifiability and category membership, but sounds within a tightly clustered group were more poorly recalled.

The work from our group, along with Bayesian approaches to multisensory integration suggests that subjective information quantified in the manner discussed in this paper provides a tractable method for including subjective information in Bayesian prior estimation. By using this type of information in the development of machine sensing approaches, it becomes possible for man and machine to have a deeper and more consistent understanding of their operational environment, potentially reducing the workload associated with communicating information between agent and human teammate.

4 Conclusions and Future Directions

The research highlighted here suggests that in order to accurately and meaningfully represent the environment to both man and machine, more information than the direct sensory stream may be required. By quantifying subjective attributes, such as similarity, that relate complex features across objective and subjective perceptual estimates, researchers can develop a better understanding of the feedback that guides behavior under the complex and dynamic conditions of the real world. Additionally, the research summarized here converges on emerging perspectives in multisensory integration, that, rather than separating out each sensory stream for modular and potentially parallel processing, the auditory and visual information are processed together as a holistic object [31]. This perspective suggests that there may be value in the further uncertainty reductions and saliency gains in including subjective factors in the characterization of stimulus events. Future research will focus on continuing to evaluate how humans and intelligent agents complete tasks in isolation and in cooperation in order to uncover the stimulus-related objective and subjective factors producing efficient and accurate behavior.

References

1. Wiener, N.: Cybernetics: Control and Communication in the Animal and the Machine. Wiley, New York (1948)
2. Seising, R.: Cybernetics, system (s) theory, information theory and fuzzy sets and systems in the 1950s and 1960s. Inf. Sci. **180**, 4459–4476 (2010)
3. Dubberly, H., Pangaro, P.: Cybernetics and service-craft: language for behavior-focused design. Kybernetes **36**, 1301–1317 (2007)
4. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)
5. Biederman, I., Bar, M.: One-shot viewpoint invariance in matching novel objects. Vis. Res. **39**, 2885–2899 (1999)

6. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision (2015)
7. Lopez-Morillas, J., Canadas-Quesada, F.J., Vera-Candeas, P., Ruiz-Reyes, N., Mata-Campos, R., Montiel-Zafra, V.: Gunshot detection and localization based on non-negative matrix factorization and SRP-hat. In: Sensor Array and Multichannel Signal Processing Workshop (SAM), pp. 1–5. IEEE (2016)
8. Khalid, M.A., Babar, M.I.K., Zafar, M.H., Zuhairi, M.F.: Gunshot detection and localization using sensor networks. In: Smart Instrumentation, Measurement and Applications (ICSIMA), pp. 1–6. IEEE (2013)
9. Deng, L., Li, X.: Machine learning paradigms for speech recognition: an overview. IEEE Trans. Audio Speech Lang. Process. **21**, 1060–1089 (2013)
10. Lippmann, R.P.: Speech recognition by machines and humans. Speech Commun. **22**, 1–15 (1997)
11. Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Rose, R.: Automatic speech recognition and speech variability: a review. Speech Commun. **49**, 763–786 (2007)
12. Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., Yu, D., Zweig, G.: Achieving Human Parity in Conversational Speech Recognition. Microsoft Research Technical Report MSR-TR-2016-71, February 2017
13. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y. Multimodal deep learning. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11) (2011)
14. Poria, S., Cambria, E., Howard, N., Huang, G.B., Hussain, A.: Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing **174**, 50–59 (2016)
15. Saproo, S., Faller, J., Shih, V., Sajda, P., Waytowich, N.R., Bohannon, A., Jangraw, D.: Cortically coupled computing: a new paradigm for synergistic human-machine interaction. Computer **49**, 60–68 (2016)
16. Brooks, J., Slayback, D., Shih, B., Marathe, A., Lawhern, V., Lance, B.J.: Target class induction through image feedback manipulation in rapid serial visual presentation experiments. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1047–1052, October 2015
17. Burr, D., Alais, D.: Combining visual and auditory information. Progr. Rain Res. **155**, 243–258 (2006)
18. Ernst, M.O., Bülthoff, H.H.: Merging the senses into a robust percept. Trends Cogn. Sci. **8**, 162–169 (2004)
19. Shams, L., Seitz, A.R.: Benefits of multisensory learning. Trends Cogn. Sci. **12**, 411–417 (2008)
20. Gunraj, D.N., Drumm-Hewitt, A.M., Dashow, E.M., Upadhyay, S.S.N., Klin, C.M.: Texting insincerely: the role of the period in text messaging. Comput. Hum. Behav. **55**, 1067–1075 (2016)
21. Gunraj, D.N., Drumm-Hewitt, A.M., Klin, C.M.: Embodiment during reading: Simulating a story character's linguistic actions. J. Exp. Psychol.: Learn. Mem. Cogn. **40**, 364–375 (2014)
22. Angelaki, D.E., Gu, Y., DeAngelis, G.C.: Multisensory integration: psychophysics, neurophysiology, and computation. Curr. Opin. Neurobiol. **19**, 452–458 (2009)
23. Roach, N.W., Heron, J., McGraw, P.V.: Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. Proc. R. Soc. Lond. B: Biol. Sci. **273**(1598), 2159–2168 (2006)
24. Leech, R., Gygi, B., Aydelott, J., Dick, F.: Informational factors in identifying environmental sounds in natural auditory scenes. J. Acoust. Soc. Am. **126**, 3147–3155 (2009)

25. Dickerson, K., Foots, A., Gaston, J.: The influence of concreteness on identification and response confidence for common environmental sounds. PLoS ONE (under review)
26. Gygi, B., Kidd, G.R., Watson, C.S.: Similarity and categorization of environmental sounds. Atten. Percept. Psychophys. **69**, 839–855 (2007)
27. Dickerson, K., Gaston, J., Foots, A., Mermagen, T.: Sound source similarity influences change perception during complex scene perception. J. Acoust. Soc. Am. **137**, 2226 (2015)
28. Gaston, J., Dickerson, K., Hipp D., Gerhardstein, P.: Change deafness for real spatialized environmental scenes. Cogn. Res.: Princ. Implic. (in press)
29. Dickerson, K., Gaston, J.R.: Did you hear that? The role of stimulus similarity and uncertainty in auditory change deafness. Front. Psychol. **5**, 1–5 (2014)
30. Dickerson, K., Sherry, L., Gaston, J.: The relationship between perceived pleasantness and memory for environmental sounds. J. Acoust. Soc. Am. **140**(4), 3390 (2016)
31. Ramenahalli, S., Mendat, D.R., Dura-Bernal, S., Culurciello, E., Niebur, E., Anderou, A.: Audio-visual saliency map: overview, basic models and hardware implementation. In: 2013 47th Annual Conference on Information Sciences and Systems (CISS), pp. 1–6. IEEE (2013)
32. Slayback, D., Files, B., Lance, B., Brooks, J.: Effects of image presentation highlighting and accuracy on target class induction (in preparation)
33. Ernst, M.O., Banks, M.S.: What determines dominance of vision over haptics? In: Proceedings of the Annual Psychonomics Meeting (2000)

Spatial Understanding as a Common Basis for Human-Robot Collaboration

D. Paul Benjamin[1(✉)], Tianyu Li[1], Peiyi Shen[1], Hong Yue[1],
Zhenkang Zhao[1], and Damian Lyons[2]

[1] Pace University, 1 Pace Plaza, New York, NY 10038, USA
{dbenjamin,yh19243n}@pace.edu
[2] Fordham University, 340 JMH, 441 E. Fordham Road, Bronx, NY 10458, USA
dlyons@fordham.edu

Abstract. We are developing a robotic cognitive architecture to be embedded in autonomous robots that can safely interact and collaborate with people on a wide range of physical tasks. Achieving true autonomy requires increasing the robot's understanding of the dynamics of its world (physical understanding), and particularly the actions of people (cognitive understanding). Our system's cognitive understanding arises from the Soar cognitive architecture, which constitutes the reasoning and planning component. The system's physical understanding stems from its central representation, which is a 3D virtual world that the architecture synchronizes with the environment in real time. The virtual world provides a common representation between the robot and humans, thus improving trust between them and promoting effective collaboration.

Keywords: Virtual world · Soar cognitive architecture

1 Introduction: Mental Models

Computer vision has had a difficult time reproducing the human ability to understand visual scene information across a wide range of applications domains and environmental conditions. There is evidence from cognitive psychology [1] that effectively leveraging context is a key aspect of this human facility. However, while there has a strong bottom-up Marr-based stream of vision research [2], the use of context has also been recognized in computer vision for a long time: at least from the Univ. of Mass. VISIONS project [3] and more recently to the linguistic-inspired Bag of Words approaches (e.g., [4]) global extensions of scale-invariant features (e.g., [5]) and others [6]. But in general these approaches still view scene recognition as a 'recognize the snapshot' problem, with little input from ongoing, long term objectives and tasks of the system. The scene understanding problem for a human is one of an embedded system leveraging sensing to fulfill its goals: sensing is strongly biased in the service of task and how the agent's and other agents' actions are expected to play out in the physical world.

Research in cognitive psychology indicates that the use of 3D models in spatial comprehension is fundamental, even in people who have been blind since birth [7]. Recent evidence in cognitive psychology [8] and neuroscience [9] supports the

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_3

proposition that simulation, the "re-enactment of perceptual, motor and introspective states" is a central cognitive mechanism that helps to provide context for planning. Shanahan [8] proposes a large-scale neurologically plausible architecture that allows for direct action (similar to a behavior-based approach) and also "higher-order" or "internally looped" actions that correspond to the rehearsal or simulation of action without overt motion.

People have evolved a set of sophisticated strategies for using limited short term memory and limited processing speed to solve extremely difficult problems, including visual comprehension. These strategies reflect an engineered structure that avoids the computationally expensive algorithms of modern computer vision. Our vision system architecture is directly inspired by this cognitive and neurobiological structure, and our goal is to try to replicate it at a functional level. One of our unfunded collaborators is a professor of neurobiology at Fordham University, who will advise us on the cognitive and neurobiological plausibility of aspects of our system and also compare our system's performance with that of the human visual system.

The human vision system does not apply equal computational resources everywhere in its visual field, but instead focuses on and analyzes just a small portion of the visual field at each moment; this is called a *fixation* [10]. After extracting the needed information from that region of the visual field, the vision system rapidly moves the eyes to a new region of the visual field for the next fixation. These rapid movements are *saccades*, which are quick movements across the visual field, and *vergences*, which change the depth of focus [10]. The effect of this organizational structure is to permit efficient use of limited computational resources. Instead of fully processing all of the sensory input and then discarding everything that is not relevant to the goals, this organization applies computational resources only to the parts of the sensory input that are likely to be relevant to the agent's goals. The key is to organize the search of the visual field in a manner that effectively gathers useful information.

Much work has been done on measuring the functioning of the human vision system and of its system of saccades and vergences [10], but there has not been a computational implementation that connects the actions of the vision system to the goals of the agent.

The research hypothesis of our work is that the movements of the vision system are those that are necessary to build a sufficiently accurate 3D world model for the robot's current goals. For example, if the goal is to navigate through a room, the model needs to contain any obstacles that would be encountered, giving their approximate positions and sizes. The vision system needs to obtain this information; other information does not need to be rendered into the virtual world.

In this way, our system prunes the information at the perception stage, using its knowledge about the agent's goals and about objects in the world and their dynamics to decide where to look and what type of information to obtain. This is in contrast to the usual approach of gathering lots of sensory information, processing it all and rendering it into a world model in a goal-independent manner, then deciding which information is necessary for decision making. This latter approach wastes a great deal of processing time processing information that is discarded in the decision-making process. Our goal is to design a fast, inexpensive vision system by emulating the functional organization of the human vision system.

This approach to spatial comprehension and modeling is based on the functional structure of the human visual system. The usual approach to the use of vision in robotics is to attempt to solve two problems [11]:

(a) Process visual data to extract all the objects and motions in the environment,
(b) Identify the results from (a) that are important and relevant to the current task.

Unfortunately, both of these steps are very expensive computationally. The first step requires processing an enormous amount of visual data, especially when the environment is very dynamic. The second step is a difficult data mining problem.

Our approach to this complexity issue is to leverage goal-directed rendering: the robot first decides which aspects of its environment are relevant, based on its task goals. This information is used to focus the cameras on specific regions of the environment and extract only the information needed for the goals. This approach is important because it has the potential to be faster and less expensive than current approaches. Our current system runs on a laptop in real time.

2 A Basic Example: Tracking a Rolling Ball

A robot wishes to predict the path of a bouncing ball so that it can intercept the ball as quickly as possible. The ball will bounce off walls that will alter its path. The robot needs to perceive the objects that the ball will hit and also perceive the ball's motion, then combine this information to produce a predicted path.

In Fig. 1 below, we see two boards placed on the floor. Our vision system detects keypoints and lines in the images, then selects a region for initial focus. The density of keypoints at the bottom of the images causes this to be selected as the region of focus. This region is denoted by the dark boxes in Fig. 2.

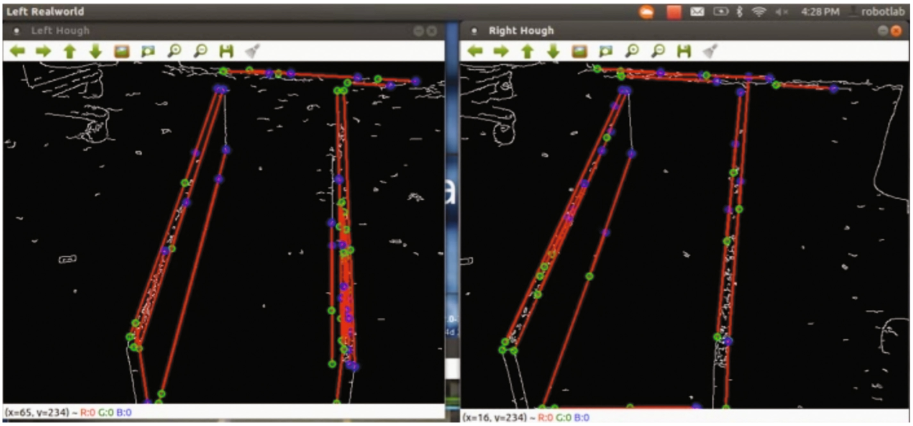


Fig. 1. Initial processing of two boards on the floor, showing keypoints and lines.

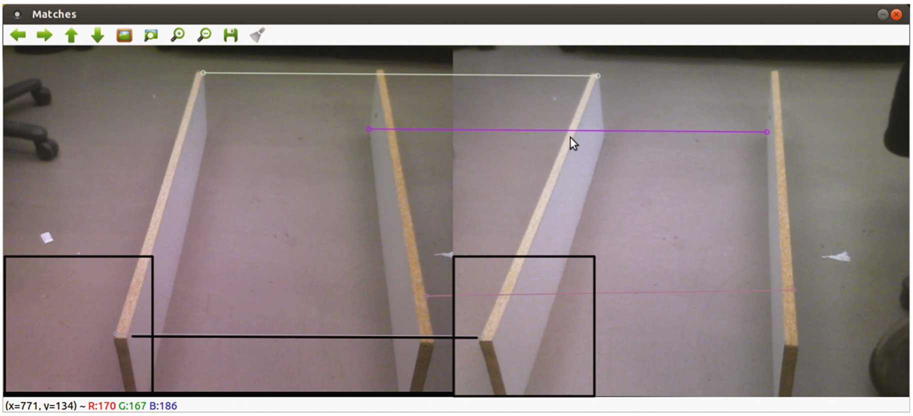


Fig. 2. The region of focus is the boxes at the lower left, yielding the correspondence between the corners.

The upper corners of the boards are detected, and registration of the left and right images produces an initial correspondence, together with position and orientation data.

A saccade is performed to the next region of focus. In this case, the closest region is chosen, which is a correspondence between the right boards. This process is repeated four times until the top corners of the boards are reached.

Segmentation information is added, producing additional correspondences. This is combined with the correspondences from the keypoints, yielding a small set of best correspondences. In Fig. 3, we see the tops of the boards correspond.

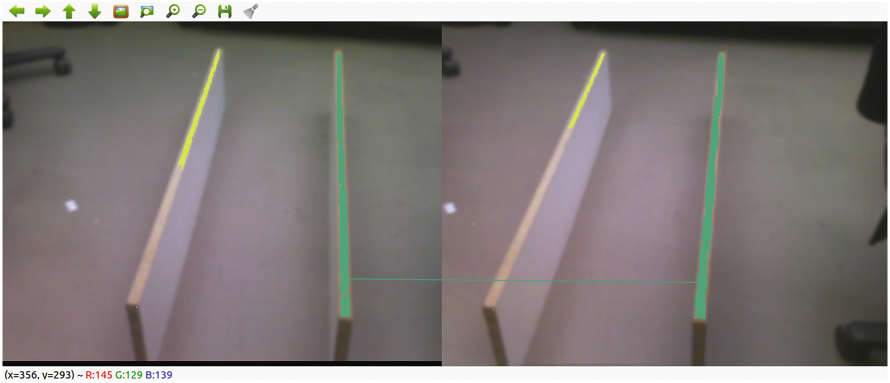


Fig. 3. Segmentation correspondence between the boards.

Finally, the boards are rendered in PhysX, and a ball is added. The ball is rolled from right to left (Fig. 4).

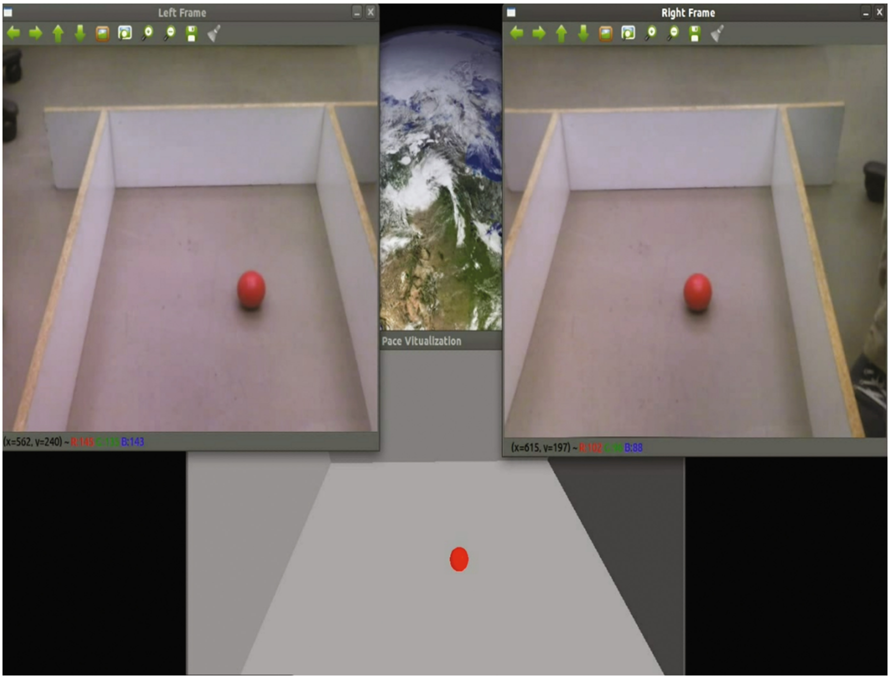

Left Frame
Right Frame
Pace Vitualization
(x=562, y=240) ~ R:145 G:135 B:143
(x=615, y=197) ~ R:102 G:96 B:88

Fig. 4. A ball is rolled between two boards. Left and right images are at top. The virtual world is at bottom.

The direction and velocity of the ball are computed over a small interval then duplicated in the virtual world. The physics engine is then run much faster than real-time, producing a predicted path for the ball. A mobile robot can use this prediction to intercept the ball efficiently.

A number of videos showing this process in various scenarios are available at http://csis.pace.edu/robotlab/videos.html.

3 Example: Tracking a Moving Car

Our current work is to monitor and predict the motion of a car so that we can drive another car beside it without any accidents. We are patterning our work on the KITTI project because of its excellent data, and our goal is to render actual traffic into the 3D model in real time. Our initial step is to use the PhysX vehicle demo for both worlds.

Figure 5 shows the use of the PhysX demo. The left column is the real world and the right column is the rendered world. In the first pair of images, the cars start synchronized. In the second pair of images, the real car begins to move, and differences are created between the images. In this demo, the virtual camera follows the car, so the differences include differently shaped regions in the background of the images and along the borders of the road.

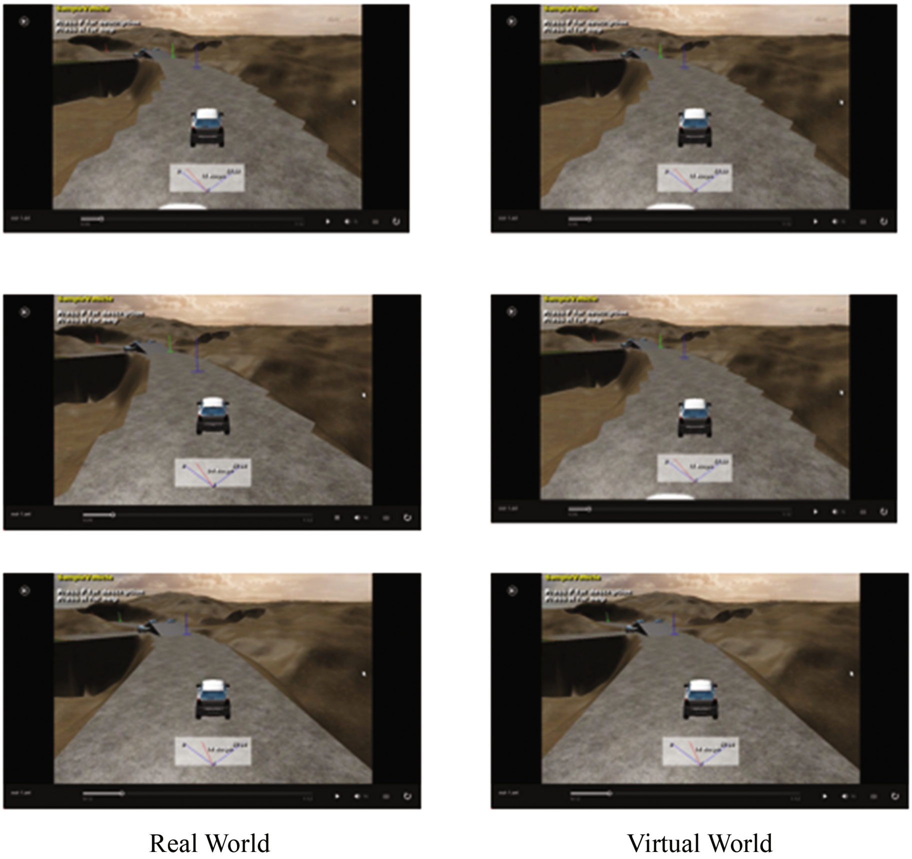


Fig. 5. The PhysX demo is used as both the real world and the virtual world, to develop the tracking capability.

The differences are not centered on the car itself; this forces the system to reason about how the differences have been caused, and to attempt to register the backgrounds.

The larger dark brown area on the left of the real image indicates that it is closer, and the car has moved towards it, and the system registers the backgrounds and adjusts the velocity of the virtual car to match the real car, shown in the bottom image.

In Fig. 6, we see another example, in which the real car turns left. In the middle pair of images, the differences caused by the turn are many. Once again, the reasoned searches to register the scenes, this time matching the cars in the background to derive the turn. The virtual car is repositioned and given the appropriate turning radius to match the real car at bottom.

Errors accumulate as the real car follows a path that is not exactly straight and the virtual car goes straight (similar to dead reckoning error). We ran a number of simulations and found that the vehicle's position needs to be updated about every 4 s on average. The computational effort required to keep the worlds registered was very small, less than 5% of total processing time.

This is not comparable to other data because nobody else seems to be doing this kind of activity modeling/comprehension, e.g. the KITTI database has no data on this.

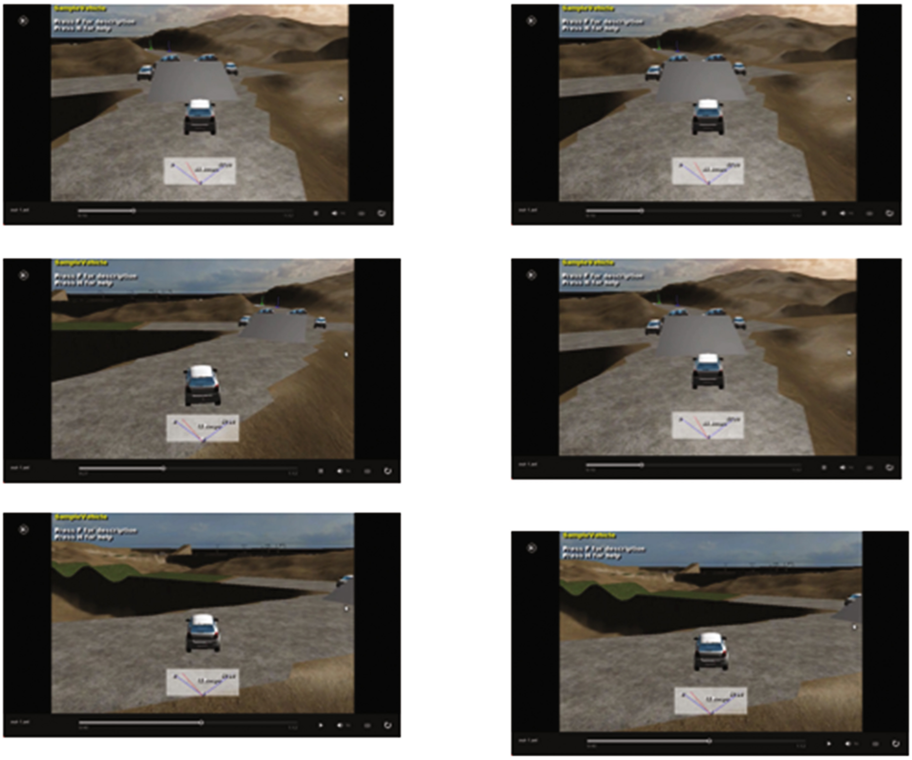


Fig. 6. Another example showing the car turning.

4 Summary

We have sketched the overall design of a cognitive computer vision system based on the structure and behavior of the human visual system. Our system builds a 3D model of a dynamic environment, updating it in real time as the world changes. Stereo cameras are moved and refocused by a cognitive architecture to build and update this model.

Further information on this work, including implementation details, can be found in [12, 13]. 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. Oliva, A., Torralba, A.: The role of context in object recognition. Trends Cogn. Sci. **11**(12), 520–527 (2008)
2. Marr, D.: Vision. W. H. Freeman, San Francisco (1982)
3. Hanson, A., Riseman, E.: Visions: a computer system for interpreting scenes. In: Hanson, A., Riseman, E. (eds.) Computer Vision. Academic Press, New York (1978)
4. Csurka, G., et al.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, vol. 1 (2004)
5. Mortensen, E., Deng, H., Shapiro, L.: A SIFT descriptor with global context. In: International Conference on Computer Vision and Pattern Recognition (2005)
6. Marques, O., Barenholtz, E., Charvillat, V.: Context modelling in computer vision: techniques, implications and applications. Multimed. Tools Appl. **51**, 303–339 (2011)
7. Ungar, S.: Cognitive mapping without visual experience. In: Kitchin, R., Freundschuh, S. (eds.) Cognitive Mapping: Past Present and Future. Routledge, London (2000)
8. Shanahan, M.P.: A cognitive architecture that combines internal simulation with a global workspace. Conscious. Cogn. **15**, 433–449 (2006)
9. Pezzulo, G., et al.: The mechanics of embodiment: a dialog on embodiment and computational modeling. Front. Psychol. **2**, A5 (2011)
10. Rayner, K.: Eye movements and cognitive processes in reading, visual search, and scene perception. In: Findlay, J.M., Walker, R., Kentridge, R.W. (eds.) Eye Movement Research: Mechanisms, Processes, and Applications, pp. 3–21. Elsevier, New York (1995)
11. Barnes, N., Liu, Z.-Q.: Embodied computer vision for mobile robots. In: ICIPS 1997, pp. 1395–1399 (1997). doi:10.1109/ICIPS.1997.669238
12. Benjamin, D.P., Lyons, D., Monaco, J.V., Lin, Y., Funk, C.: Using a virtual world for robot planning. In: SPIE Conference on Multisensor, Multisource Information Fusion (2012). http://csis.pace.edu/robotlab/pubs/SPIE2012.pdf
13. Lyons, D., Nirmal, P., Benjamin, D.P.: Navigation of uncertain terrain by fusion of information from real and synthetic imagery. In: SPIE Conference on Multisensor, Multisource Information Fusion (2012). http://csis.pace.edu/robotlab/pubs/LyonsNirmalBenjamin2012.pdf

Wearable Robotics, Industrial Robots and Construction Worker's Safety and Health

Rita Yi Man Li[1(✉)] and Daniel Ping Lung Ng[2]

[1] Real Estate Research Center / Real Estate and Economics Research Lab,
Hong Kong Shue Yan University, Hong Kong, China
ymli@hksyu.edu

[2] Department of Business Administration, Hong Kong Shue Yan University,
Hong Kong, China

Abstract. Workers' health always has an indispensable relationship with safety at work. Many different approaches are implemented on sites to improve workers' health condition. In this research study, we review robot and wearable robotics' applications on construction workers. Although there are many different types of robotic tools on sites, we conduct studies on some of the most updated tools such as high-tech robotic equipment.

Keywords: Wearable robotics · Ergonomic tool arms · Construction Safety · Health

1 Introduction

A myriad of dangerous and clumsy environmental factors in construction industry leads to poor and unsafe working conditions on sites. Besides, workers' fatigues due to long working hours and human errors are common causes that lead to construction accidents [1, 2]. In view of these, various types of robots are developed to alleviate the workers' health problems so as the various safety issues.

Robots can be applied in different kinds of construction projects on sites. They are used for placing concrete, lifting steel bars, spraying the exterior wall, window and floor cleaning [3]. Adopting robotic systems has a wide range of merit, for instance, better quality, productivity and safety. Employing robotic technologies accelerates the construction of high-rise buildings, which is also the current trend in the construction industry [4].

Japan is one of the leading country in applying robotic systems to automate construction tasks. SMART system (Shimizu Co.), ABCS system (Obayashi Co.), T-UP method (Taisei Co.), AMURAD system (Kajima Co.), and ROOF PUSH-UP construction method (Takenaka Co.) are the implementation of construction factory in the construction automation [4].

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_4

2 Three Types of Robots in the Construction Industry

2.1 Traditional Robots

Controlled by computers or other kinds of stimulus on site, robots are used to construct the superstructure of buildings autonomously [5]. Different types of robots are used in different types of construction. For example, climbing robots have been used in tall building maintenance, bridge and highways maintenance [6]. The underwater construction robot for heavy work is driven by a hydraulic system which is robust to external impact [7]. Likewise, installation of heavy building materials such as exterior curtain wall panels is often hazardous and complicated, a large amount of manpower is often needed traditionally [8]. Robots for installing window panels can help. Hence In recent year, robot is applied in various kinds of construction process on site (Table 1):

Table 1. Construction and infrastructure activities with the help of robots [9].

Types of construction activities	Example
Infrastructure	Road construction
	Tunneling
	Bridge construction
	Construction and deconstruction of power plants and dams, power plants
	Container port
Construction	Vertically and horizontally oriented buildings
	Housing construction
	Construction in artic zones, sea and deep sea, space, desert
	Building service and maintenance

2.2 Wearable Robotics

AWN-03. The AWN-03 provides back support, senses workers' motion and sends a signal to motors which rotate the gears [10]. The Suit AWN-03 embraces user's shoulder, waist and thigh [11]. In essence, it assists workers' movement when they lift and hold heavy items. It raises worker's upper body support, pushes their thighs and reduces workers' lower back stress by 15 kg [10].

The battery power pack last for six hours and each of the robot suit is sold for about (US$8,100). It is expected that there is an increase in demand for AWN-03 amid the labor shortage problems and graying of workers in the construction industries [11] (Fig. 1).

The AWN-03 provides back support, senses workers' motion and sends a signal to motors which rotate the gears [10]. The Assist Suit AWN-03 is strapped around a user's shoulder, waist and thigh [11]. In essence, it assists workers' movement when they lift and hold heavy items. It raises worker's upper body and pushes their thighs. Hence, it reduces workers' lower back stress by 15 kg [10].

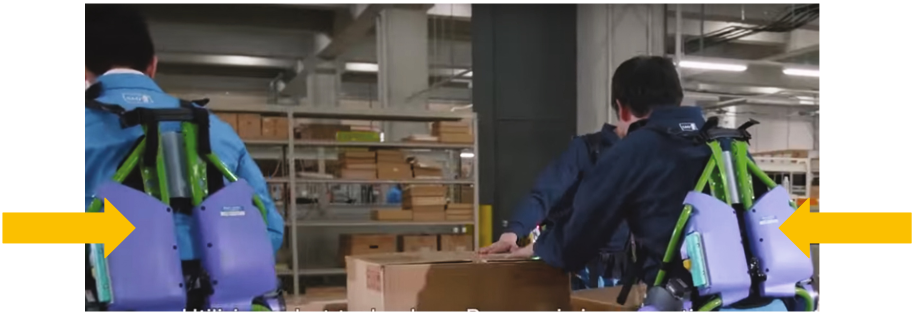


Fig. 1. Panasonic ActiveLink AWN-03 [10]

The battery power pack last for six hours and each of the robot suit is sold for about (US$8,100). It is expected that there is an increase in demand for AWN-03 amid the labor shortage problems and graying workers in the construction industries [11].

FORTIS Exoskeleton. Another wearable exoskeleton FORTIS strengthens users' strength; and costume. Similar to Iron Man, the tool is unpowered and light in weight. The external structure enhances user's endurance. It provides advantages to workers when they lifting heavy loads such as rebar and use industrial tools. It is configured to transfer loads through the skeleton to the ground when the workers stand or kneel. It creates a weightlessness feeling when wearers are maneuvering or carrying heavy objects. The exoskeleton's ergonomic design moves naturally with the wearer and able to adapt to different body heights and types [12]. Capable of supporting up to a thirty-six-pound tool, it is designed to the bucket of a cherry picker or a man lift. It is used to

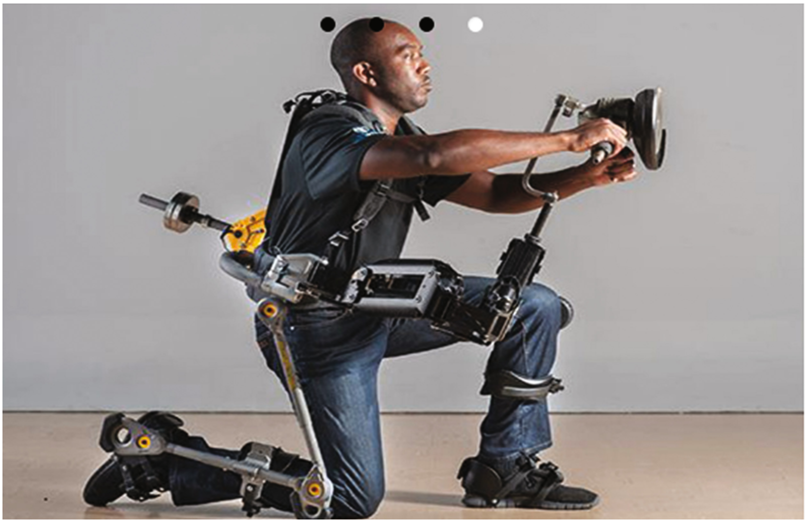


Fig. 2. FORTIS exoskeleton [13]

support large tools which may be fatigue to operate horizontally and overhead such as demo hammers, grinders, rivet busters etc [13] (Fig. 2).

Robotic Arms. Signma Ergonomics [14] suggests that a range of ergonomic tool arms has been designed to to weightlessly maneuver heavy tools. The arms rely on spring tension to balance the weight of tools used for sanding, riveting, drilling and so on. The Ekso Bionics Zero G arms can hold up to 19 kg and balance the weight. Hence, it allows the user to accurately, freely and safely maneuver the load in any direction without fatigue or injury. The ergonomic tool arms have vast numbers of mounting options which suit various different types of works such as portable gantry's carts, linear rail and jib arms. These systems require little maintenance, inexpensive inputs such as compressed air or electricity (Fig. 3).

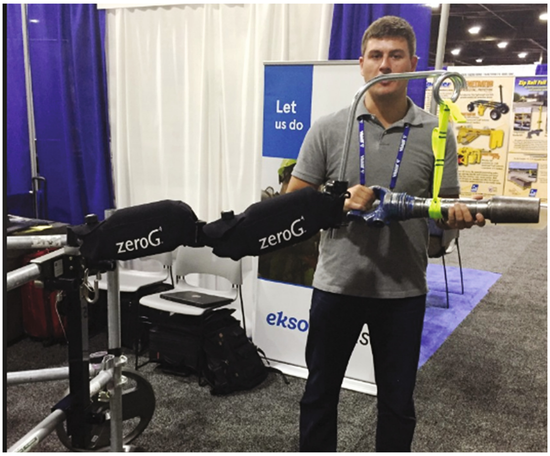


Fig. 3. Ekso Bionics Zero G arms [14]

3 Robots in Construction Industry

Heaps of the robots were related to industrial applications between 1960s' and 90s'. They were mainly used to rationalize manufacturing production. Modern robots are ubiquitous and robust. They support, nurse and accompany humans. The US market is expected to grow at 15 percent annually while the Chinese market will increase by 17 percent. Service robot market has become more important in the first decade in the 21st century [15].

4 Application of Wearable Robotics in Hong Kong Construction Industry

One of the largest Hong Kong local construction firm has applied robot to reduce the number of workers, enhance the productivity and safety. In Murray Building Hotel

Redevelopment project, the installation robot reduced the number of workers on site by 25%. Gammon Construction also purchased two sets of exoskeleton from Japan which protect their workers improve the efficiency and reduce risk. It is good to protect worker's back and waist in long period of time. However, the cost of these technology are high, it is hard for small-size company to use these technology [16].

Many of the workers in construction industry are older than 50 years old in Hong Kong. In view of the wearable robotics such as exoskeleton and robotic arm's ability to reduce strain and lighten heavy loads, applications of these tools inevitably benefit the older workers. As robotics technology continues to develop, coupled with the rapid development in programming and robot maintenance techniques, this certainly opens up new opportunity for younger staff, who are often considered to be the group that welcome the modern technology to look for innovative applications such as wearable robots [16].

Nevertheless, some workers worry that robots eventually take jobs away from us or lead to some of the scenarios in a science fiction movie. Yet, many countries and cities in the developed world experience the problem of labor shortage as less people enter the industry. Therefore, robots are simply doing their jobs in filling the gap left by due to a dwindling workforce [16].

5 Conclusion

In the modern days, robots are used on sites for structural and maintenance work, window panel installation and under water construction works. They can do high hazardous works on sites and save much labour costs. Wearable robotics raises worker's upper body and reduces workers' lower back stress. Robotic tool arms give strengths to workers and reduce workers' fatigue. As many of the construction happen on sites due to fatigue and difficult works, various types of robotic equipment reduce the likelihood of accidents on sites. Despite some workers may worry that workers will be replaced by robots and are unemployed, developed countries or cities often face the problems of labour shortage, this certainly provides a solution for construction companies and workers can work for the less hazardous work. On the other hand, wearable robotics fastens the pace of works, make them work more efficiently; reduce the workers back or arm force and the likelihood of accidents due to fatigue.

Acknowledgement. This paper is supported by University Grant Committee, Faculty Development Scheme, Construction safety index for skyscrapers in Hong Kong: A Multi-criteria decision-making approach (Project Reference No.: UGC/FDS15/E01/15).

References

1. Li, R.Y.M., Poon, S.W.: Construction Safety. Springer, Berlin (2013)
2. Li, R.Y.M.: Construction Safety and Waste Management: An Economic Analysis. Springer, Germany (2015)
3. Strukova, Z., Liska, M.: Application of automation and robotics in construction work execution. J. Interdiscip. Res. **2**, 121–125 (2012)
4. Chu, B., et al.: Robot-based construction automation: an application to steel beam assembly (part I). Autom. Constr. **32**, 46–61 (2013)
5. Niu, Y., Lu, W., Liu, D.: The application scenarios of smart construction objects (SCOs) in construction. In: Wu, Y., et al. (eds.) Proceedings of the 20th International Symposium on Advancement of Construction Management and Real Estate, pp. 969–980. Springer, Singapore (2017)
6. Tavakoli, M., et al.: Design and prototyping of a hybrid pole climbing and manipulating robot with minimum DOFs for construction and service applications. In: Climbing and Walking Robots: Proceedings of the 7th International Conference CLAWAR 2004, pp. 1071–1080. Springer, Berlin (2005)
7. Kim, H., et al.: Active control for rock grinding works of an underwater construction robot consisting of hydraulic rotary and linear actuators. In: Duy, V.H., et al. (eds.) AETA 2016: Recent Advances in Electrical Engineering and Related Sciences: Theory and Application, pp. 713–722. Springer, Cham (2017)
8. Lee, S.Y., et al.: Human-robot cooperation control for installing heavy construction materials. Auton. Robot. **22**(3), 305 (2006)
9. Bock, T.: The future of construction automation: technological disruption and the upcoming ubiquity of robotics. Autom. Constr. **59**, 113–121 (2015)
10. Panassonic: No More Power Barriers with Panasonic Assist Robots (2016). http://news.panasonic.com/global/stories/2016/44969.html. Accessed 10 Mar 2017
11. Kyodo: Panasonic to sell robot suits starting in September (2015). http://www.japantimes.co.jp/news/2015/07/03/business/tech/panasonic-sell-robot-suits-september/. Accessed 10 Mar 2017
12. Dude: Fortis Exoskeleton (2014). http://www.dudeiwantthat.com/fitness/equipment/fortis-exoskeleton.asp. Accessed 11 Mar 2017
13. Frane, D: 7 Cool and Unusal Tools From the 2016 STAFDA Show (2016). http://www.toolsofthetrade.net/power-tools/7-cool-and-unusual-tools-from-the-2016-stafda-show_o. Accessed 10 Mar 2017
14. Signma Ergonomics: Zero G Ergonomic Tool Arm (2017). https://www.sigmaergonomics.com/products/zero-g/. Accessed 10 Mar 2017
15. Haidegger, T., et al.: Applied ontologies and standards for service robots. Robot. Auton. Syst. **61**(11), 1215–1223 (2013)
16. Gammon: The Record The Rise of Robotics Gammon Technologies and the Changing Face of Construction (2017). http://www.gammonconstruction.com/uploads/files/press/the_record/The%20Record_2017%20issue%201.pdf. Accessed 10 Mar 2017

Advanced Displays and Intuitive Interfaces

Using the Soundpainting Language to Fly a Swarm of Drones

Nadine Couture[1,2](✉), Sébastien Bottecchia[1], Serge Chaumette[2], Mateo Cecconello[1], Josu Rekalde[3], and Myriam Desainte-Catherine[2]

[1] ESTIA, 64210 Bidart, France
Nadine.Couture@u-bordeaux.fr,
S.Bottecchia@estia.fr, M.Cecconello@net.estia.fr
[2] University of Bordeaux, 33400 Talence, France
{Serge.Chaumette,
Myriam.Desainte-Catherine}@u-bordeaux.fr
[3] University of Basque Country, 48940 Leioa, Biscay, Spain
Josu.Rekalde@ehu.eus

Abstract. The goal of this position paper is to explore an existing structured gestural language, called Soundpainting, to control and direct a swarm of autonomous drones. Soundpainting already integrates the notion of groups of entities and makes it possible to address one single entity of a set/subset, still being able to address the set as a whole. The key point is that Soundpainting has been designed for directing a set of improvising live performers, and, thus, we link this ability of improvisation to the capacity of decision of the autonomous drones. Indeed, Soundpainting allows a real exchange and an adaptive dialogue between the soundpainter and the group, enabling contextual interpretation by each individual, and generating rich interaction and dialogue. With a systematical analysis of the Soundpainting gestures, we show that it is perfectly adaptable to the context of the flying swarms of drones.

Keywords: Unmanned autonomous vehicles · Human-machine interface 3D gesture-based spatial interaction · Soundpainting · Swarm of drones · Smart and empowering interfaces

1 Introduction and Motivation

The starting point of our thinking is to create and perform an opera called Rain of Music. It is an Art and Science creation project in the form of an Opera made up of sound, robots and light. One of the main objectives is the spatialization of the sound in three dimensions. For this purpose we will use drones as moving sound sources. These drones will embed small speakers which will be responsible for emitting the dynamic range of a more complex soundtrack. These small flying robots can respond to choreographic commands, where movement, sound and lighting all interact. This opera proposes a change in the traditional distribution of sound where the sound sources come from fixed points (box speakers) in the space. In this opera, where the sound

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_5

sources are in constant movement, a new scientific and artistic challenge is presented. The Doppler effect, for example, which is the apparent change of frequency of a wave produced by the relative movement of the source in relation to its observer, takes on an important place in the composition and in the listening. This system will permit new experiments in sound spatialization by contrast with lots of research projects focusing on the spatialization of mobile sound sources by the use of immobile loud-speakers. We want to study the perception and rendering of mobile sound sources in 3 dimensions. As drones are becoming very common in our days artist are also willing to investigate their potential. The music pioneer John Cale, with the speculative architect Liam Young made them fly but also dance and sing. In 2014, they made a Drone Orchestra in The Barbican, London (UK), during the Digital Revolution Exhibition [1]. The question rather the piloting of a swarm of drones would make sense for real world applications in other contexts than the artistic domain, that we are considering here, still needs to be studied. We definitely believe that directly controlling a swarm by gestural interactions would not be feasible for real world applications on theatres of operations. Still, combining gestures with augmented reality glasses would make it possible to "see" a remote swarm or a swarm flying within a building could reveal useful for the army or for disaster relief for instance. To the best of our knowledge, there is no such tool available today.

In this position paper we explore the Soundpainting [2–4] approach, an existing structured gestural language, in order to control and direct a swarm of autonomous drones. Soundpainting already integrates the notion of groups of entities and makes it possible to address one entity of a set/subset, still being able to address the set as a whole. The key point is that Soundpainting has been designed for directing a set of improvising live performers, and, thus, we link this ability of improvisation to the capacity of decision of the autonomous drones. Indeed, the Soundpainting allows a real exchange and an adaptive dialogue between the soundpainter and the group, enabling contextual interpretation by each individual, and generating rich interaction and dialogue.

The remainder of this paper is organized as follows. First, we set this work with respect to related work. Then, we explore the expressive power of the Soundpainting as it can be understood by drones and we propose a set of rules to fly a swarm of drones. We conclude with pros and cons that we found and with some directions for future work.

2 State of the Art

In 2003, Draper et al. [5] compared manual and gestural input with a speech input, when operators had to control drones. The study showed that using voice commands can significantly improve an operator's ability to control subsystems. Yet nowadays, technological advances like Kinect® from Microsoft [6] allowed new modalities of interaction with a machine, passing the GUI to the NUI (Natural User Interface as defined in [7], causing a new exploration of gesture command mode. Among the studies that give the possibility to control all or parts of a swarm of drones, Monajjemi et al. [8] presented the first example of a HRI (Human Robot Interaction) with a vision-mediated gestural interface. With an un-instrumented human, it is possible to create, modify and command teams of drones. "*To create a team the user focuses attention on an individual robot by simply looking at it, then adds or removes it from the current team with a motion-based hand gesture*". Sanna et al. [9] introduce NUIs to command a drone because "*NUIs constitute a direct expression of mental concepts*". Based on this, Mashood et al. [10] proposed a set of eleven gestural postures that are supposed to enable intuitive control of mobile robots (for example, lean left to yaw left) and will extend the work to flight formation of UAVs. It appears here a proposal of structured gestures for control, but the authors did not conclude about their effectiveness. More recently, in 2014, Nagi et al. [11] proposed human-swarm interaction using spatial gestures where they use support vector machines classifiers. Following the path of these last studies, Pourmehr et al. [12] presented a control solution that combines voice and gesture depending of the context. One can, for example direct two drones to take off by designating or naming them ("*You two, take off!*", "*Not You!*"). It is then possible to dynamically create groups comprising a desired number of entities. In [13], in 2016, Morgan presented a swarm of autonomous drones (with an optimized trajectory algorithm) that react to an operator's arm movements like an "*orchestra to a director's prompts*". This study highlights a new algorithm for collision avoidance and "basic" gestural commands but does not offer a real structured gesture control command system. Beyond gesture or voice, La Fleur et al. [14] reported "*a novel experiment of BCI (Brain Computer Interface) controlling a robotic quadcopter in three-dimensional (3D) physical space using non invasive scalp electroencephalogram (EEG) in human subjects*".

As we can see in this review of the state of the art, there is an emergence of several attempts to create a grammar of gestures with the aim to control all or parts of a drone swarm. But, all of them ignore the existence of a gestural control method that would be able to lead a group or a part of it like the Soundpainting can do it.

Close to our artistic aim (a swarm of drones embedding loud speakers), in addition to the work of Cale and Young presented previously, we identified a UAV carrying a speaker for a festival in Belgium [15]. It was an association between Spotify and Base (a Belgian mobile operator). The "*Party Drone*" plays music over the entrance and the "*rest*" area, it offers free music to the festival attendees. Very recently, the "*Drone 100*" project, wich set a Guinness World Record for most UAVs airborne simultaneously, has been performed by Futurelab and Intel spectacle, and has been led by Horst Hoertner [16]. They had four pilots, each controlling a swarm of 25 UAVs. An orchestra was playing on the ground and up there the 100 drones were doing a new firework experience, a Light show, painting 3D shapes and messages in the sky.

The Soundpainting is a gestural language that has been invented by Walter Thompson in 1974 at Woodstock; it is meant for live composition. Walter Thompson proposed a language based on gestures and a well-defined grammar for conducting a large ensemble of improvising musicians without the use of any score. The Soundpainting is an interesting approach because it already integrates the notion of groups of entities and makes it possible to address one entity of a set/subset, still being able to address the set as a whole. It furthermore offers the possibility to apply an operation to different groups/subgroups. Additionally, it must be kept in mind that one of the key features when considering swarms of UAVs is that the decision process must (as of today) integrate a man in the loop; this is mandatory when decisions that require reactivity and adaptation to the environment are involved. Soundpainting relies on a man in the loop. Furthermore, the Soundpainting is positioned as an universal multi-disciplinary live composing sign language for musicians, actors, dancers and visual artists. One advantage is that the Soundpainter (the composer) can command, in live, a “swarm” of artists. To allow that, and it is a second advantage, the Soundpainting is a structured gestural language, that has been enriched by experience over time. Furthermore, the Soundpainting could be a natural as well as an efficient way to control a swarm of autonomous drones, because improvisation is an important part of that field. Could we imagine a swarm of drones getting creative? The freedom that the soundpainter would give to the swarm is a very interesting aspect for the control but also for the artistic part. The Soundpainting is then well-suited to direct and coordinate, by the use of gestures, autonomous agents. With a transversal methodology to direct both artistic and technological parameters, we claim that the Soundpainting opens up a significant direction of research into the interaction between the movements of the robots-drones, the audience and the 3D location of the sound. As far as we know, there is no tentative to fly a swarm of drones by gestures from the Soundpainting.

3 The Expressive Power of the Soundpainting is Understandable by Drones

We recall that, in formal language theory, a grammar is a set of rewrite rules used to generate strings. The rules describe how to form strings (from the language’s alphabet) that are valid according to the language’s syntax. The alphabet A, in the Soundpainting, is a set of gestures, classified in four subsets. A gesture g_{who} in the *who* subset indicates who is chosen by the soundpainter. For example: whole group, only you, group number 1. A gesture g_{what} in the *what* subset indicates what the soundpainter wants to be done. For example: define the action for the designated musicians: hold a note, laugh. A gesture g_{how} in the *how* subset indicate how the soundpainter wants the action to be done. For example: in the case of sound, loud, fast, high or low register. A gesture g_{when} in the *when* subset indicates when the soundpainter wants the action to start and/or stop. For example: now, stop.

The subsets *what* and *how* form the category “sculpture”, i.e. the type of material. The subsets *who* and *when* form the category “function”. There are 6 sub-subsets within the 4 subsets (see [4] page 8). The language syntax is defined as follow. A string s is a valid Soundpainting string if and only if,

$$s = g^{+}_{who} \cdot g^{+}_{what} \cdot g_{how} \cdot g_{when} | \, g^{+}_{who} \cdot g^{+}_{what} \cdot g_{when} \tag{1}$$

where the operator $\cdot$ is not commutative. When the g_{how} gesture is omitted, it is the target that decides of the quality and of the property of the material. Note that the grammar is not as strict as computer scientists are used to (with compilation grammar for example). For example, in Soundpainting, for a specific context, one can accept the sentence $g_{who} \cdot g_{what} \cdot g_{what} \cdot g_{when}$. As well, more than one gesture in the same category can occur in a sentence, for example several g_{who}, but the order of the categories has to be respected. Let us give some details of two examples of gestures.

Whole Group: arms above your head, create a large circle by joining your fingertips, see Fig. 1. This command specifies that everybody has to perform the action. Loudness: arms in front of you, hands forming a V. The performers modify the loudness according to the movement of the hands. If they move up, the volume of the sound is louder and vice versa. This language has been extended to the domains of dance and theatre and is now very popular. Soundpainting has been taught in public and private schools in many countries and soundpainters are numerous.

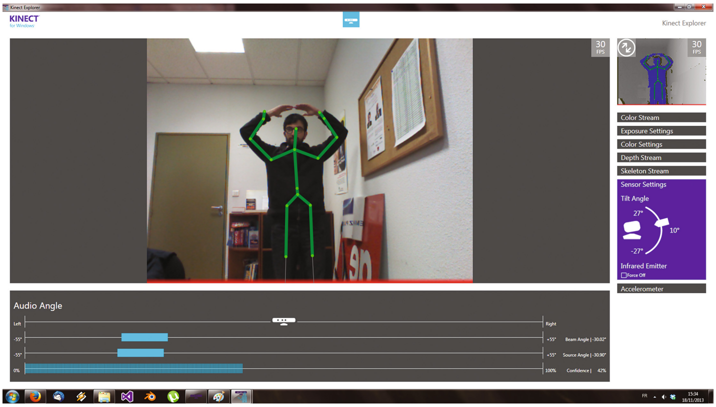


Fig. 1. The “whole group” Soundpainting gesture recognition with a Kinect.

Our objective is to explore the expressive power of the Soundpainting in the context of the flying swarms of drones, considering the piloting of both the movements of the drones and the sounds made by speakers embedded on the drones. To do so, we have, with a systematic and exhaustively approach, taken every gestures of the Soundpainting described in the two books [2, 4]. We therefore investigated the 114 gestures of the two books and for each of them, we propose a corresponding drone task. We record these gestures for the drone pilot in the 7 following tables. In column 1, we indicate the page number and the studied book where the gesture is defined. The name of the gesture is in column 2. In column 3, we give our proposition of the task that will be done by a drone, a set of drones or the whole swarm. In column 4, we indicate the scope of the task: to control the drone movement (MVT), to control the sound embedded in the drone (SND), or to control both of them. For some gestures, a secondary subset is indicated in [2, 4], for them, we draw dedicated tables. To do so, we conducted an interdisciplinary research in the fields of motion control, swarms of drones, human machine interface, arts and sciences. We used a participative design method, allowing collective creativity for a few weeks.

For the majority of the gestures, the adaptation is quite natural and represents a consensus. For them, we consider that the explicative sentence in the table is sufficient for the understanding (Table 1).

For few gestures listed in the above *what* table (Table 2), the drone task is based on an interpretation. We thus explain our choices bellow.

Table 1. Category *who*

Ref	Name	Drone task	Purpose
[2, p. 16]	whole group	all drones of the swarm	MVT SND
[2, p. 16]	brass/…/ percussion	all drones whith the same characteristic (5 possibilities)	MVT SND
[2, p. 18]	groups	a group of drones	MVT SND
[2, p. 19]	rest of group	all drones in stand by (even the ones on the floor) and executing no function	MVT SND
[2, p. 20]	watch me	the drones become non autonomous	MVT SND
[2, p. 45]	performer doesn't undertand	the drone sends a signal to the soundpainter when it doesn't understand the gesture (ex: red light)	MVT SND
[2, p. 46]	performer can't do this	the drone sends a signal to the soundpainter when it can't do the action (ex: blue light)	MVT SND
[2, p. 137]	configuration	the drone has recorded configurations and has to take one of them	MVT SND

Table 2. Category *what*

Ref	Name	Drone task	Purpose
[2, p22]	pitch up/pitch down	up and down the drone	MVT
[4, p28]	actor/dancer	all drones without speakers	MVT
[4, p62]	stand/sit	the drone takes off or lands	MVT
[4, p63]	movement	the drone, or group makes predefine movement	MVT
[4, p95]	break	the rest of the swarm leave some room for one drone (or for a group) to make a specific content	MVT
[4, p97]	change places	switch places with an other drone or just go to an other location	MVT
[4, p101]	gestures	the drone can make movements like pitch, roll or yaw	MVT
[4, p103]	horizontal movement	move on the horizontal plan	MVT
[4, p104]	jump	the drone goes up fast very shortly, like he was jumping	MVT
[4, p105]	mirror	two drones act face to face doing a mirror effect	MVT
[4, p113]	tableau	create a static picture, 3 s	MVT
[4, p113]	vertical movement	the drone moves up and down	MVT
[4, p23]	scanning	random in a set of musical phrases when the arm of the soundpainter is in front of the drone	SND
[2, p24]	pointillisme	short and arrhythmic sounds	SND
[2, p27]	minimalism	play a few notes and repeat them without variation	SND
[2, p28]	change	change drone sequence with a random sound	SND
[2, p36]		a way to indicate one sound of 4	SND

(*continued*)

Table 2. (*continued*)

Ref	Name	Drone task	Purpose
	laugth/…/ whistle		
[4, p54]	scream	the drone will make a ”screaming” noise	SND
[4, p96]	breathing	the drone makes a breathing noise	SND
[4, p98]	crossover	switch what a group is playing with an other group	SND
[4, p101]	expletives	the drones play stored insults	SND
[4, p109]	phrase	the drone will play a sentence	SND
[4, p112]	sprinkles	need a group. One of the drones will make a short and ponctual noise	SND
[4, p134]	block scanning	same as SCANNING but the soundpainter defines an area with several drones	SND
[2, p21]	long tone	the drone will play a long note or a wide movement	MVT SND
[2, p26]	relate to	the drone chooses its relations ship and choose how to follow	MVT SND
[2, p29]	memory	the configuration is memorized	MVT SND
[2, p32]	hit	short and intense sound or movement	MVT SND
[2, p35]	synchronize	synchronisation of drones	MVT SND
[2, p38]	continue	the drone keeps doing what it’s doing	MVT SND
[2, p39]	with	addition of different gestures (operator and)	MVT SND
[2, p39]	this (is)	use to put a characteristic to an other gesture	MVT SND
[2, p40]	erase	use to signal that the last gestures are cancelled	MVT SND
[2, p40]	wait	a designated drone (or a group) have to stop and wait	MVT SND
[4, p34]	default - rehearsal	define what the drone has to do when it’s on stand by, defined before the show starts	MVT SND
[4, p34]	default - performance	define what the drone has to do when it’s on stand by, defined during the performance	MVT SND
[4, p48]	freeze	the drone plays the last music like a “long tone” (SND) and/or the drone stops moving (MVT)	MVT SND
[4, p49]	anaphora	the drone repeats once the last couple of words, or movement then goes on	MVT SND
[4, p61]	layer scanning	we can’t use a SCANNING during a PLAY content, except if we use a LAYER SCANNING, the SCANNING will be superimpose to the PLAY	MVT SND
[4, p100]	drone	a slow ripple in a narrow space in between long tones	MVT SND
[4, p102]	go back to	go back to what the drone was playing/acting right before	MVT SND

(*continued*)

Table 2. (*continued*)

Ref	Name	Drone task	Purpose
[4, p102]	go on to	a sequence is send to the drone, it executes. It could be stopped only by the Soundpainter	MVT SND
[4, p104]	match	match with what an other drone is doing	MVT SND
[4, p107]	palette/palette punch	a predefine sequence of words, movement, music its played. For punch the sequence is only a few seconds	MVT SND
[4, p108]	universal palette	make a new set up and environment of the scene, defined before the show	MVT SND
[4, p111]	silence/stillness	the drone will stop playing and/or moving	MVT SND
[4, p114]	zero go onto zero	the drone will stay still and silent, then makes a small sentence or movement then go back to zero	MVT SND
[4, p135]	launch mode	after LAUNCH MODE is set, all content gestures will be executed immediately by the drones	MVT SND
[4, p136]	reverse (block) scanning	the drone stops playing and moving when the arm of the soundpainter is in front of it. Resume when the arm is gone	MVT SND
[4, p136]	tear up	use to stop LAUNCH MODE and SNAPSHOT	MVT SND
[4, p137]	back to the top	go back to the beginning of a sequence	MVT SND
[4, p138]	prepare	the drone will prepare itself to act as the following sequence	MVT SND

For the gesture MIRROR (line 10), there are two ways to sign:

$$g_{who} \cdot g_{what} \cdot g_{when} \; and\; a\; few\; seconds\; latterg'_{who} \cdot MIRROR \cdot g_{who} \cdot g_{when} \tag{2}$$

or

$$g_{who} \cdot g'_{who} \cdot MIRROR \cdot g_{what} \cdot g_{when} \tag{3}$$

In the first case the second drone has to recognize the movement of the first drone and act the same. Whereas, in the second case, the couple of drones has to decide cooperatively and autonomously which one is the leader. The latter is more complex to implement because of the synchronization phase that is required to make the decision.

For the gesture SPRINKLES (line 23) the key point is the autonomy of the set of drones, that must develop a collective intelligence in order to choose, within the swarm, one of them. Here again, a synchronization process is required to achieve an election among the drones.

In the Soundpainting, the gesture LAYER SCANNING (line 39) has been added to prevent a semantic error. Indeed, it could make sense, for the Soundpainter, to give an action order to the swarm (or swarm subset) and at the same time, to want that, at the moment the arm-scanner is pointing to one (or more) performers -for us drones-, those perfomers-drones do something else, and then come back to the first order. As it is invalid to concatenate the two orders *PLAY.SCANNING* because *PLAY* is in *when* and

SCANNING is in *what*, and also because the performer-drone does not know how to do after the scanning, the gesture LAYER SCANNING has been created to superimpose an action.

Finally, for the gesture LAUNCH MODE (line 48), we can notice that there is no g_{when} gesture at the end of the string. It is an irregular syntax that is accepted in the Soundpainting. We can consider it as a grammatical exception.

For the gesture CHANGE ADD (line 4, in the previous Table 3), concerning the sound, we interpreted it as cutting the current sound, play another sound, come back to the previous (current) one (Tables 4, 5, 6, 7).

Table 3. Category *what*, secondary category *how*

Ref	Name	Drone task	Purpose
[4, p11]	turns	the drone, a set of drones, or the swarm, will make turns on itself	MVT
[4, p128]	only	specify that the drone will play with out adding any content	SND
[4, p129]	open	cancel ONLY	SND
[4, p132]	change add	add a little thing on the content actually playing	SND
[4, p132]	change subtract	subtract a little thing on the content actually playing	SND
[4, p127]	morph	gradual transformation of the content	MVT SND

Table 4. Category *how*

Ref	Name	Drone task	Purpose
[4, p35]	direction fader	tell the drone where it has to go	MVT
[4, p52]	facing	the designated drones face each other	MVT
[4, p69]	space fader	the drones will gather or spread over the space	MVT
[4, p121]	in place	the drone stays static and can only pitch, roll and yaw	MVT
[4, p123]	floor	the drone will flight hedgehopping over the stage	MVT
[4, p41]	volume fader	volume up or down	SND
[4, p42]	tempo fader	increase or slow the tempo	SND
[4, p32]	beats	use to define the rhythms	SND
[4, p117]	classic/…/rap hip hop feel	play a specific type of music	SND
[4, p125]	level fader	indicate the range of the sound (from high-pitched to deep sound)	SND
[4, p133]	change this	keep going but change something by the soundpainter	SND

(*continued*)

Table 4. (*continued*)

Ref	Name	Drone task	Purpose
[4, p122]	develop	develop a content twice faster than in Point to Point or Scanning	MVT SND
[4, p122]	duration fader	tell how long the drone have to play/act	MVT SND
[4, p124]	head tempo	tell the drone on witch tempo it will do its content	MVT SND
[4, p127]	more space fader	implement more or less silence/stillness in the content	MVT SND
[4, p131]	stop	stop moving but keep playing	MVT SND

Table 5. Category *how*, secondary category *what*

Ref	Name	Drone task	Purpose
[4, p52]	facing	the designated drones face each other	MVT
[4, p121]	in place	the drone stays static and can only pitch, roll and yaw	MVT
[4, p117]	classic/…/rap hip hop feel	play a specific type of music	SND
[4, p133]	change this	keep going but change something by the soundpainter	SND

Table 6. Category *how*, secondary category *when*

Ref	Name	Drone task	Purpose
[4, p122]	duration fader	tell how long the drone have to play/act	MVT SND

For some of the 114 gestures, the corresponding tasks, cannot be achieved considering the current abilities of autonomous drones, it is somewhat “too complicated for a drone”. This is the case for the gestures point to point, stab freeze, extended technics, improvise, shape line and snapshot from [2], and for the gestures word, word salad, costume, dialogue, half word, narrative/melody, organically develop, play can’t play, badly/goodly, intent & feel, broken heart intent, bum intent, romantic vamp intents, asshole[1] fader, mimic fader, status fader, from [4]. The gestures contact, prop, body, initiate with (body part), face and open/close mouth from [4] are also not feasible by drones because of no body part to move, no face, and the drones can’t touch each other. This is also the case of all the gestures of Thompson’s book vol. 2 [3], that we studied

[1] We have kept the word used page 120 of [4].

Table 7. Category *when*

Ref	Name	Drone task	Purpose
[2, p43]	play	the drone, group, swarm start to play when the gesture ends	MVT SND
[2, p43]	off	the drone, group, swarm stop to play when the gesture ends	MVT SND
[2, p44]	enter slowly	start playing with a delay of 5 s approximatly	MVT SND
[2, p44]	exit slowly	stop playing with a delay of 5 s approximatly	MVT SND
[2, p45]	finish your idea	the drones finish what they are doing, delay 1 min	MVT SND

as well. It is dedicated to develop skills, adroitness of the soundpainter and his capacity to interact with performers as well as the ideas and philosophy of the Soundpainting, all this is clearly "too complicated for a drone".

For this first approach of adapting the Soundpainting, we compelled ourselves to stay as close as possible to the definition given by Walter Thompson. That led us to decide for several gestures, that the command is "too complex for a drone". But, if we drift a little from Thompson state of mind, we could adjust the gestures for drones. For example, while the gesture "dialogue" cannot be done by a standard drone, if we embed "swarm intelligence" in the UAVs, then two drones could have a chat and, thus, a "dialogue". It is one of the perspectives of our work.

4 Conclusion

With a systematic analysis of all the 114 Soundpainting gestures from [2, 4] we shown that the Soundpainting gestural language created by Walter Thompson in 1974 and regularly enriched since this date, is perfectly adaptable to the context of flying swarms of drones. Therefore, after checking the technical feasibility our approach, we have proposed a Gesture Grammar, based on 86 gestures, to pilot swarms of drones. We are aware that our proposed solution has some limitations. For example the fact that the Soundpainting language will put constraints on scenography and, by doing so, creates some limits on an artistic point of view. In choreography where machines, humans and interfaces all coexist, the use of the Soundpainting has a double risk. First, it can lead to confusion in knowing what each of the "actors" is doing. Second, while the actions of the swarm of drones are spatially mobile, the soundpainter is static and, as a part of the show, leads the spectators to focus on one point in space. This results in a contradiction on a scenographical point of view. To overcome these risks, we have to look for a creation methodology where the different artistic and technological disciplines converge.

Innovative evolution of the robots-drones opens new artistic applications for the artists. When it comes to the objectives, artists must contextualize that the cultural creations are not enough considered as an object of study by the technological research.

This has led artists to develop new research methods and new groups. Our project, Rain of Music is typical of this approach. It involves the study and transfer of new artistic possibilities based on 3D technological multi-phony and multi-casting for performance and cultural activities. Assisting the design of a multimedia choreography in 3D implies to study coordination in time and in 3D space between media, especially in an interactive context. For this purpose, we intend to extend the i-score system [17] based on experiments with artists in real conditions of live performances.

On a more general context, our perspective, beyond the definition of a Soundpainting based gestural interaction used to control individual UAVs and swarms of UAVs, is to experiment it within a simulator, to implement the operations within flying platforms, and to run real scenario. We will hopefully come with answers to the following questions: How relevant gestures are? How reliable is their detection?

Acknowledgements. Authors would like to thank the "Initiative d'Excellence" of Bordeaux University and Euskampus for their support, Baptiste Audemard (ESTIA Engineering School, France) for his contribution in the formatting of this article, Joseph Canou (ESTIA, France) for his advices in the design of drones, and Shlomo Dubnov (University of California, San Diego, US), Pierre Cauchart (SCRIME, Talence, France), Jean-Michaël Celerier (LaBRI et Blue Yeti, Bordeaux, France) for their fruitful discussions.

References

1. John, C., Liam, Y.: Loop ≫ 60hz. The Barbican, London (2014)
2. Thompson, W.: Soundpainting: the art of live composition. In: Thompson, W. (ed.) Workbook 1 (2006)
3. Thompson, W.: Soundpainting: the art of live composition. In: Thompson, W. (ed.) Workbook 2 (2009)
4. Thompson, W.: Soundpainting: the art of live composition. In: Thompson, W. (ed.) Workbook 3 (2014)
5. Draper, M., Calhoun, G., Ruff, H., Williamson, D., Barry, T.: Manual versus speech input for unmanned aerial vehicle control station operations. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 47, pp. 109–113. SAGE Publications (2003)
6. Kinect homepage. http://developer.microsoft.com/en-us/windows/kinect/hardware
7. Wigdor, D., Wixon, D.: Brave NUI World: Designing Natural User Interfaces for Touch and Gesture, 1st edn. Morgan Kaufmann Publishers Inc., San Francisco (2011)
8. Monajjemi, V.M., Wawerla, J., Vaughan, R., Mori, G.: Hri in the sky: creating and commanding teams of Uavs with a vision-mediated gestural interface. In: Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS'13), Tokyo, Japan (November 2013)
9. Sanna, A., Lamberti, F., Paravati, G., Henao Ramirez, E.A., Manuri, F.: A Kinect-Based Natural Interface for Quadrotor Control, pp. 48–56. Springer, Berlin (2012)
10. Mashood, A., Noura, H., Jawhar, I., Mohamed, N.: A gesture based kinect for quadrotor control. In: 2015 International Conference on Information and Communication Technology Research (ICTRC), pp. 298–301 (May 2015)

11. Nagi, J., Giusti, A., Gambardella, L.M., Caro, G.A.D.: Human-swarm interaction using spatial gestures. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14–18, 2014, pp. 3834–3841. IEEE (2014)
12. Pourmehr, S., Monajjemi, V., Vaughan, R., Mori, G.: "You two! Take Off!": Creating, modifying and commanding groups of robots using face engagement and indirect speech in voice commands. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2013)
13. Morgan, D., Subramanian, G.P., Chung, S.J., Hadaegh, F.Y.: Swarm assignment and trajectory optimization using variable swarm, distributed auction assignment and sequential convex programming. Int J Rob Res **35**, 1261–1285 (2016)
14. LaFleur, K., Cassady, K., Doud, A., Shades, K., Rogin, E., He, B.: Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain computer interface. J. Neural Eng. **10**(4), 046003 (2013)
15. Base, Spotify: The party drone. Belgium festivals (2014)
16. Hoertner, H.: Drone 100 Project. Flugplatz Ahrenlohe, Tornesch (2015)
17. i-score homepage: http://i-score.org

Issues and Advances in Anomaly Detection Evaluation for Joint Human-Automated Systems

Cory A. Rieth(✉), Ben D. Amsel, Randy Tran, and Maia B. Cook

Pacific Science and Engineering, San Diego, CA, USA
{CoryRieth,BenAmsel,RandyTran,MaiaCook}@Pacific-Science.com

Abstract. As human-managed systems become more complex, automated anomaly detection can provide assistance—but only if it is effective. Rigorous evaluation of automated detection is vital for determining its effectiveness before implementation into systems. We identified recurring issues in evaluation practices limiting the conclusions that can be applied from published studies to broader application. In this paper, we demonstrate the implications of these issues and illustrate solutions. We show how receiver operating characteristic curves can reveal performance tradeoffs masked by reporting of single metric results and how using multiple simulation data examples can prevent biases that result from evaluation using single training and testing examples. We also provide methods for incorporating detection latency into tradeoff analyses. Application of these methods will help to provide researchers, engineers, and decision makers with a more objective basis for anomaly detection performance evaluation, resulting in greater utility, better performance, and cost savings in systems engineering.

Keywords: Automation evaluation · Anomaly detection · Tennessee Eastman process simulation · Receiver operating characteristic

1 Introduction

Complex systems must be monitored to detect deviations from normal behavior to avert major problems. This is true in many domains, including unmanned vehicle control, robotics, industrial process control, and cybersecurity. These deviations, or anomalies, must be managed; accurate and timely anomaly detection is the first step in effective anomaly management. If operators can detect anomalies early and reliably, more time is available to diagnose, intervene, and thus prevent serious incidents [1]. For example, minor abnormalities in telemetry data might be early indicators that an unmanned vehicle is malfunctioning. If caught early, operators may be able to land and salvage the vehicle. To assist human operators in detecting problems, many automated anomaly detection approaches have been developed, including those employing simple threshold-based alarms, multivariate statistics [2] and machine learning [3].

Accurate expectations about the performance of automated anomaly detection are critical for many roles in the development of complex systems using this automation. Human factors practitioners and system designers need reliable performance estimates so they can accurately convey to operators, through user interfaces or other means, the automation's

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_6

effectiveness, limitations, and recommended conditions of use. Automation researchers need accurate estimates of performance as a basis for refining and improving the automation. Program managers need to estimate return on their investments for proposed automation. System engineers need objective evaluation methods so performance requirements can be defined and communicated to developers.

What is desired from anomaly detection performance evaluation, and what evaluation methods exist? Performance measures should be objective and unbiased, effectively characterize performance for multiple system development roles, and generalize across different anomaly detection problems. Clear high-level metrics, such as return on investment of implemented automation are often desired and widely understood. Unfortunately, these metrics are application-specific and difficult to objectively estimate. However, lower-level measures of performance are well established, general across application domains, and inform estimations of high-level impacts. Evaluating detection performance is important across disciplines, and while terminology varies (e.g., anomalies, faults, outliers), the overall basic evaluation practices and measures have commonalities, for example, measuring true (true detection rate, hit rate, sensitivity, recall) and false detections (false detection/alarm rate, specificity, precision).

Table 1. Summary of issues that were common in evaluations in anomaly detection publications, with solutions. These issues and solutions are further explored in Sects. 2, 3, and 4.

Issue (section discussed)	Implication	Solution
True detection rate reported at single threshold; setting threshold to fix false detection rate is imperfect (2)	Cannot compare performance if both true and false detection rates vary; single threshold results are not generalizable and mask performance crossovers	Use ROCs or other analysis of threshold tradeoffs
Few or single examples used for training and testing with time-varying data (3)	Biased performance estimates from between-example variability	Use multiple training and test examples (dataset provided)
Detection latency often unreported or reported only at single threshold (4)	Inability to evaluate detection latency tradeoffs with true and false detection rate	Apply latency-based techniques reviewed and introduced here

As part of a larger ongoing effort to study wider human-autonomy integration issues in anomaly management, we conducted an initial cross-domain sample of published evaluations of automated anomaly detection in unmanned systems, process control, and intrusion detection domains. Despite the maturity of detection evaluation theory and practice, three issues casting doubts about the reliability and generalizability of results were repeatedly observed: (1) How generalizable are results across applications given the inherent tradeoffs between true and false detections? (2) For time-varying data, what are the impacts of using limited data examples for evaluation? (3) How can we evaluate detection latency considering accuracy tradeoffs? Our goal in this paper is to raise awareness of these issues and provide solutions to them. Table 1 summarizes each issue (and the section describing it), the resulting implications, and proposed solutions. We used a common benchmark dataset and two common multivariate automated anomaly detection approaches: principal component analysis (PCA) and dynamic principal component analysis, DPCA). Our specific focus is on evaluation of anomaly *detection*. This is the necessary first step in a larger anomaly management process involving diagnosis and action selection, that may involve multiple automated methods. Next, we briefly review relevant concepts, introduce anomaly detection evaluation procedures, and introduce the Tennessee Eastman process simulation.

In complex systems, anomaly detection will involve multiple observed variables, and anomalies may manifest not only in single variables, but in the relationship between those variables. Generally, anomaly detection automation reduces multiple input variables to a single univariate measure of normal operating conditions. First, normal (anomaly-free) data is used to "train" the automation, establishing what is normal so that anomalies stand out in contrast. A threshold is then applied to the automation output to separate normal from abnormal behavior. If the output exceeds the threshold, it is considered a detection response. If not, it is a non-response. This threshold can be set in multiple ways, for example, from a formula or from observed data. Once the automation is trained and the threshold set, new testing data are used to assess automation performance. Testing data with anomalies are used to measure how often the automation correctly detects anomalies. Testing data without anomalies are used to measure how often the automation incorrectly reports detection with no anomaly present. See [4] for additional details on the application of several anomaly detection methods.

The issues discussed in this paper were found in many, but not all, publications on anomaly detection approaches. However, for brevity and concreteness, we focus on anomaly detection evaluated using the Tennessee Eastman process (TEP) simulation [11]. The TEP simulation is a commonly used benchmark for comparing different anomaly detection methods for process control monitoring [2, 4–10]. The simulation produces multivariate time-varying datasets with several different anomalies ("faults") that can be introduced. Despite the ability of the TEP to produce multiple datasets, a single pre-generated dataset is typically used in the literature [12]. We refer to this as the *standard dataset*, which includes a training and testing dataset both without faults and with each fault type (available at http://web.mit.edu/braatzgroup/TE_process.zip). We generated additional TEP data that is identical to the standard dataset except for the random seeds used. To facilitate future investigations we have made available (at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6C3JR1) the additional datasets used in this publication.

2 Evaluating Tradeoffs in True and False Detection

When evaluating anomaly detection automation, reported results often stress true detection rate, either fixing or separately reporting false detection rate. However, considering tradeoffs in true detection rate over a range of false detection rates can reveal convergence, divergence and even crossovers in automation performance, and enable results to be useful to a wider variety of applications. Further, ignoring these tradeoffs can have dire consequences. As a poignant example from the field of eyewitness identification, the finding that a non-standard procedure for eyewitness police lineups reduced false detection rates [13] motivated changes in police lineup procedures. However, researchers [14] later found that after controlling for tradeoffs between true and false detection (using methods discussed below), the new procedure was inferior to the original procedure. Thus, the policy changes promoted by the original conclusions were not only unnecessary, but likely reduced effectiveness and public trust in scientific experts. Although this example illustrates the importance of tradeoffs for *human* rather than automated detection, the same principles apply.

In the anomaly detection automation literature using the TEP data, detection rates are typically reported in a table with columns for each detection automation and rows for each fault type. When reported, detection latency and false detection rate are presented similarly. Yin et al. [4] compared the true detection and false detection rates of nine different anomaly detection approaches—the most comparisons we found in a single paper. However, variations in the tradeoffs between true detection rate and false detection rate make it difficult to compare across anomaly detection approaches. If one detection approach yields both a higher true detection rate and a lower false detection rate than another approach, it would be preferred. But what if the automation with the highest true detection rate also has a higher false detection rate? In [4], true detection rates of PCA and DPCA were 0.55 and 0.67, respectively, for fault 16—a DPCA advantage. However, DPCA also had a much higher false detection rate of 0.10 versus 0.06 for PCA. It would be erroneous to conclude that PCA should be preferred if false detections are more of a concern. The tradeoff between true detections and false detections can be trivially manipulated by changing the threshold used. From these results one cannot determine which automated approach would perform better if the tradeoffs between true detection and false detection rate were controlled.

Fortunately, there are well-established methods for evaluating such tradeoffs across the full range of detection thresholds. By recording true and false detection rate while varying the threshold from very strict values where no detections are recorded (0 detection rate and 0 false detection rate) to values where the detection and false detection rates are both 1, a complete picture of achievable tradeoffs can be obtained. Plotting these results with the observed false detection rate on the x-axis, and the true detection rate on the y-axis, is called a receiver operating characteristic (ROC) curve [15]. Other similar analyses such as precision-recall curves offer the same benefits. An ROC curve closer to the upper left corner indicates better performance. Figure 1 shows ROC curves for two faults in the standard TEP dataset, from which DPCA and PCA anomaly detection performance can be compared across thresholds (throughout the paper, squared prediction error is used for detection, and DPCA and PCA parameters and training procedures are identical to [2]). With an ROC plot in hand, and a desired detection or false detection rate, it is possible to generalize findings to applications with different requirements for true or false detection rate. Note that it is important to consider enough thresholds so that a clear curve is drawn out; otherwise, performance between tested thresholds may be underestimated. While the entire ROC curve can include parts of the tradeoff space where false detection rates are unacceptably high, ROC plots can be restricted to examine important regions, e.g., insets in Fig. 1. Detection automation resulting in ROC curves that are always closer to the upper left can be confidently preferred (e.g., DPCA in Fig. 1 left panel). However, results are not always clear cut. As visible in the Fig. 1 right panel inset, PCA exhibits slightly higher true detection rates at low false detection rates. For higher false detection rates, however, there is a crossover in performance, and DPCA performs better. As a modification of generic ROC curves, true and false detection rates can be transformed to rates over *time* (e.g., detections per day) to make results more relatable (e.g., to communicate with high-level stakeholders).

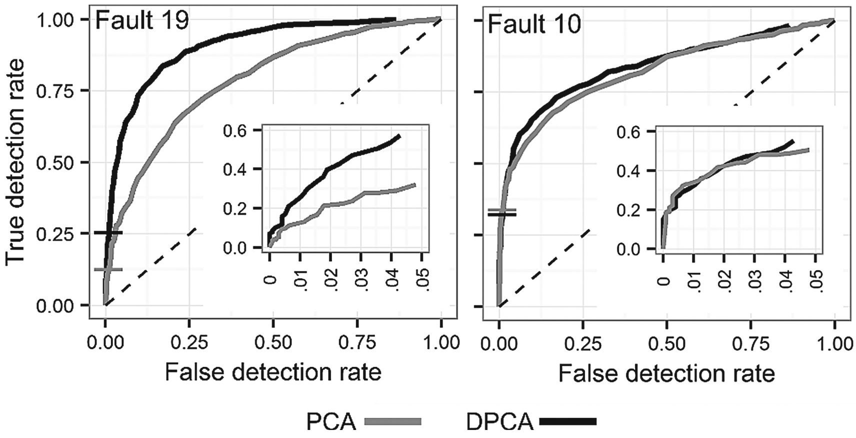


Fig. 1. Limiting evaluation to only true detection rates at single, fixed false detection rate (data from [2] shown as horizontal gray and black ticks) hides both performance at other false detection rates and potential crossovers between detection methods shown by ROC curves. Anomaly detection performance using PCA (*gray lines*) and DPCA (*black lines*) for faults 19 (left panel) and 10 (right panel) of the standard TEP dataset. The diagonal line indicates chance performance. Inset ROC plots zoom in on the low false detection region of the plot.

What if a single performance metric for each approach is desired? ROC curves and related analyses can be summarized in a single metric, for example by taking the area under the curve. However, this measure abstracts over the entire ROC curve and can therefore be misleading when used to compare detection automation that would only be operating in a subset of ROC space (e.g., false detection rates <.05), and it hides crossovers in performance between approaches. Another approach in the literature using TEP is to fix false detection rate and focus on true detection rate. However, this also hides performance crossovers at different false detection rates, limiting generalizability. For example, Russell et al. [2] report fault 10 detection rates of 0.341 for PCA and 0.335 for DPCA, a small PCA advantage, shown as horizontal ticks in Fig. 1. These single true detection rates only provide a narrow view of performance across all possible tradeoffs. From these results, someone seeking to implement an anomaly detection method for an application where higher true detection rates are needed and false detection rates are permissible might choose PCA. However, from the ROC curves in Fig. 1, right, DPCA would be a better choice.

Another issue is that attempting to fix false detection rate is not always reliable in practice. This calls into question comparisons based only on minor differences in detection rate performance. This is particularly problematic when analytical methods are used, because they often rely on statistical assumptions (e.g., normality) that may not apply. Multiple researchers have reported inaccuracies of analytical fixing of thresholds for the TEP datasets [2, 16]. Russell et al. [2] found empirical false detection rates as high as 0.28 when false detection rates were analytically “fixed” at 0.01. Consequently, they set thresholds based on empirical false detection rates. However, even empirically set false detection rates vary when independently measured (observed false detection rates between 0.009 and 0.016 were recorded by [5]). Evaluation using ROCs avoids these issues by always measuring and comparing true and false detection rate across thresholds.

3 Importance of Multiple Independent Datasets

In this section, we demonstrate that using a single training dataset and testing dataset (i.e., the standard TEP dataset) can result in biased performance evaluation. It can often be difficult or impossible to obtain enough real-world data to estimate performance with high precision [17]; further, it may be difficult to divide a stream of real-world data into normal and anomalous cases. With a simulation, however, the evaluator can both generate as much data as they desire, and control the occurrence of anomalies. The standard TEP dataset contains two examples of each fault and of normal no-fault data—one shorter example intended for training and a longer example intended for testing. Detection rates are obtained by examining performance over multiple samples in time (note both the data and univariate automation outputs are non-independent in time [6]). Effectively, anomaly detection automation is evaluated from a single training dataset and a single test example for each fault.

In theory, using only a single example of training and/or test data, even with several time samples, will result in biased measurements of performance due to variability between different simulation runs. This brings into question whether the results obtained using the standard TEP dataset are biased, and if so, by how much. An example of variability in the TEP simulation is plotted in Fig. 2. The reactor pressure variable (‘xmeas_7’) during fault 2 is shown for the standard dataset (thick black line) and 25 additional simulation runs. Note both the overall variability in the different runs, and the delay in the rise and peak of the standard dataset compared to other data.

If results from the single training and test examples in the standard TEP simulation dataset result in biased measurements, performance should differ when trained and/or tested with multiple new data sets. Figure 3 plots results from an experimental comparison of PCA-based detection (faults 10 and 16), trained and tested with different combinations of training and testing data from the single examples in the standard TEP dataset

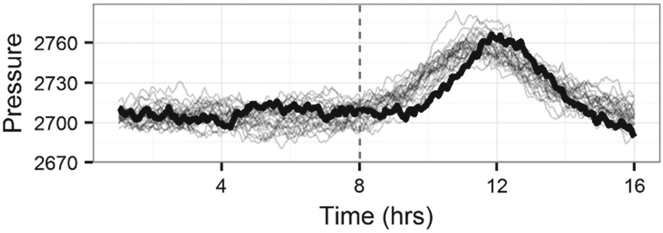


Fig. 2. Variability in the TEP simulation shown comparing data from the standard TEP dataset to additional simulation data. The plot shows reactor pressure (‘xmeas_7’) before (region left of dashed line) and during the occurrence of fault 2 for 25 new TEP simulation datasets (*thin gray lines*), and the single standard dataset commonly used for evaluation (*thick black line*).

and averaged over 500 new independently generated TEP simulation datasets. Performance biases resulting from both the single training and testing datasets were found. Bias due to using the single training example of the standard dataset was measured by comparing detection performance of automation tested on 500 new testing datasets, but trained on either the standard dataset, or 500 new training datasets. Training with the standard dataset resulted in performance overestimates of 29% and 27% for faults 10 and 16, respectively, at a 0.01 false detection rate. Though these are extreme examples, this performance overestimation from the standard training dataset occurred in 75% of faults, indicating that the standard training dataset was "more normal than normal". In some cases, the standard *testing* dataset also produced biased detection rates. This can be measured by comparing measured performance for automation trained on 500 newly generated training sets, but tested either on the standard testing set or with 500 newly generated TEP testing datasets. The measured detection rates (again for a false detection rate of 0.01) using the standard testing dataset (versus 500 simulated datasets) was underestimated by 6% in fault 10 but overestimated by 23% in fault 16. The standard testing dataset produced fault detection rates that were inflated by at least 5% in three faults, and deflated by over 5% in another three faults. These biased performance estimates could be at least as large as differences between automation approaches, and could adversely impact automation selection and implementation decisions. We have yet to explore the possibility of differential biases between automated detection approaches, which could lead to incorrect conclusions about which approach should be preferred.

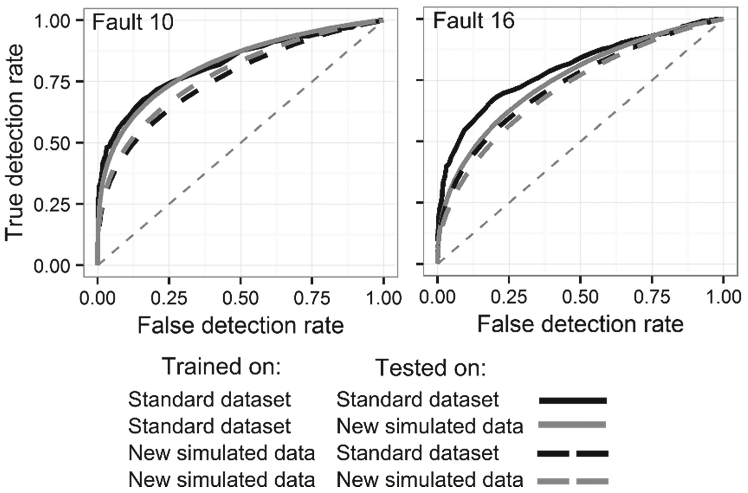


Fig. 3. Observed biases in performance measured from single training and testing examples relative to average performance over many examples. The overall bias for faults 10 (left panel) and 16 (right panel) is seen by comparing the ROCs within each plot for automation trained and tested using the standard TEP dataset (solid black) versus automation trained and tested using many new TEP simulation examples (dashed gray lines). Comparing the two gray lines, again within each plot, illustrates bias specifically due to the single *training* example, and comparing the two dashed lines illustrates bias due to the single *testing* example.

Fortunately, these biases can be easily avoided by training and testing automated detection with several independent datasets. In situations where the data-generating process cannot be simulated and automation must be trained on historical data, resampling methods (cross-validation, bootstrapping) can reduce bias.

Finally, note that when using multiple testing datasets, there is the possibility of treating detection rate as a proportion of cases where detection threshold was exceeded across datasets, rather than a proportion of detections over the threshold (as is the case above and in the TEP anomaly detection literature). Computing detection rate across data examples has the advantage that the measure is not dependent on detection latency (i.e., producing two independent measures of performance), and there is a clear detection event with which to associate latency. Conversely, detection rate as proportion of detections over threshold is a more sensitive measure and may be preferred in the face of limited training and testing data.

4 Incorporating Detection Latency in Tradeoff Analyses

Detecting anomalies as early as possible is important so that human operators have time to investigate, diagnose, and act. In situations where anomaly detection is an ongoing time-varying process (e.g., missile early warning systems, unmanned air vehicle (UAV) in-flight sensing of potential collisions), the early detection of an anomaly can be just as important to detecting the anomaly at all. Accordingly, measurement of detection latency, the time between the onset of an anomaly and when it is detected, is an important part of anomaly detection evaluation. For static data, such as isolated network connection attempts, detection latency is primarily driven by computational complexity of the testing approach and thus is unrelated to true and false detection rate tradeoffs. However, in time-varying data, such as the TEP, detection latency is strongly influenced by the detection threshold. Consider an anomaly that is initially undetectable but becomes more apparent over time. A more conservative threshold that decreases both detection rate and false detection rate will generally increase detection latency. In this section, we review methods of relating detection latency to detection thresholds that could be applied to the evaluation of anomaly detection automation. Some methods are adapted from other domains, and some are introduced for the first time here.

Despite the importance of detection latency, it is rarely reported. Among the papers using the standard TEP dataset, only two report detection latency [2, 5]. When reported, detection *latency* suffers from the same problems as detection *rates*: latencies are based on a single detection threshold, and are from a single test example. For example, it is not possible to determine if improving detection latency by allowing higher false detection rates is worthwhile. Because ROC analyses (Fig. 4A) do not include detection latency, they cannot answer this question.

There are methods from various fields, particularly epidemiology, that start to address the relationship between detection latency and accuracy that could be adapted to evaluate anomaly detection automation, including Activity Monitoring Operating Characteristic (AMOC) curves, 3D ROCs, reaction time modeling, and utility score calculations. AMOC curves [18] plot detection latency (normalized to a 0–1 scale) against false detection rate (see Fig. 4D for a variant). Two downsides to AMOCs are that detection rate, a key metric for understanding performance, is not represented and that detection latency is only included as a summary statistic. Another solution is to add detection latency as a third dimension to create a 3D ROC *surface* plot. Visual comparison of multiple surfaces in 3D plots would be more difficult than comparing ROC curves, but the volume under these surfaces can be computed and compared [19]. However, volume under the ROC surface has the same limitations of area under the ROC curve: it obscures tradeoffs and performance at specific detection rates. For detailed analysis of detection performance, methods from reaction time modeling could be adopted [20]. While these methods could yield significant insights, they can be more technically challenging to apply, and abstracted from more common metrics. Another alternative is to develop explicit costs associated with detection latency, missed detections, or false detections, from which a utility score can be calculated (e.g., [21]). However, objective and accurate specification of costs, even if possible, requires narrow focus on and substantial expertise in an application domain.

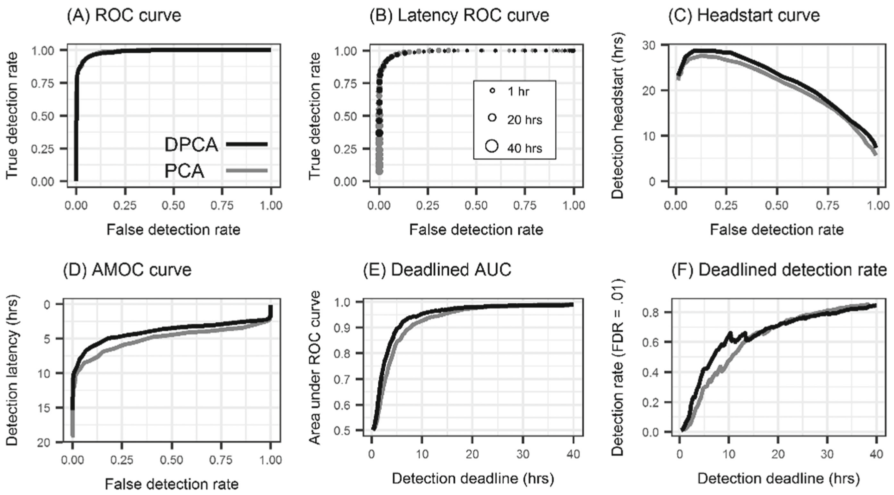


Fig. 4. Analysis of detection latency can reveal differential performance not otherwise apparent from ROC curves (A). Five different methods for evaluating detection model performance with varying strengths and weaknesses are presented (B-F). TEP fault 19 detection results from PCA (black line) and DPCA (dark gray line) detection models applied to 500 new simulated datasets are shown. See text for explanation of each method. Unlike Figs. 1 and 3, these plots use detection rate over full examples, rather than over time samples, but this is not a requirement.

We additionally developed novel evaluation methods to relate detection rates to detection latency. A simple method is to plot an ROC (Fig. 4A) above a modified AMOC (unnormalized and using median latency) (Fig. 4D). From these two plots, one can see the tradeoffs between true detection rate, false detection rate, and median detection latency. Additionally, for one value of any metric, the corresponding values of the others can be visually estimated. Another approach is to modify the ROC with a detection latency summary statistic as an additional graphical attribute, such as point color or point size (Fig. 4B). This representation of latency is quick to implement, easy to understand, and good for comparing a few models quickly, but may become too cluttered to simultaneously compare more than a few detection techniques. A third approach is to directly compare true detection and false detection latencies in "races" between anomaly present and absent trial pairs. At a given threshold, the better approach will have a greater "headstart" (anomaly absent minus anomaly present detection latency). This headstart can be plotted against false detection rate (Fig. 4C).

One shortfall in many of the above methods is that they reduce full distributions of detection latency to a single metric. Another possible evaluation method that provides a more complete picture of the full latency distribution was inspired by work in biostatistics [22]. At increasing latency deadlines from anomaly onset, the area under the ROC curve (aROC) is computed and plotted (Fig. 4E). Similarly, at each latency, the detection rate at a fixed false detection rate (e.g., 0.01; Fig. 4F) can be plotted. These methods are relatively easy to interpret, can be summarized to single values (as in [19]), give a good sense of performance across a range of latencies, and can be useful for comparing different anomaly detection approaches. However, like evaluation from aROC or at a single detection rate, these deadlined plots can obscure performance crossovers and results either cannot be generalized to other false detection rates (deadlined single detection rate) or do not allow for inspection at specific false detection rates (deadlined aROC curve). Note that all of these methods could be easily adapted to precision recall curves. Overall, consideration of several different evaluation analyses and metrics, rather than always relying on one or two, is advantageous.

5 Discussion

The costs and frequency of incidents and mishaps across domains (e.g., [23, 24]) underscore the criticality of effective anomaly management. Improving joint human-automated system performance by introducing automated anomaly detection requires generalizable, unbiased, and thorough evaluation practices. We identified three common shortfalls in evaluation practices, demonstrated how they reduce the ability to confidently draw conclusions about detection performance, and offered guidance and novel methods to improve evaluation practices.

The first common shortfall was reporting true detection rate for a single detection threshold, either in combination with a separate false detection rate or attempting to fix false detection rate. Without a full awareness of the tradeoffs between measured true and false detection rates, it is difficult to determine anomaly detection automation's usefulness in helping humans appropriately classify events as normal or anomalous. This difficulty is especially evident when applying results across applications requiring different false detection rates. Using the standard TEP dataset, we demonstrated how ROC curves can reveal crossovers where the "best" model varied over false detection rates. We also demonstrated the limitations of single detection metrics in published studies compared to ROC analyses that characterize true detection rates across the full spectrum of false detection rates, allowing generalization of performance results.

The second issue was the use of small datasets, or even single examples for evaluation. This can bias performance evaluation due to the unknown discrepancies between sample statistics and the larger population parameters. We compared performance averaged over 500 independent runs of the TEP simulation to performance measured from the single examples of the standard TEP dataset. The results revealed general performance overestimations resulting from the standard training dataset, and varied biases when using the standard testing dataset. The combined effects led to 10% or greater biases to performance estimates in 30% of fault types.

The third issue, relevant for detecting anomalies unfolding in time-varying data, concerns reporting of detection latency evaluation. In the few publications that do include detection latencies, tradeoffs between detection rate and latency as a result of varying the detection threshold were not assessed. Further, while ROC and related analyses assess tradeoffs between true and false detections, they do not include detection latency and cannot address three-way tradeoffs. In this paper, we highlighted existing approaches to incorporating detection latency into evaluation, and introduced novel approaches. Our hope is that raising awareness of these latency evaluation methods will facilitate the development of standard practices for latency-based analyses.

While we have focused on automated anomaly detection, and specifically efforts using the TEP simulation, the evaluation methods reported here apply broadly to *any* evaluation of detection performance, including detection performance of human operators aided by automated anomaly detection [25]. These methods can help researchers and practitioners better compare anomaly detection automation and stimulate the development of more effective automation approaches. Additionally, this work can provide decision makers with an objective basis for automation selection and implementation decisions. Further, sound assessment of detection performance is foundational for tools to support diagnosis and intervention for wider anomaly management.

Acknowledgments. This work was sponsored by the Office of Naval Research, Human & Bioengineered Systems (ONR 341), program officer Dr. Jeffrey G. Morrison under contract N00014-15-C-5003. The authors thank Drs. Harvey Smallman and Shannon Toftely for their advice and assistance. The views expressed are those of the authors and do not reflect the official policy or position of the Office of Naval Research, Department of Defense, or US Government.

References

1. Burns, C.M.: Towards proactive monitoring in the petrochemical industry. Saf. Sci. **44**, 27–36 (2006)
2. Russell, E.L., Chiang, L.H., Braatz, R.D.: Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. Chemomther. Intell. Lab. Syst. **51**, 81–93 (2000)
3. Sommer, R., Paxson, V.: Outside the closed world: on using machine learning for network intrusion detection. In: 2010 IEEE Symposium on Security and Privacy, pp. 305–316 (2010)
4. Yin, S., Ding, S.X., Haghani, A., Hao, H., Zhang, P.: A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. J. Process Control **22**, 1567–1581 (2012)

5. Mahadevan, S., Shah, S.L.: Fault detection and diagnosis in process data using one-class support vector machines. J. Process Control **19**, 1627–1639 (2009)
6. Rato, T.J., Reis, M.S.: Fault detection in the tennessee eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR). Chemomther. Intell. Lab. Syst. **125**, 101–108 (2013)
7. Lee, J.-M., Yoo, C., Lee, I.-B.: Statistical process monitoring with independent component analysis. J. Process Control **14**, 467–485 (2004)
8. Cheng, C.-Y., Hsu, C.-C., Chen, M.-C.: Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes. Ind. Eng. Chem. Res. **49**, 2254–2262 (2010)
9. Hsu, C.-C., Chen, M.-C., Chen, L.-S.: Integrating Independent Component Analysis and Support Vector Machine for Multivariate Process Monitoring. Comput. Ind. Eng. **59**, 145–156 (2010)
10. Lee, J.-M., Qin, S.J., Lee, I.-B.: Modified independent component analysis for multivariate statistical process monitoring. IFAC Proc. **39**, 1133–1138 (2006)
11. Downs, J.J., Vogel, E.F.: A plant-wide industrial process control problem. Comput. Chem. Eng. **17**, 245–255 (1993)
12. Chiang, L.H., Russell, E.L., Braatz, R.D.: Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Chemomther. Intell. Lab. Syst. **50**, 243–252 (2000)
13. Lindsay, R.C., Wells, G.L.: Improving eyewitness identifications from lineups: simultaneous versus sequential lineup presentation. J. Appl. Psychol. **70**, 556 (1985)
14. Gronlund, S.D., Mickes, L., Wixted, J.T., Clark, S.E.: Conducting an eyewitness lineup: how the research got it wrong. Psychol. Learn. Motiv. **63**, 1–43 (2015)
15. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**, 861–874 (2006)
16. Lee, J.-M., Qin, S.J., Lee, I.-B.: Fault detection and diagnosis based on modified independent component analysis. AIChE J. **52**, 3501–3514 (2006)
17. Khalastchi, E., Kalech, M., Kaminka, G.A., Lin, R.: Online data-driven anomaly detection in autonomous robots. Knowl. Inf. Syst. **43**, 657–688 (2014)
18. Fawcett, T., Provost, F.: Activity monitoring: noticing interesting changes in behavior. In: Proceedings of Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 1, pp. 53–62 (1999)
19. Kleinman, K.P., Abrams, A.M.: Assessing surveillance using sensitivity, specificity and timeliness. Stat. Methods Med. Res. **15**, 445–464 (2006)
20. Wagenmakers, E.-J., van der Maas, H.L.J., Dolan, C.V., Grasman, R.P.P.P.: EZ does it! Extensions of the EZ-diffusion model. Psychon. Bull. Rev. **15**, 1229–1235 (2008)
21. Lavin, A., Ahmad, S.: Evaluating real-time anomaly detection algorithms–the Numenta anomaly benchmark. In: 2015 IEEE 14th International Conference on Machine Learning and Application (ICMLA), pp. 38–44 (2015)
22. Heagerty, P.J., Zheng, Y.: Survival model predictive accuracy and ROC curves. Biometrics **61**, 92–105 (2005)
23. Hewitt, P., Lane, M.A.: Human factors engineering delivers ROI. Control Glob. 1–2 (2017)
24. Tvaryanas, A.P., Thompson, W.T., Constable, S.H.: Human factors in remotely piloted aircraft operations: HFACS analysis of 221 mishaps over 10 years. Aviat. Space Environ. Med. **77**, 724–732 (2006)
25. Sorkin, R., Woods, D.D.: Systems with human monitors: a signal detection analysis. Hum. Comput. Interact. **1**, 49–75 (1985)

Maintain and Regain Well Clear: Maneuver Guidance Designs for Pilots Performing the Detect-and-Avoid Task

Kevin J. Monk[1](✉) and Zachary Roberts[2]

[1] NASA Ames Research Center, Moffett Field, CA 94035, USA
kevin.j.monk@nasa.gov

[2] San Jose University Research Foundation, Moffett Field, CA 94035, USA
zachary.s.roberts@nasa.gov

Abstract. In order to support the future expansion and integration of Unmanned Aircraft Systems (UAS), ongoing research efforts have sought to produce findings that inform the minimum display information elements required for acceptable UAS pilot response times and traffic avoidance. Previous simulations have revealed performance benefits associated with DAA displays containing predictive information and suggestive maneuver guidance tools in the form of banding. The present study investigated the impact of various maneuver guidance display configurations on detect-and-avoid (DAA) task performance in a simulated airspace environment. UAS pilots' ability to maintain DAA well clear was compared between displays with either the presence or absence of green DAA bands, which indicated conflict-free flight regions. Additional display comparisons assessed pilots' ability to regain DAA well clear with two different guidance presentations designed to aid in DAA well clear recovery during critical encounters. Performance implications and display considerations for future UAS DAA systems are discussed.

Keywords: Human-systems integration · Human factors · UAS · User interface design · Detect and avoid · Safety · Aviation · Aeronautics · Displays

1 Introduction

Current day applications of Unmanned Aircraft Systems (UAS) primarily involve military operations in restricted airspace. However, civil and public-use UAS are expected to fly alongside manned commercial aircraft unrestricted across other airspace classes within the National Airspace System (NAS) in the coming years. Subject matter experts from academia, industry, and government are currently developing minimum operational performance standards (MOPS) in order to maintain safety of flight with the integration of UAS in the NAS [1, 2]. Currently, federal regulations require manned pilots to "see and avoid" other aircraft to remain "well clear" [3]. Since UAS pilots are positioned at a ground control station (GCS) without the ability to visually detect potential threats from inside the cockpit, they will require a "detect and avoid" (DAA) system that provides the information necessary to identify a threat and make an appropriate maneuver with the command and control interface [4]. The minimum amount of

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_7

information on the DAA display needed to detect conflicts, determine a resolution, and avoid losses of well clear safely has been the focus of ongoing research within NASA's UAS integration in the NAS (UAS-NAS) project.

Previous research has explored the minimum DAA display requirements necessary to perform UAS pilot tasks. Predictive displays that include avoidance zones, color-coded alerting, intruders' relative closest-point-of-approach (CPA), and directional icons with conflict alerting have been shown to reduce losses of DAA well clear (DWC) and minimize the severity of collision hazards when they do occur [5, 6]. A survey assessing pilots' visual information preference identified intruder state information, visual alerts, and DAA maneuver recommendations as important information elements [7].

Human-in-the-loop simulations have revealed performance benefits with displays containing advanced conflict resolution tools integrated into the vehicle control interface on the ground control station (GCS) [8–10]. Specifically, suggestive maneuver guidance in the form of banding that provided continuous indications of the predicted threats (i.e., losses of DWC) at nearby headings and altitudes have yielded the most desirable benefits compared to an informative display [11, 12]. While DAA banding guidance that indicated the predicted threat severity level was accepted as a requirement in the Phase 1 MOPS for the DAA system, it was not specified whether it is necessary to incorporate green, conflict-free bands that highlight flight regions absent of any predicted loss of DWC threats. Furthermore, the suggestive maneuver guidance in previous evaluations informed maneuvers for maintaining well clear with sufficient time, but did not always provide guidance to aid in *regaining* well clear in the more severe cases where a near midair collision (NMAC) was imminent. Recovery guidance presented to pilots was not directly assessed for its effects, and there were too few losses of DWC with the banding displays to make post hoc statistical comparisons. Therefore, the present study utilized a test setup that allowed for a more complete evaluation of the conflict-free maneuver guidance bands and various well clear recovery guidance design concepts with respect to pilots' ability to maintain and regain well clear.

2 Method

2.1 Participants

Ten pilots were recruited for participation in the current experiment. Six of ten were active duty UAS pilots (M_{age} = 36 years old), with averages of 1600 h of manned flight experience and 1400 h of unmanned flight experience. The other four were commercial pilots (M_{age} = 30 years old) with an average of 9000 h of manned flight experience in civil airspace.

2.2 Simulation Environment

Ground Control Station. The GCS used for this study was the Vigilant Spirit Control Station (VSCS), which was developed by the Air Force Research Laboratory (AFRL) [13]. For the purposes of the current study, VSCS provided a Tactical Situation Display

(TSD) that displayed ownship, mission route, DAA maneuver guidance, and traffic information over a moving map. The TSD supported an autopilot control interface that allowed pilots to change their altitude and heading without manipulating the pre-filed flight plan route. Heading holds could be executed via the graphical compass rose interface or keypad inputs to a steering command window. The compass rose interface allowed pilots to click-and-drag an arrow shaped heading bug to the desired heading rather than manually input numbers. Altitude values could be changed either by manually typing in desired values or by using two small arrows ("spinners") that would increase/decrease altitudes by 500 feet (ft.) per mouse click. Pilots uploaded commands to the aircraft by clicking the "Send" button located within the steering window. On a separate monitor only visible by the researcher, another component of the Vigilant Spirit software ('Vigilant Spirit Simulation') allowed researchers to manually launch the pre-scripted encounters toward the ownship.

DAA System. The multi-level alerting structure was constructed through the Java Architecture for DAA Modeling and Extensibility (JADEM v.5.4.1) [14]. Intruders equipped with transponders (i.e., cooperative) were displayed on the TSD at sensor ranges of 15 nautical miles (nmi) laterally and ±5000 ft. vertically, while intruders with RADAR-only equipage (i.e., non-cooperative) were detected at a lateral range of 8 nmi with the same vertical range. Color-coded symbology was applied to all aircraft within sensor range to provide pilots with indications of individual threat severity, based on whether they were currently predicted to penetrate the spatial and temporal thresholds pre-defined for the current study (see Table 1). Direct auditory alerts were presented as the threat severity levels of the intruders increased. The intruders' relative altitude and vertical trend were also constantly visible once they were within sensor range. Other intruder elements that appeared within the data tag at the onset of a conflict alert included call sign (cooperative intruders only), ground speed, absolute altitude, and vertical velocity.

Table 1. Multi-level alerting scheme

Alert level	Separation criteria	Time to loss of DWC	Icon	Aural alert verbiage

DAA warning alert	HMD = 0.75 nmi
	ZTHR = 450 ft.
	modTau = 35 s
		25 s	▲	"Traffic, Maneuver Now, Traffic, Maneuver Now"

Corrective DAA alert	HMD = 0.75 nmi
	ZTHR = 450 ft.
	modTau = 35 s
		55 s	●	"Traffic, Avoid"

Preventive DAA alert	HMD = 0.75–1.0 nmi
	ZTHR = 450–700 ft.
	modTau = 35 s
		N/A	○	"Traffic, Monitor"

None (target)	Within surveillance field of regard	N/A	△	N/A

The DAA maneuver guidance ('Omni Bands'), also generated by the JADEM DAA system, provided pilots with a form of conflict resolution using dashed lines ('banding') that predicted whether particular heading or altitude values were predicted to cause loss of DWC threats if flown at that time. The horizontal bands probed relative headings within 270° around ownship and appeared on the inner range ring of the moving map. The vertical bands probed altitudes within ±2,000 ft. of ownship on the altitude tape to the right of the TSD. The maneuver guidance bands were constantly updating to reflect the most up-to-date flight state information, as JADEM did not account for ownship or intruder intent. The heading and altitude bands were color-coded in correspondence with the predicted threat level from the alerting structure. Headings and altitudes with yellow bands were predicted to lead to a loss of DWC (as defined in Table 1) with an intruder aircraft within 25–55 s. Red banding indicated that a particular heading or altitude would lead to a loss of DWC within 25 s or less. Thus, regions with yellow or red banding were to be avoided, as maneuvers toward these areas would trigger at least one Corrective DAA or DAA Warning alert, respectively. Safe flight regions that would remain well clear with intruders were indicated by either the presence of green banding or the absence of banding, depending on the condition (see "Experimental Design").

Once resolution options for remaining well clear were no longer achievable by ownship, the bands would fully saturate to red and well clear recovery guidance was presented on the TSD. Though likely that the well clear boundary had already been penetrated in these worst-case scenarios, it was considered necessary to provide pilots with some form of maneuver guidance as a last resort to help minimize the severity of the separation loss and regain DWC. In order to achieve this, the well clear recovery guidance calculated the direction that would lead to the maximum separation at the CPA. The underlying computations used to suggest conflict resolutions were based on the Generic Resolution Advisor and Conflict Evaluator (GRACE) algorithm [15], which evaluates multiple intruders for threats based on the aforementioned separation standards. The GRACE maneuver selection logic generated conflict-free solutions with considerations made to current intruder flight states and, in well clear recovery cases, presented either the lateral or vertical maneuver suggestion with the lowest 'NMAC cost'.

The recovery guidance was displayed to pilots both textually and graphically on the TSD. The graphical representation of the maneuver recommendation varied among trials (see "Experimental Design"). The textual guidance for both recovery displays was shown at the top of the TSD inside of a green border, labeled with commands of either 'Turn Right', 'Turn Left', 'Climb', or 'Descend'. The recovery guidance text would switch to 'Maintain' to inform pilots that they may remain at their current heading and altitude for the time being once they reached the flight state necessary to maximize their separation with the surrounding intruder(s). Once ownship regained well clear with the intruder(s) and the bands were no longer saturated red, the recovery guidance as a whole was no longer displayed.

2.3 Experimental Design

The current experiment utilized a one-way and repeated measures design to examine the impact of green DAA bands (With vs. Without) and band saturation display options (Limited Suggestive vs. Directional) on pilots' DAA task performance.

Green DAA Bands. In the previous simulation that introduced Omni Bands [11], green banding was used to denote safe flight regions that would not result in a loss of DWC if flown at that time. Headings and altitudes that were not probed (i.e., the 90° behind ownship, >3,000 ft relative altitude) did not have any banding. During the open-ended portion of the debriefs, pilots voiced varied opinions on the usefulness of the green bands for conflict avoidance, with some stating that they added display clutter in conditions using Omni Bands (albeit quantitative analysis of questionnaire responses with regard to display clutter did not reveal statistical significance). The present study sought to examine whether the presence (or absence) of green DAA banding had any impact on pilot performance and response times; thus, it was added as a between-subjects manipulation. Half of the pilots saw green DAA bands which differentiated safe regions from those that were un-probed or would create conflict (at the Corrective DAA or DAA Warning levels), while the others did not have green DAA bands displayed to them (Fig. 1). The pilots that did not have use of the green bands saw a blank, defaulted (grey) presentation of their inner range ring and altitude tape (similar to when un-probed) until there was a potential conflict within sensor range that triggered yellow and/or red bands on the display. Pilots without green DAA bands were instructed to avoid conflicts by flying into regions with no banding present.

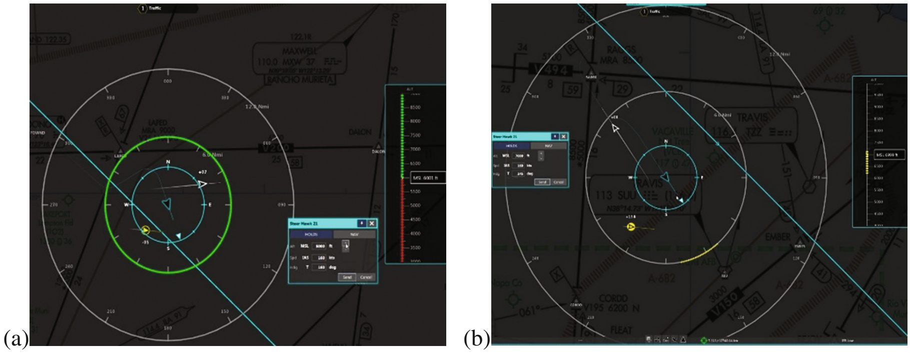
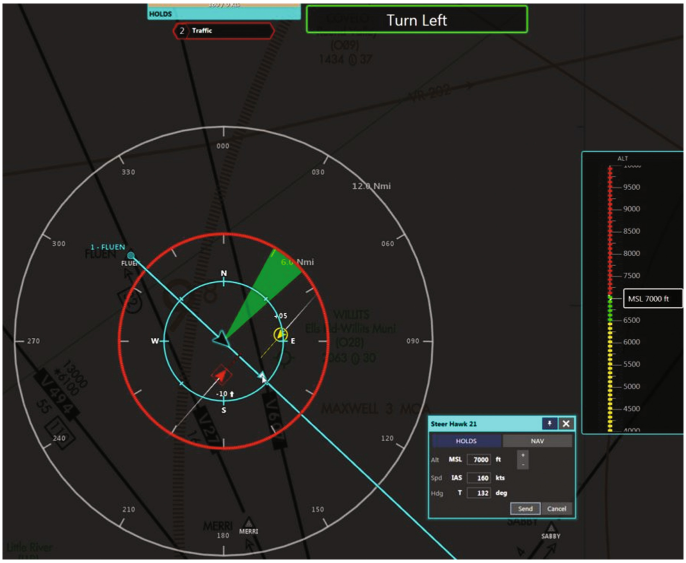


Fig. 1. TSD with (a) and without (b) green bands displayed.

Well Clear Recovery Guidance. The Well Clear Recovery guidance display(s) appeared when a loss of DWC could no longer be avoided. There were two graphical representations of guidance presented to pilots for regaining well clear: Limited Suggestive and Directional. The Limited Suggestive recovery guidance displayed the range of optimal headings or altitudes to fly in order to maximize separation (Fig. 2). Low and high bounds of a recommended altitude or heading range were provided to achieve a timely regain of well clear. A green 'wedge' encompassing the suggestion range



Steer Hawk 21

HOLDS | NAV

Alt MSL 7000 ft

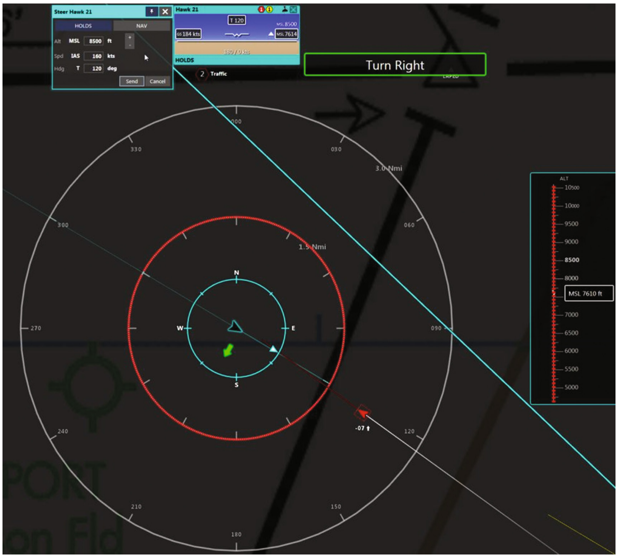
Spd IAS 160 kts

Hdg T 132 deg

Send Cancel

Fig. 2. Limited suggestive well clear recovery guidance.

appeared next to ownship and extended out to the range rings when the algorithm recommended a turn; pilots were to comply by flying to headings within the suggestive wedge. If the recovery algorithm recommended a vertical maneuver for collision avoidance, pilots were to aim for altitudes within the green wedge that appeared on the altitude tape.



Steer Hawk 21

HOLDS | NAV

Alt MSL 8500 ft

Spd IAS 160 kts

Hdg T 120 deg

Send Cancel

Fig. 3. Directional well clear recovery guidance.

The Directional recovery guidance simply indicated the suggested maneuver type by displaying an arrow in the recommended direction of the maneuver (Fig. 3). A green arrow appeared pointing to either the left or right of ownship when the recovery guidance was recommending a turn. For vertical maneuver recommendations, an up or down arrow appeared to the left of the altitude tape to suggest a climb or descent. Directional recovery guidance did not specify a specific range of headings or altitudes to choose from, instead allowing the pilot to determine the size of the maneuver in the recommended direction and sense.

2.4 Procedure

Training. Pilots began the day by filling out demographics and informed consent forms. They were then given a short briefing on the experiment before being trained on the basic functionalities within VSCS. Once pilots demonstrated proficiency with the vehicle control inputs required to maneuver the simulated aircraft, they were trained on the various components of the DAA system (above). Pilots assigned to the display configuration with no green bands for remaining well clear were trained with slides that excluded any mention of green bands. Each experimental scenario was preceded by a training run that allowed pilots to practice interaction with the upcoming well clear recovery display configuration. The banding options for remaining well clear varied between subjects, while the recovery display options varied between trials (within subjects).

DAA Pilot Task. Pilots completed four 40-minute scenarios—two with each well clear recovery guidance display. The order of presentation was counterbalanced between participants. Pilots were instructed to navigate a simulated MQ-9 Reaper along a route line while avoiding well clear violations and NMACs with nearby intruders. The scenarios consisted of 20 encounters lasting approximately two minutes each. Sixteen of the 20 encounters were scripted to lose well clear absent any pilot action, with eight of them involving conflicts that blundered into ownship and forced an immediate well clear violation. Since there were hardly any losses of DWC observed for displays utilizing green DAA banding in the previous simulation [11], it was necessary to introduce severe encounters that allowed for the evaluation of pilots' ability to regain well clear with each recovery band display option. Pilots were instructed not to begin editing their trajectory until the onset of a DAA alert required them to do so. Once pilots complied with guidance and successfully avoided an aircraft, they were to return to course and fly along the route line until their next encounter was triggered.

2.5 Measures

Initial Response Time (Initial RT). Initial Response Time refers to the amount of time it took for pilots to initiate a navigational edit into the GCS after the onset of a Corrective DAA or DAA Warning alert.

Total Edit Time. Total Edit Time refers to the amount of time it took for pilots to complete their final upload into the navigation interface after starting their initial edit.

Total Response Time (Total RT). Total Response Time refers to the full amount of time it took for pilots to upload their final resolution after the onset of a Corrective DAA or DAA Warning alert.

Loss of DWC (LoDWC) Severity. As of the present study, the overall severity of each well clear violation was identified by the DAA Well Clear Penetration Integral (DWCPI) metric [16], which combined the amount of time spent within the well clear threshold and the minimum geometric separation at CPA into a single measure. The greater the DWCPI magnitude for a given encounter, the more severe the loss of DWC event was considered.

3 Results

3.1 Green DAA Bands

The response time metrics were analyzed across the two banding displays using a one-way Analysis of Variance (ANOVA), with an alpha level of .05. The analysis included the encounter cases that required pilots to remain DWC (i.e., when positive maneuver guidance bands were available). The LoDWC severity results for this variable were not tested for statistical significance, as there was only one loss of DWC occurrence among all nominal encounters with each banding display.

Initial RT. There was a significant main effect of green bands on initial response times, $F(1, 281) = 13.10$, $p < .05$. Initial RTs were, on average, 1 s quicker with the No Green Bands display (M = 5.00 s, SE = 0.32 s) compared to the Green Bands display (M = 6.00 s, SE = 0.25 s).

Total Edit. There was only a marginal effect of green bands on total edit times, $F(1, 283) = 3.74$, $p = .054$. Pilot completed their edits, on average, 0.86 s quicker with the Green Bands display (M = 3.87 s, SE = 0.26 s) compared to the No Green Bands display (M = 4.73 s, SE = 0.37 s).

Total RT. There was no significant effect of green bands found on total response time, $p > .05$.

3.2 Well Clear Recovery Type

The response time and separation metrics were analyzed across the two well clear recovery displays using a repeated measures ANOVA, with an alpha level of .05. Statistical comparisons were made across encounters that presented well clear recovery guidance to regain well clear, including the critical cases that blundered into ownship.

Initial RT. There was no significant difference in initial response times found between the Limited Suggestive (M = 3.98 s, SE = 0.59 s) and Directional (M = 3.61 s, SE = 0.36 s) display configurations, p > .05.

Total Edit. There was no significant difference in total edit times found between the Limited Suggestive (M = 5.20 s, SE = 1.32 s) and Directional (M = 5.71 s, SE = 1.44 s) display configurations, p > .05.

Total RT. There was no significant difference in total response times found between the Limited Suggestive (M = 9.17 s, SE = 1.25 s) and Directional (M = 9.31 s, SE = 1.32 s) display configurations, p > .05.

LoDWC Severity. While loss of DWC events were, on average, slightly less severe with the Limited Suggestive display configuration (M = 0.86, SE = 0.19) compared to the Directional display (M = 1.34, SE = 0.32), the difference in DWCPI magnitude was nonsignificant, p > .05.

4 Discussion

4.1 Conflict-Free Bands for Maintaining DAA Well Clear

The results suggest that DAA guidance in the form of banding is effective at aiding the pilot responsibility of remaining well clear, regardless of whether green bands are implemented to highlight the well clear regions. Initial RT was the only response time metric that yielded a (significant) difference of over one second between conditions. Pilots utilizing the green bands started their initial edits following a DAA Corrective alert 1.5 s slower on average. There may be slightly less of a processing delay when simply monitoring the onset of conflict bands versus detecting continuous changes in the color of bands constantly visible on the display. Nonetheless, conflicts were successfully avoided at a nearly equal rate across banding displays overall. There was only 1 LoDWC (<1% of total encounters) with each banding display, and the LoDWC proportion was comparable to previous analyses observing non-blunders at nominal encounter ranges sufficient to remain well clear [16]. The Phase 1 DAA MOPS require a distinction between the yellow corrective and red warning guidance bands, while the implementation of green (or any color) conflict-free maneuver guidance bands are considered optional [17].

4.2 Recovery Guidance for Regaining DAA Well Clear

Well clear recovery display type failed to significantly impact any of the response time or separation variables in the present study. Response times were nearly identical between recovery displays, as there was a difference of a half-second or less on every response time metric. It should be noted that no large response time differences between recovery types were expected, as well clear recovery did not appear until the pilot could no longer maintain DAA well clear. Once the bands were fully saturated red and the threat severity reached the critical Warning level, an immediate maneuver in compliance with the well clear recovery bands was the expected pilot action. While pilots were trained to comply with the guidance, it was left to their discretion whether it was deemed appropriate. Minimal decision-making was required when pilots made immediate maneuvers in compliance with the guidance as expected. Pilots complied with the well clear recovery guidance to regain well clear in 359 of the 365 (98%) LoDWC occurrences. Five of the 6 non-compliance cases involved vertical resolution uploads being made instead of the recommended turn (possibly in anticipation of a subsequent vertical resolution advisory), and one was due to the pilot preferring a turn in the opposite direction. Compliance rates were identical between displays. Subjective ratings gathered from post-simulation questionnaires were also nearly equal between the Limited Suggestive and Directional displays (preferred by 60% and 40% pilots, respectively).

Loss of DWC events were slightly less severe when using the Limited Suggestive guidance, which presented a specific solution range to pilots at the onset of recovery bands. While the precise recommendations slightly decreased the time spent within the well clear threshold, differences in LoDWC severity were not significant. The Limited Suggestive and Directional well clear recovery guidance displays available in the present study reduced LoDWC severity by 78% and 64%, respectively, compared to the previous analysis of the DAA system without recovery guidance [16]. The recovery displays appear to be equally effective at aiding the pilot task of regaining DAA well clear against intruders at critical ranges, and both were referenced as viable guidance options for maximizing horizontal and/or vertical miss distance during a loss of DWC event in Phase 1 of the DAA MOPS [17]. In conclusion, multiple design concepts are acceptable for maintaining and regaining DWC when the guidance corresponds with the alerting logic.

References

1. RTCA: Terms of Reference: RTCA Special Committee 228 Minimum Performance Standards for Unmanned Aircraft Systems. RTCA Inc., Washington (2013)
2. Federal Aviation Administration (FAA): Integration of Civil UAS in the NAS Roadmap, 1st edn. FAA, Washington (2013)
3. Code of Federal Regulations: 14 CFR, Part 91, Sec. 91.113 (2004)
4. Santiago, C., Mueller, E.R.: Pilot evaluation of a UAS detect-and-avoid system's effectiveness in remaining well clear. In: 11th USA/Europe Air Traffic Management Research and Development Seminar, Lisbon (2015)
5. Bell, S., Drury, J., Estes, S., Reynolds, C.: GDTI: A ground station display of traffic information for use in sense and avoid operations. In: 2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC), Williamsburg, VA (2012)
6. Friedman-Berg, F., Rein, J., Racine, R.: Minimum visual information requirements for detect and avoid in unmanned aircraft systems. In: Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, Chicago, IL (2014)
7. Draper, D.H., Pack, J.S., Darrah, S.J., Moulton, S.N.: Human-machine interface development for common airborne sense and avoid program. In: Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, Chicago, IL (2014)

8. Fern, L., Rorie, R.C., Pack, J., Shively, R.J., Draper, M.: An evaluation of DAA displays for unmanned aircraft systems: the effect of information level and display location on pilot performance. In: Proceedings of 15th AIAA Aviation Technology, Integration, and Operations Conference, Dallas, TX (2015)
9. Rorie, R.C., Fern, L.: The impact of integrated maneuver guidance information on UAS pilots performing the detect and avoid task. In: Proceedings of the 59th Human Factors and Ergonomics Society Annual Meeting (2015)
10. Monk, K.J., Fern, L., Rorie, R.C., Shively, R.J.: Effects of display location and information level on UAS pilot assessments of a detect and avoid system. In: Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting, Los Angeles, CA (2015)
11. Rorie, R.C., Fern, L., Shively, J.: The impact of suggestive maneuver guidance on UAS pilot performing the detect and avoid function. In: AIAA Infotech@ Aerospace, San Diego, CA (2016)
12. Monk, K.J., Roberts, Z.: UAS pilot evaluations of suggestive guidance on detect-and-avoid displays. In: Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, Washington, DC (2016)
13. Feitshans, G.L., Rowe, A.J., Davis, J.E., Holland, M., Berger, L.: Vigilant spirit control station (VSCS)—'The Face of COUNTER'. In: Proceedings of AIAA Guidance, Navigation and Control Conference Exhibition, Honolulu, HI (2008)
14. Santiago, C., Abramson, M., Refai, M., Mueller, E., Johnson, M., Snow, J.: Java Architecture for Detect and Avoid (DAA) Modeling and Extensibility (JADEM) (2015)
15. Abramson, M., Mohamad, R., Confesor, S.: A generic resolution advisor and conflict evaluator (GRACE) in applications to detect-and-avoid (DAA) systems of unmanned aircraft. In: Proceedings of 17th AIAA Aviation Technology, Integration, and Operations Conference (2017)
16. Mueller, E., Santiago, C., Watza, S.: Piloted "Well Clear" performance evaluation of detect-and-avoid systems with suggestive guidance. NASA/TM-2016-219396 (2016)
17. RTCA: Minimum Operational Performance Standards (MOPS) for Unmanned Aircraft Systems (UAS) Detect and Avoid (DAA) systems. RTCA Inc., Washington (2017)

Evaluation of Early Ground Control Station Configurations for Interacting with a UAS Traffic Management (UTM) System

Arik-Quang V. Dao[1(✉)], Lynne Martin[1], Christoph Mohlenbrink[2], Nancy Bienert[2], Cynthia Wolter[2], Ashley Gomez[2], Lauren Claudatos[2], and Joey Mercer[1]

[1] NASA, Moffett Field, CA, USA
{Quang.V.Dao,Lynne.Martin,Joey.Mercer}@NASA.gov
[2] San Jose State University, San Jose, CA, USA
{Christoph.P.Mohlenbrink,Nancy.Bienert,Cynthia.Wolter,
Ashley.N.Gomez,Lauren.E.Claudatos}@NASA.gov

Abstract. The purpose of this paper is to report on a human factors evaluation of ground control station design concepts for interacting with an unmanned traffic management system. The data collected for this paper comes from recent field tests for NASA's Unmanned Aircraft System (UAS) Traffic Management (UTM) project, and covers the following topics; workload, situation awareness, as well as flight crew communication, coordination, and procedures. The goal of this evaluation was to determine if the various software implementations for interacting with the UTM system can be described and classified into design concepts to provide guidance for the development of future UTM interfaces. We begin with a brief description of NASA's UTM project, followed by a description of the test range configuration related to a second development phase. We identified (post hoc) two classes in which the ground control stations could be grouped. This grouping was based on level of display integration. The analysis was exploratory and informal. It was conducted to compare ground stations across those two classes and against the aforementioned topics. Overall, subjective ratings showed no differences with respect to workload and communication, but ratings for situation awareness and effectiveness of the procedures favored integration of displays.

Keywords: Unmanned traffic management · Human-systems integration · Systems engineering · Human factors

1 Introduction

Unmanned aircraft, commonly referred to as "drones", are aircraft designed to operate with the absence of an onboard pilot. These aircraft may be part of an unmanned aircraft system (UAS) that allows a pilot to manually control the vehicle remotely, or provide strategic guidance when the aircraft is autonomous. To name a few, UAS have applications in search and rescue, infrastructure inspections, goods delivery, recreation, and media. Due in part to the diversity of such applications and their relative affordability, the market for UAS is expected to grow significantly. The Federal Aviation

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_8

Administration (FAA) has projected that sales of UAS of all sizes and types in the United States will grow from 2.5 million in 2016 to 7 million units in 2020 [1]. If these aircraft are eventually deployed, the total number of concurrent UAS operations in the U.S. is expected to be up to 250,000 by 2035 - approximately 175,000 of those will be for commerce alone [2].

Early experience from manned aviation has made the case for an appropriate level of organization, if safety is to be achieved with increasingly congested air traffic [3]. This meant that the air traffic had to be managed. For manned air-traffic, this is done through a system where human controllers are responsible for maintaining separation between aircraft. This system has proven to be very safe. The disadvantage is that the capacity of the airspace will be limited by human cognitive resources - mainly workload. NASA studies, e.g. [4], conducted to examine solutions for meeting growing demand for air transportation services found that sustained high capacity is achievable if automation assisted controllers in separating aircraft. We expect these findings to remain relevant to UAS operations. However, a system for managing UAS operations will need to consider accommodating an unmanned aircraft fleet size that is forecast to be 35 times that of manned aircraft in 2020. Given these considerations, NASA's vision for UAS Traffic Management (UTM) will not mirror the traditional manned air traffic management system. Instead, UTM will be designed to allocate active management of aircraft from human controllers to automation. Humans in this system will then serve as supervisors providing strategic level input [3]. Beyond the U.S., the European aviation community has also recognized the importance of UTM, and acknowledged that rising demand for UAS operations could be addressed by UTM and related technologies. NASA's UTM concept of operations offers initial guidance for testing such technologies.

According to the NASA UTM concept of operations, the UTM will be designed to safely enable large-scale small UAS (i.e., unmanned aircraft less than 50 lb) operations in low altitude (i.e., below 500 feet) Class G airspace [3]. This will be achieved by providing technical capabilities to UAS operators and stakeholders. Technical capabilities will take the form of information products and services. These products and services will be tested as part of the UTM concept in NASA's research platform [5, 6]. The research platform supports research with both live and simulated aircraft. The simulation capabilities have enabled the testing of off-nominal interactions between virtual and live aircraft in field tests, and it will be the primary tool for feasibly evaluating large scale UAS operations as the tests grow in complexity. This increasing complexity, as well as scope, is reflected in NASA's plan for testing the UTM concept - distinguished by four technical capability levels (TCL). TCL 1 and 2 have already been tested.

In TCL 1, operations were conducted over remotely populated areas (e.g., rural operations) [7]. The traffic density was very low, with 4 aircraft available in early field tests and at most 2 concurrent operations; each in its own volume that reserved all airspace above and below. The aircraft remained within visual line of sight throughout the operation. The UTM services provided only vetting of operations against conflicts between operations and other constraints, such as existing airports; the information that it provided conveyed whether an operation was accepted into the system and if it was rejected, why it was rejected.

In TCL 2, the technical capabilities from TCL 1 were carried over and the concept was extended to the evaluation of industrial applications of UAS operations over sparsely populated areas. It included a mixture of visual line of sight and beyond visual line of sight operations. TCL 2 included other enhancements such as: alerting for airspace intrusion, alerts to contingency management, and segmented flight planning that allowed stratification of operational volumes among other things. TCL 3 and 4 have not been conducted yet. They will build on the capabilities of TCL 1 and 2 to include: operations over increasingly populated areas, moderate and then high UAS traffic densities, interactions between manned and unmanned operations, as well as large-scale contingency management. The reader is encouraged to see Johnson et al. [7] and Kopardekar et al. [3] for a more detailed discussion of NASA's concept of operation and test plans. The findings reported in this paper will focus on the most recent test - TCL 2.

In TCL 2 we advanced the concept by introducing the ability for operators to plan and schedule beyond visual line of sight (BVLOS) operations. To accomplish BVLOS, a number of capabilities had to be in place for the UAS operators. They needed displays that provided information about where their aircraft were relative to operational boundaries, and other air traffic. This was accomplished either on a few integrated displays or across separate displays. The configuration of these displays as part of the floor plan for the UAS ground stations influenced the size of the flight crews and their roles. In this paper, we describe how our human factors measures varied as a function of these ground station configurations, and offer some interpretation on how those differences reflected their effectiveness in the field tests. The analysis we presented here was informal and intended to be exploratory. In the next section we consider the field test configuration and the UTM architecture before discussing the ground control stations and the human factors measures used to evaluate them.

2 Test Range and Scenarios

TCL 2 flight tests were bound to uncontrolled airspace 2 miles north of the active runway at Reno-Stead Airport (RTS). The range was a flat, dry, desert basin surrounded by steep mountains. UAS flight crews were positioned between 4 of 5 total ground control station (GCS) locations in the area (Fig. 1); simulated aircraft were injected into the test from GCS 1 when the test scenarios required it. The flight test director (FTD) coordinated flights on the field, and was located immediately east of GCS 3 on the south end of the test range.

Four different scenarios, each with a different back story, prompting different combinations of events, were performed. Scenarios shared at least one event (Table 1) and were designed to represent interactions likely to occur if UTM was to be implemented in the future. These interactions included those that involved intruder aircraft and contingency management operations. Scenarios were 30 min long with up to 5 concurrent flights over the test range. Actual flight durations were between 6 and 23 min. The flights took place over the course of 9 days, and daily between 8:30 am to 12:30 pm to take advantage of favorable weather and wind conditions.

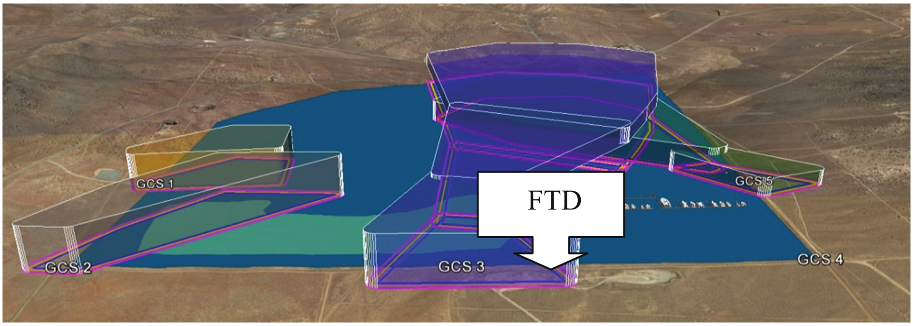


Fig. 1. UAS test range north of Reno-Stead Airport. Three-dimensional extrusions above the map represent operational boundaries for each of the ground control stations.

Table 1. Flight test scenarios

	Scenario 1 Agriculture	Scenario 2 Lost Hiker	Scenario 3 Ocean	Scenario 4 Earthquake
BVLOS	X	X	X	X
Multiple BVLOS	X		X	
Altitude stratified VLOS	X	X		X
Altitude stratified BVLOS			X	
Intruder aircraft tracking	X		X	
Intruder aircraft conflict alert	X		X	
Rogue aircraft conflict alerts	X			
Dynamic re-routing		X		X
Contingency management alerts			X	X
Public safety operation		X		
Simulated aircraft		X		X

3 Flight Crew Roles and Responsibilities

A total of eleven flight crews participated over the duration of the TCL 2 flight tests. Two of those were NASA crews. The remaining 9 came from different industry partners. Crew-members were composed of individuals who operated together regularly. The size of regular crews varied between 2 and 5 members, where the most common crew size was four. Larger size crews of 5 were able to assign just one role to each member. Individuals in smaller crews served more than one role. The available roles were; pilot-in-control (PIC), client operator, ground control station operator (GCSO), on-site software engineer, and launch technician (Table 2). On-site software engineers were not required, but were present in some flight crews. Launch technicians were a practical requirement of the type of aircraft platform, i.e., fixed-wing aircraft. Auxiliary members expanded the regular crew to provide support specific to the TCL 2 flight test, but on occasion assist in pre-flight and post-flight procedures. These roles were; UTM representative, visual observer, observer controller, and human factors observer. Table 3 shows how the aforementioned roles were distributed across crew members.

Table 2. An exhaustive list of TCL 2 flight crew roles.

Regular roles	Auxiliary roles
(a) Pilot-in-control (PIC)	(f) UTM representative
(b) Client operator	(g) Visual observer
(c) Ground control station operator (GCSO)	(h) Observer controller
(d) On-site software engineer	(i) Human factors observer
(e) Launch technician [fixed-wing only]	

Table 3. Distribution of roles across flight crews.

	FC1	FC2	FC3	FC4	FC5	FC6	FC7	FC8	FC9	FC10	FC11
1	a	ac	a	a	a	a	abc	a	a	a	ac
2	b	e	c	b	c	c	c	bc	bc	bcd	b
3	h	bf	bf	c	bd	bf	f	f	f	e	f
4	g	g	g	f	f	d	g	g	d	f	e
5	i	h	h	g	g	g	h	d	d	g	e
6		i	i	h	h	h	i	h	g	h	d
7				i	i	i		i	h	i	g
8									i		h
9											i

*Columns headers are for flight crew. Rows represent individual members. The letters within cells are assigned roles as defined in Table 2. Some members supported more than one role

The PIC was responsible for the operation of the aircraft, and had ultimate authority over the operation of the aircraft. For two flight crews, the PIC doubled as a GCSO, where the aircraft flew autonomously on a pre-defined flight plan. The client operator created, and interfaced with the UTM system to submit flight geometries, as well as send and receive notifications. Examples of the geometries submitted by client operators are illustrated in Fig. 1 as 3-dimensional extrusions above the map. The GCSOs coordinated with the PIC to run pre-flight and post-flight checklists; they generated the flight plans and monitored the aircraft from their GCS display during the operations. In some cases, on-site software engineers also served as GCSOs and client operators on top of providing expertise in the troubleshooting and tweaking of proprietary clients. Launch technicians assembled and directed aircraft launch hardware for catapult fixed-wing platforms, and assisted with retrieval of aircraft upon their return and landing.

UTM representatives doubled as client operators by relaying UTM system information to the flight crew as necessary; their permanent role was to coordinate with the flight test group for adherence to research protocols. The visual observer visually tracked the aircraft to assist with separation between other aircraft, fauna, and terrain. The observer controller maintained radio contact with the test site authorities and the flight test director to coordinate and acquire authorization for take-off at their specific location. In most cases human factors observers watched passively and interacted with the flight crew for data collection purposes, but in a few instances they also provided assistance to the flight crew for miscellaneous tasks, such as equipment setup and packing.

4 Ground Control Station Configurations

The crews operated a mixture of fixed-wing and rotor-wing UAVs. Aircraft flew autonomously on routes defined through a mission planner or manually for take-off and landing by the pilot-in-control. UAS ground stations interacted with the UTM system via a client. All of the clients deployed by the flight crews were in their early development states. At minimum, the clients were required to be able to submit a flight volume to the UTM system and then receive and display messages from the system. For NASA flight crews, the UTM representative submitted volumes on their behalf and verbally relayed the messages. Mission volumes (Fig. 1) were sometimes submitted separately in the client and then redrawn on a moving map display to monitor their aircraft conformance to the boundaries. For some flight crews this was done by a single individual on a single display or by two individuals across multiple displays.

For flight crews where a single individual managed both the client and the ground control station, interacting with the UTM system through the client became an integrated part of their pre-flight and post-flight procedures. Interactions with UTM system provided GCSOs with information about the success or failure of their volume submission; if the volume was rejected, they had access to notifications that explained why, and were able to make adjustments to their missions accordingly. When the client operator and GCSO roles were assigned to separate individuals, volume submissions and mission planning were not integrated and additional effort had to be made to coordinate the operations between the two. For example, a client operator may need to wait for available gaps in the existing GCSO regular procedures to deliver important UTM system notifications, which resulted in some take-off delays. It is based on this notion of integration between the client operations and GCSO that we applied a post hoc grouping of the ground control station configurations. A flight crew operated an integrated ground control station if a single individual occupied both the client operator and ground control station roles. In Table 1, these were flight crews with the letters “b” and “c” located within the same cell - flight crews FC7 to FC10. All other flight crews carried non-integrated ground control stations. We conducted our analysis according to this classification.

5 Procedure

Each flight crew attended the flight test for three days. The first day consisted of a briefing and time for the crews to set up and test their equipment. The second two days were flight days when the crews flew warm-up flights and then a selection of the four scenarios that are described above. The warm-up flights served to verify connectivity between the aircraft and the ground station, and the client to the UTM system. Generally, crews flew two warm-up flights and two scenarios per flight day.

As part of the flight tests, a five-person human factors team collected data from the participants about their experiences flying in one location with multiple partners. They collected qualitative data from each group in the flight test as one researcher observed each crew. Data were collected in a number of ways, through observations of the participants during flights, brief questionnaires at the end of each scenario, and end of day group interviews. All these methods focused on five specific topics: flight crew workload during different phases of the flight, flight crew situation awareness, flight crew interaction, usage of the UTM clients and usability of the UTM system information. Participants were asked to rate their workload and situation awareness after each flight on a rating scale from 1 to 7 (very low to very high). Time permitting, they were also asked to discuss the flight they had just made. End of day interviews were focus group sessions, where flight crews discussed topics related to UTM. UTM reps also took part in a separate session and discussed the same topics. We focus our paper on the results generated from the subjective ratings collected for workload, situation awareness, flight crew communication, and how well flight crews thought the overall procedures, including the client operations, were integrated with those associated with the operation of the GCS. Analysis of the remaining human factors data are still underway and will be reported in a separate paper.

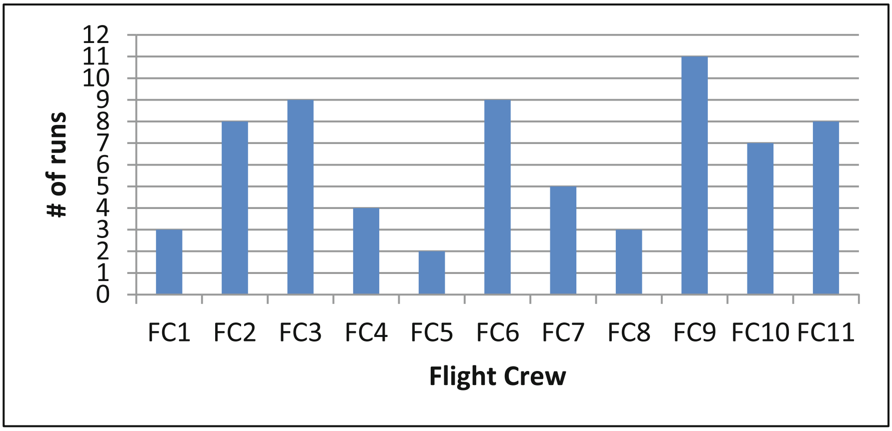


Fig. 2. Distribution of flights across flight crew.

The 7 point Likert format rating scales were administered to GCSOs and PICs between each run. A run was defined by the start and stop of single test scenario; this was announced by the flight test director. Across 9 days, these ratings were collected for a total of 69 runs. These runs were not distributed evenly across flight crews due to schedule availability, unforeseen circumstances that made it unsafe to fly, or equipment failure. The distribution of runs is shown in Fig. 2. Between integrated and non-integrated ground control configurations there were 26 and 43 runs respectively. Unfortunately, although interview and observation data was acquired for Flight Crew 4 (FC4), we were unable to acquire workload ratings from them for various reasons; consequently, the results reported here do not include input from FC4. Overall, the collection of ratings from the flight crews was inconsistent at best-so the total number of ratings does correspond to the number of runs.

Due to the safety risks associated with interacting with flight crews during live flight tests, human factors observers were instructed to administer data collection instruments only when safe to do so and at the discretion of the flight crews. In some instances, human factors observers were not able to collect data. The analysis we show here is informal. The results we present in this paper serve only to assist with organizing and describing various flight crew configurations for informing future studies, and to highlight some potential display design challenges. We reserve formal investigations for future in-lab simulations where experimental control can be exercised.

6 Results

Sixteen workload ratings were collected for integrated GCSs versus 24 for non-integrated GCSs. Figure 3 shows that mean rated workload between the two crew-types was equally moderate. Average rated workload for integrated GCSs was 3.9 (*SD* = 0.7) and 3.6 for non-integrated (*SD* = 1.5).

Overall quality of communication was reported to be very good for integrated and non-integrated GCSs (Fig. 4), where the mean for integrated crews was 6.7 (*SD* = 0.5) and 6.6 for non-integrated (*SD* = 8). Lack of availability for ratings from the flight crew and hardware failure resulted in some missing data here. There was one missing data point for the integrated group (*N* = 15) and 7 missing data points for the non-integrated group (*N* = 17).

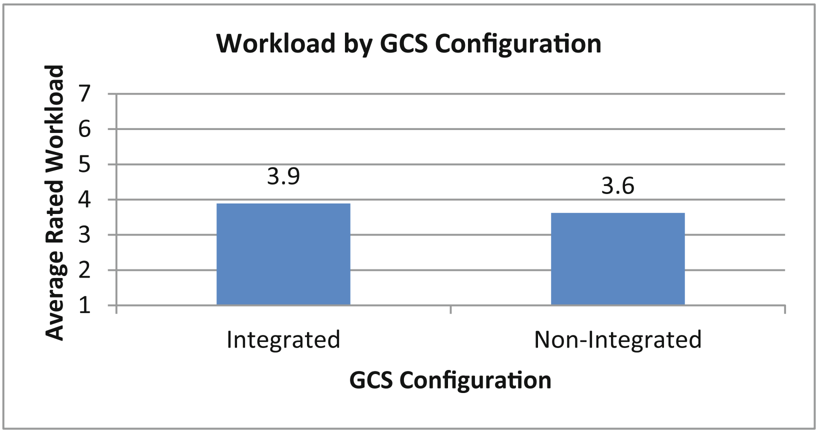


Fig. 3. Average workload rating by GCS configurations.

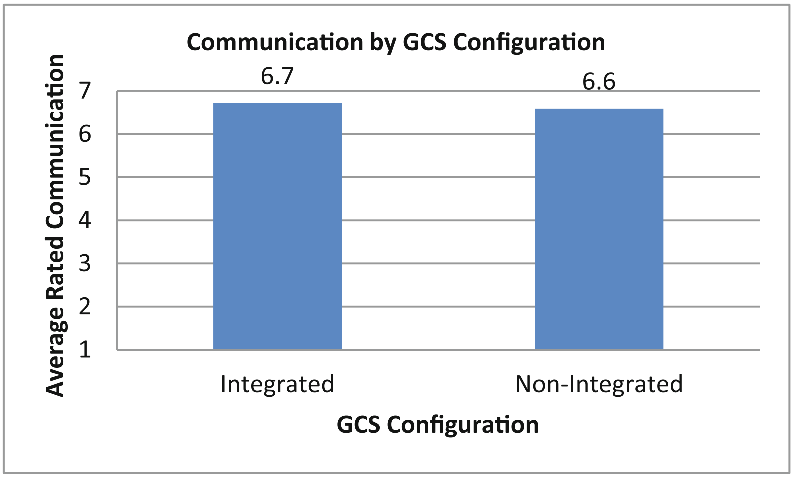


Fig. 4. Average communication rating by GCS configuration.

Consistent with overall quality of communication, on average the GCSOs indicated that there was little difficulty in relaying UTM information to the rest of the flight crew (Fig. 5). This was the same for both integrated ($M = 2.3, SD = 2.4$) and non-integrated ($M = 2.4, SD = 2.2$). Again, lack of availability for ratings resulted in missing data with the integrated group ($N = 13$) and software issues inhibiting launch in the non-integrated group reduced the number of ratings significantly ($N = 9.0$).

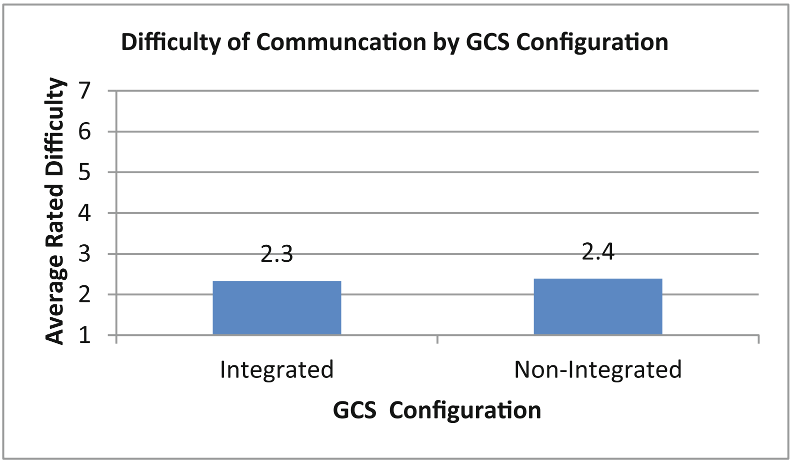


Fig. 5. Average difficulty of communication rating by GCS configuration.

We see the greatest differences between integrated and non-integrated crews with situation awareness (Fig. 6). In the integrated group GCSOs felt that their situation awareness was good ($M = 5.1, SD = 2.1$). In contrast, GCSOs in the non-integrated group believed their situation awareness was mediocre ($M = 3.9, SD = 2.0$). Fourteen ratings were acquired for integrated versus 18 for non-integrated crews. It is conceivable here, that with integrated displays, it was easier to maintain situation awareness because GCSOs did not need to divert attention from their map displays to view UTM system notifications.

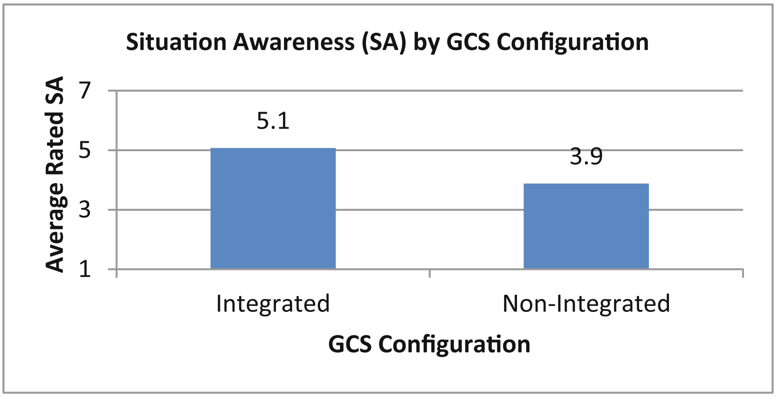


Fig. 6. Average situation awareness by GCS configuration.

Figure 7 shows ratings for how well GCSOs believed overall procedures were integrated with procedures for the ground station. The average rating for this dimension in the integrated group is 6.2 (*SD* = 1.5) and 5.4 (*SD* = 1.3) for the non-integrated group. The less positive view is not surprising. If a single operator was considering information from both the client and ground control station, it is reasonable that procedures for both the displays fall in line with a single set of procedures - even if the client and ground control were on separate, but proximal displays. Some additional reconciling of procedures would be expected if the client and ground control station operations were handled

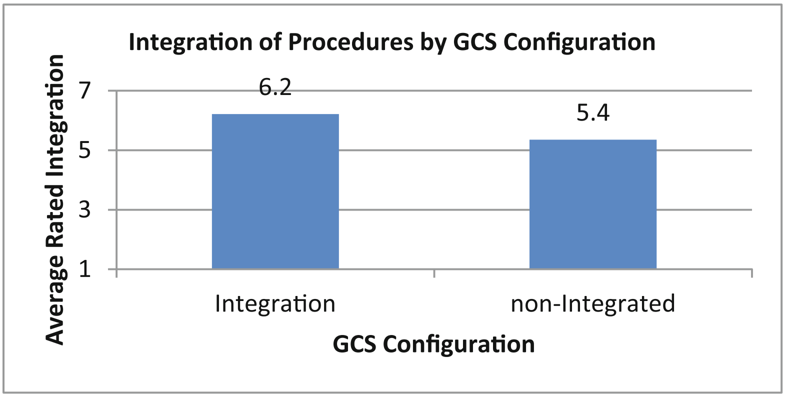


Fig. 7. Average rated integration of procedures by GCS configuration.

between two separate individuals. For this set of ratings we had 15 ratings for integrated and 18 for non-integrated crews.

7 Conclusion

In this paper we offered a brief introduction to the UTM project and activities associated with the development of the UTM concept. In our results, we expressed that due to limitations in data collection opportunities and the unpredictable circumstances of a live flight test environment a considerable amount of data was missing from our explorative analysis of the UTM GCS configurations. However, TCL 2 offered an opportunity to make human factors contributions to the UTM development effort. We were able to field and evaluate early versions of the UTM ground control stations, as well as vet equally early versions of our data collection instruments and procedures. Collectively the data will inform human factors efforts in future UTM tests.

Overall, the subjective ratings revealed that with respect to workload and communication, there was little-to-no difference between the integrated and non-integrate GCS configurations. GCSOs felt workload was moderate in most circumstances. For communication, both GCS configurations were rated equally good. We speculate from field observations that when client and GCSO roles were not integrated, UTM representatives were able to compensate for any issues that germinated from that lack of integration.

Although the UTM representatives were an artifact of the flight test and did not operate regularly with the flight crew, the flight crews regarded them as essential members of the team. In almost all cases, because they had an established rapport with the UTM representatives, flight crews requested that the representatives initially assigned to them in shakedowns remain with them throughout the field tests. As we consider development of the displays and applications for UTM, it may be useful here to more closely evaluate the role of the UTM representative and how aspects of this role can be incorporated into automation used to assist interaction with the UTM system. It may also be useful, when the concept is extended in later TCLs, to evaluate the role of operators who will control large fleets - where the economy of employing a single operator will be pertinent and issues such as task switching will be a key concern to public safety and industrial organizations.

Ratings collected for situation awareness and integration of procedures seem to favor integrating the displays and combined roles for client operator and GCSO. This seems reasonable for commercial applications mentioned above, and for recreational activities where it would be economically impractical to require a multi-person crew to operate an aircraft. The principle challenge will be to identify what information the UTM system can provide for such displays without cluttering or intruding on central mission planning features.

As the development of the UTM system moves forward, the human factors effort will need to pay particular attention to assembling information requirements that will be sensitive to the privacy concerns of people and industry, but at the same time facilitate the sharing of an appropriate amount information to support safe and effective use of the UTM airspace.

Acknowledgements. We extend our gratitude to all of the flight crews who participated in TCL 2 for their patience and support of the UTM human factors effort.

References

1. FAA: FAA Aerospace Forecasts. https://www.faa.gov/data_research/aviation/aerospace_forecasts/
2. Unmanned Aircraft Systems (UAS) Service Demand 2015–2035 Literature Review & Projections of Future Usage, Cambridge, MA (2013)
3. Kopardekar, P., Rios, J., Prevot, T., Johnson, M., Jung, J., Robinson, J.E.I.: Unmanned aircraft system traffic management (UTM) concept of operations. Am. Inst. Aeronaut. Astronaut. (2016)
4. Prevot, T., Mercer, J., Martin, L., Homola, J., Cabrall, C.D., Brasil, C.L.: Evaluation of high density air traffic operations with automation for separation assurance, weather avoidance and schedule conformance. In: Proceedings of the 11th AIAA Aviation Technology, Integration, and Operation (ATIO) Including AIA (2011)
5. Prevot, T., Homola, J., Mercer, J.: From rural to urban environments: human/systems simulation research for low altitude UAS Traffic Management (UTM). In: Proceeding of the 16th AIAA Aviation Technology, Integration, and Operations Conference Washington, DC, 13–17 June, pp. 1–13 (2016)
6. Homola, J., Prevot, T., Mercer, J., Bienert, N., Gabriel, C.: UAS traffic management (UTM) simulation capabilities and laboratory environment, pp. 1–7 (2016)
7. Johnson, M., Jung, J., Rios, J., Mercer, J., Homola, J., Prevot, T., Mulfinger, D., Kopardekar, P.: Flight test evaluation of a traffic management concept for unmanned aircraft systems in a rural environment. In: Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM 2017) (2017)

User-Centered Design and Evaluation of Smartphone-Controls for the Movement Control of Mobile Platforms

Robert Fingerle, Lars Krause, Hugo Tausch, and Carsten Wittenberg(✉)

Faculty of Mechanics and Electronics, Heilbronn University, Max-Planck-Str, 39,
74081 Heilbronn, Germany
carsten.wittenberg@hs-heilbronn.de

Abstract. Mobile platforms like mobile robots are becoming more relevant in different domains such as production industry or rescue services. Although these platforms are intended to work autonomously, it is still necessary to interact with them. This research project focuses on the input control for steering mobile robots platform via a smartphone or smart device. Three different types of controls (Virtual Joystick, 4 key directional pad and 8 key directional pad) were developed and evaluated by almost 40 participants in two evaluation rounds. Results indicate that the joystick control was rated as the best input control.

Keywords: User analysis · Human factors · Human-robot interaction · Mobile devices · Smartphone controls · User-centered evaluation

1 Introduction and Preliminary Studies

In the future, mobile platforms like mobile robots will play a major role. The application of mobile robots covers many industries such as logistics and agricultural industry. The Faculty of Mechanics and Electronics at Heilbronn University in Germany acquired different mobile platforms and participates in various robotics competitions.

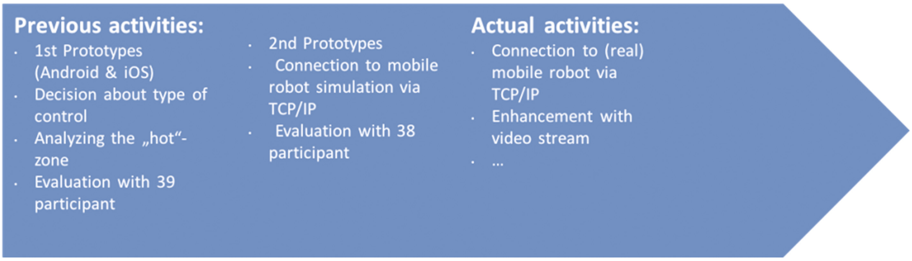
Different research activities focus on the control of the movement of mobile platforms. In one project the inclination sensor of a Wii controller was used for controlling the moving direction and the speed of the mobile platform. Given that the Wii-Controller and the control unit on the platform did not perform as intended. In a second step the Wii-Controller was replaced by a smartphone using a WLAN connection for the communication between the smartphone and the control unit of the mobile platform [1]. The inclination sensor was used as the interaction device for controlling the movement of the mobile platform. This solution had the disadvantage that the smartphone display is not completely visible during the human-platform interaction.

But for a variety of uses additional information are useful during the human-platform interaction. As an example, a video stream taped by the mobile platform can be necessary for appreciating the actual situation. As a consequence, the actual project was started to define controls for the smartphone touch-display for controlling the platform.

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_9

2 Project Plan

Three different goals were defined in the project (Fig. 1). The first goal of the project was to work out the optimal input control for the movement or cruise control of mobile platforms – based on a user-centered evaluation. For this evaluation, similar prototypes for iOS and Android were developed. Beside the decision about the input control also the optimal position on the smartphone screen (the “hot” zone) were determined.



Previous activities:

- 1st Prototypes (Android & iOS)
- Decision about type of control
- Analyzing the „hot“-zone
- Evaluation with 39 participant

- 2nd Prototypes
- Connection to mobile robot simulation via TCP/IP
- Evaluation with 38 participant

Actual activities:

- Connection to (real) mobile robot via TCP/IP
- Enhancement with video stream
- ...

Fig. 1. Project plan

The prototypes for the second phases (once more developed for iOS and Android) were used for the evaluation of the usability of the controls combined with a mobile platform simulation. The prototypes were compliant to connect to a mobile platform or a simulation via a TCP/IP functionality.

The current activities focus the integration of the smartphone control into the human-robot system and the extension of streaming video functionality.

3 First Evaluation for the Preferred Input Control

As mentioned before the main goal of this project was to define an ideal input control for a smartphone application to steer a mobile platform like a mobile robot. Based on a similar study in a not-industrial context [2] in a first step three possible control types were designed:

1. (Virtual) Joystick (Fig. 2)

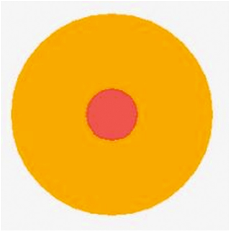


Fig. 2. Developed input control: (Virtual) Joystick

2. 4 key directional pad (Fig. 3)

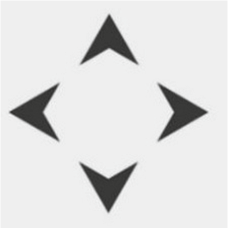


Fig. 3. Developed input control: 4 key directional pad

3. 8 key directional pad (Fig. 4)

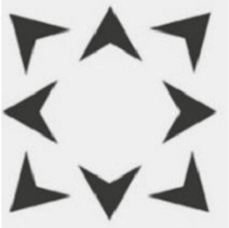


Fig. 4. Developed input control: 8 key directional pad

The input controls were integrated in two prototypes (one for each operating system). In the lower part of the touchscreen the input controls were located (the participants chose their optimal position, Chapter 4). In the upper part of the touchscreen a blue point was displayed (Fig. 5). The participants should move this point via the input control. The prototypes were shown to 39 participants (20 for Android, 19 for iOS, mainly technical students). These first prototypes had no connection either to a robot than to a robot simulation. The context of the use of these controls was verbally and textually explained to the participants. The participants had to rate the following attributes of the controls:

1. Sense while interaction
2. Ease of Use
3. Joy of Use
4. Understandability
5. Meeting the user's expectations

The following figures shows the rating of the participants (1 = very bad, 5: very good). First the participants were asked if they feel comfortable with the different controls (Fig. 6). The rating of the directional key pads is quite central; the rating of the joystick is more positive. A possible interpretation of the rating of the directional keypads is that these controls allow only a discrete movement of the robot.

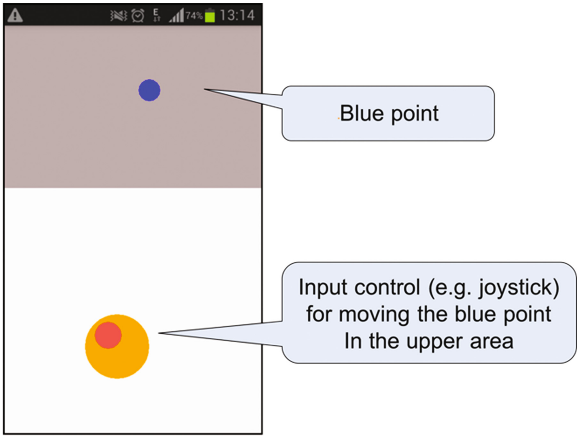


Fig. 5. Screen layout of the first prototype (example with joystick)

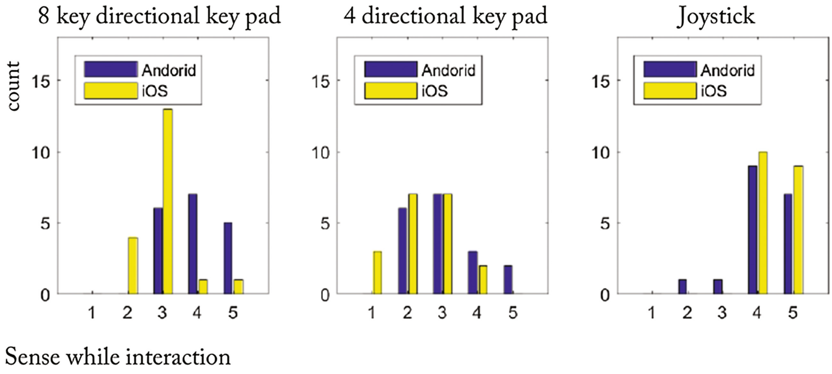


Fig. 6. Rating about the sense of the participants while interacting with the input controls

The second question focused the Ease of Use of the controls (Fig. 7). The results are similar to the previous question. The third question was aimed to the Joy of Use (Fig. 8). It seems that the "fun factor" of all three controls is not well-marked. The question about the understandability shows good results for all three controls (Fig. 9). The participants rated the understandability of the joystick a little better.

The last question of this set of issues was supposed to find out if the user expectations about the controls were fulfilled (Fig. 10). The joystick was rated better than the directional key pads. The participants gave as reasons among others that the directional key pads allow no increments e.g. of speed and course control.

In summary, it can be stated that the joystick is the preferred input control for the cruise control of mobile platforms like mobile robot. Main point of criticism regarding the directional key pads was the limitation based on the discrete positions of the key pads. These keys allow only binary states (e.g. no speed vs. full speed).

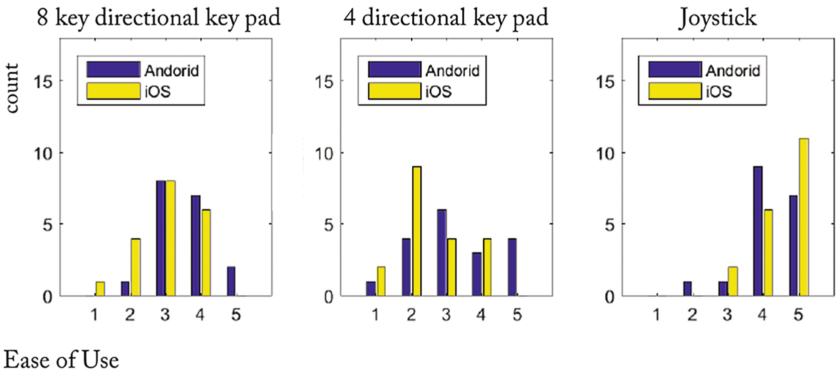


Fig. 7. Rating about Ease of Use

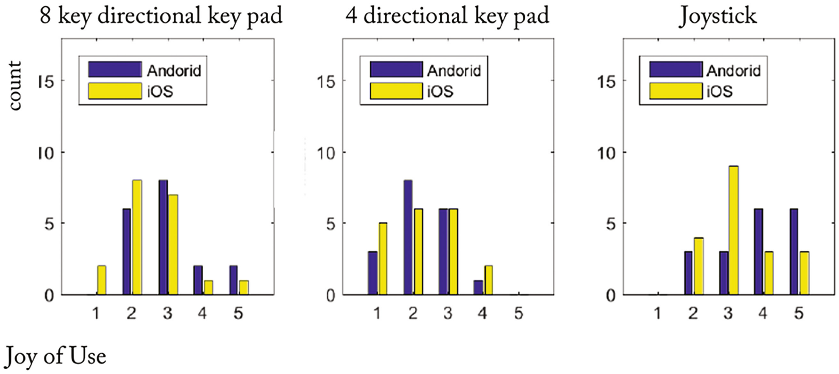


Fig. 8. Rating about Joy of Use

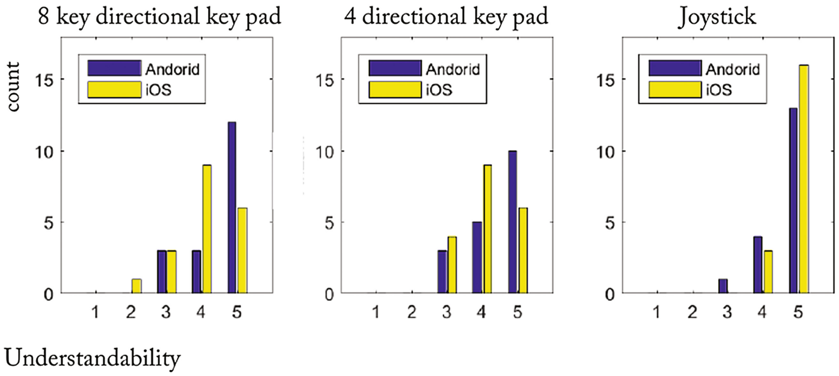


Fig. 9. Rating about understandability of the controls

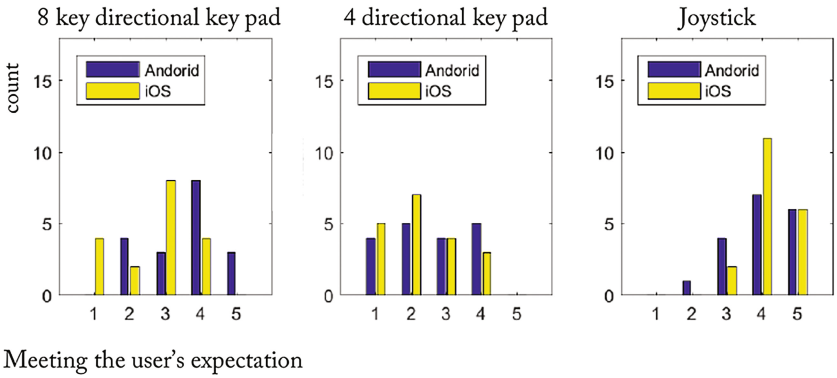


Fig. 10. Rating if the controls meet the user's expectations

4 Preferred Position of the Selected Control

The rough position of the joystick is determined by the reachability with the thumb without covering the upper area on the smartphone touchscreen. This upper area is foreseen for further functions like the live video streaming or grapping things with the robot (Fig. 11). This grapping function may be also video based.

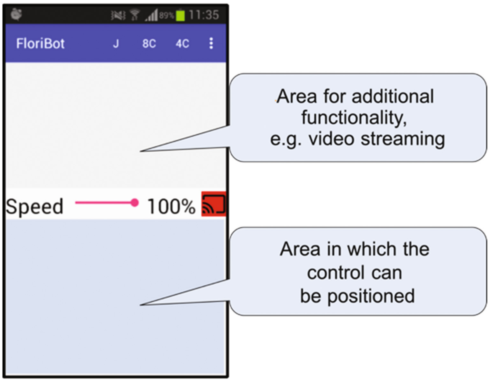

11:35
FloriBot
J
8C
4C
Area for additional functionality, e.g. video streaming
Speed
100%
Area in which the control can be positioned

Fig. 11. Planned screen layout

The evaluation was performed on two different smartphone types: Samsung S3 mini for the Android prototype and iPhone 6 s for the iOS prototype. The smartphones were chosen because of their availability. Figure 12 shows an illustration. The following figures shows the preferred control positions (center of the control) for both smartphones (Figs. 13 and 14). An advantage of using two different platforms with different sizes is also to get an idea about the influence of the size of the touchscreen and the smartphone. The results show no prominent differences between the Android and iOS prototypes except a broader distribution at the Android prototype (Samsung S3 mini).

Further research work will be done focusing this point. Heilbronn University has ordered different smartphones with different sizes and will continue research about this point.

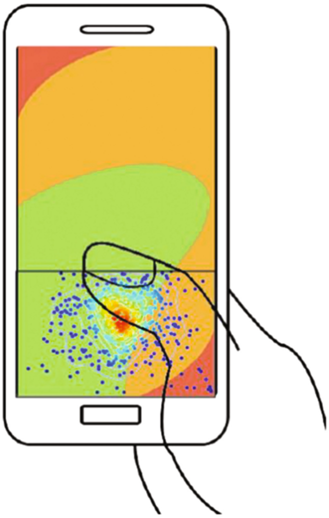


Fig. 12. Illustration of the survey about the central points

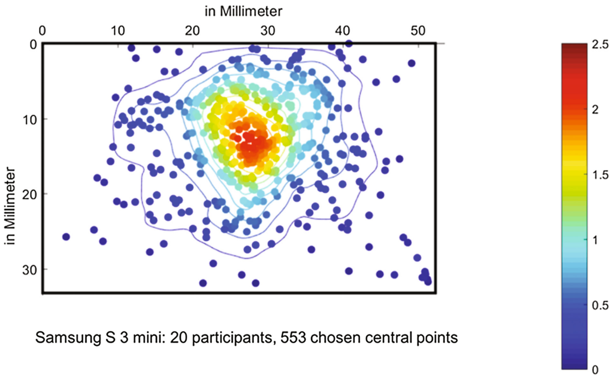


Fig. 13. Results of the preferred central points, Samsung S3 mini

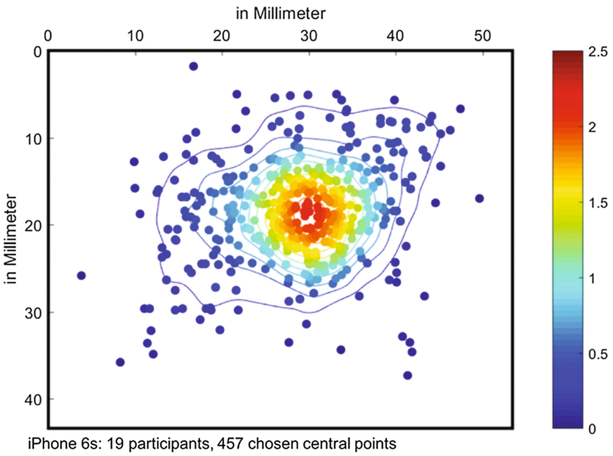


Fig. 14. Results of the preferred central points, iPhone 6 s

5 Revised Prototypes and Re-evaluation

Based on the results of the first step the applications were enhanced for the cruise control of the mobile platform. Therefor the applications got TCP/IP functionality for the communication either with a “real” mobile robot or with a robot simulation.

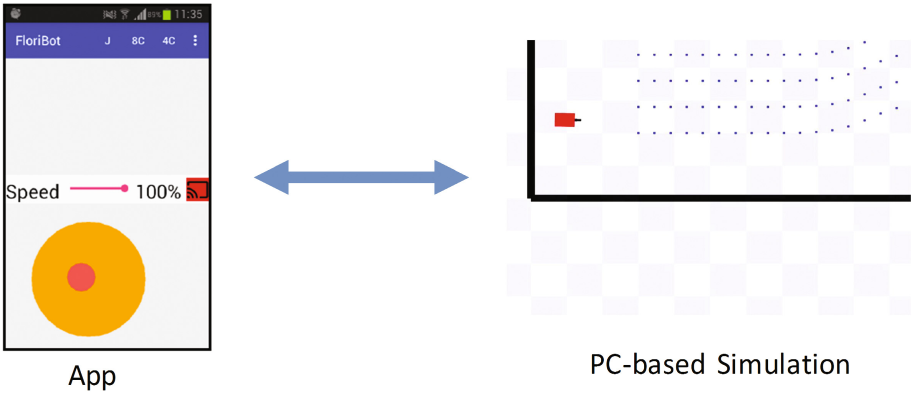


Fig. 15. Application (left, Android prototype) and PC-based simulation (right)

Reasoned in the independence from the mobile platform this evaluation was performed with a PC-based simulation. The participants of this study hat to move the virtual mobile platform (the red rectangle) through the dotted lines. At the end of each path the platform had to be turned around and steered into the next path (Fig. 15). This task is quite more complex than moving the blue point in the first part.

Once again all three input controls (Virtual Joystick (Fig. 2), 4 key directional pad (Fig. 3), 8 key directional pad (Fig. 4) were presented to the 38 participants of this part of the study.

The participants had to answer the following questions:

1. Difficulty to steer the robot
2. Steering the robot without seeing the control
3. Sense while interaction
4. Ease of Use
5. Joy of Use
6. Understandability
7. Meeting the user's expectations

The questions 1 and 2 are new, the other questions are similar to the first evaluation.

The Figs. 16, 17, 18, 19, 20, 21, 22 shows the result of the evaluation with the applications connected to the simulation.

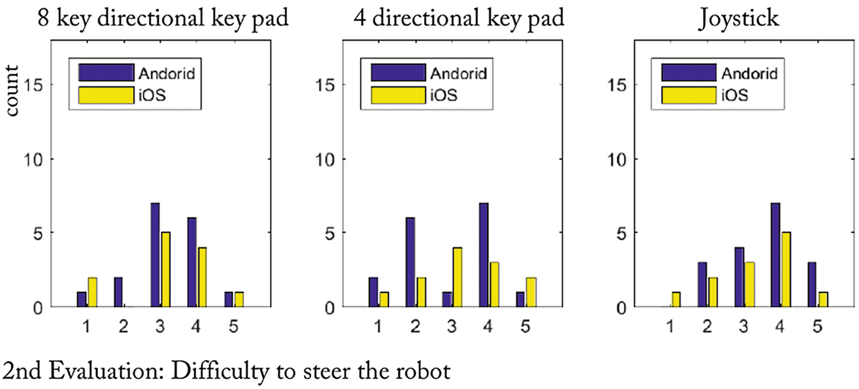


Fig. 16. Results: Subjective difficulty to steer the simulated robot

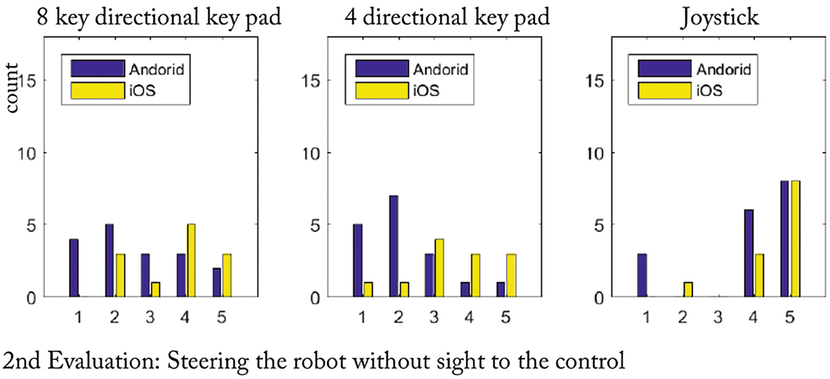


Fig. 17. Results: Feedback about steering the simulated robot without sight to the control

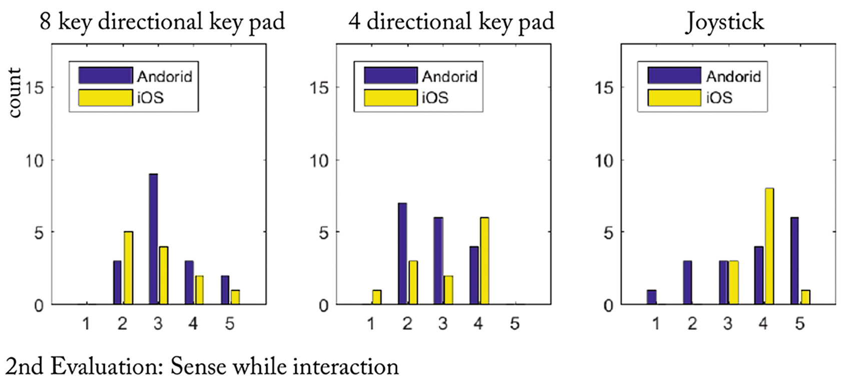


Fig. 18. Rating about the sense of the participants while interacting with the input controls in the second evaluation

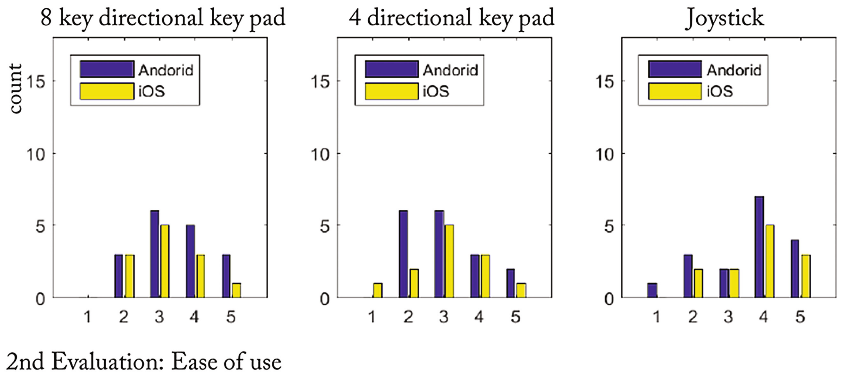


Fig. 19. Rating about the ease of use in the second evaluation

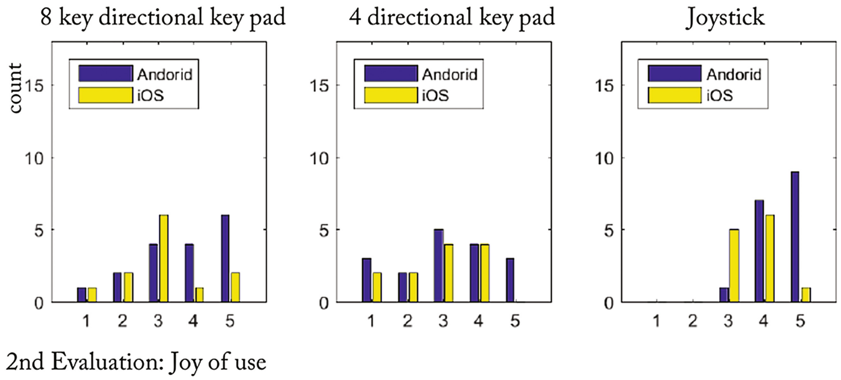


Fig. 20. Rating about the joy of use in the second evaluation

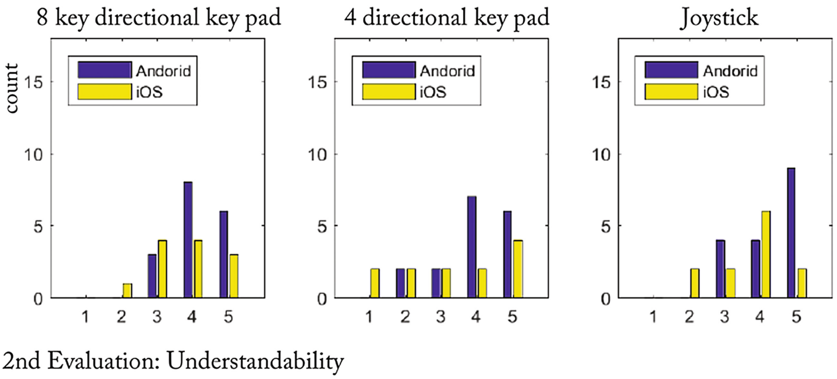


Fig. 21. Rating about the understandability of the interaction in the second evaluation

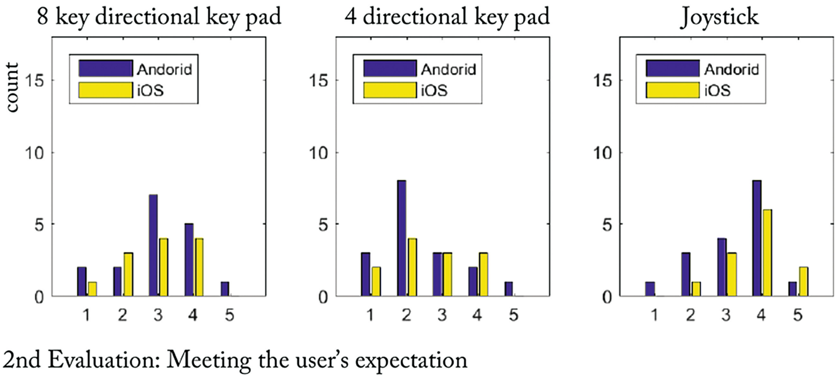


Fig. 22. Rating if the controls meet the user's expectations in the second evaluation

The participants feedback regarding steering the robot through the lines shows again an advantage of the joystick (Fig. 16). This impression was strengthened by the rating of steering the robot without taken a look to the touchscreen with the control – the participants watched only the simulation at the PC screen. The virtual joystick won also this round (Fig. 17).

The rating of the other questions relativizes the results of the first evaluation. The virtual joystick still won also the second round but the advance is shrunken (Figs. 18, 19, 20, 21, 22).

In summary, it can be said that the virtual joystick was rated in two evaluation rounds as the best interaction control for steering a mobile robot via a smartphone. The second evaluation with a steering task came closer to the reality than the first round with only the "blue point" – task.

It is important to mentioned that both evaluations focused only the subjective impressions of the participants. Objective items like time to finish a task or errors during the tasks were not recorded and used for analysis. This could be the content of further research activities.

6 Current Activities and Next Steps

The virtual joystick will be the selected input control for steering the mobile robots. The developed application will be enhanced focusing the following key points:

- Connection to the real mobile platform and testing it with the real mobile robot
- Integrating a video stream from the mobile robot (in the upper area of the application)
- Analyzing the user acceptance of different movement modes (e.g. crab-mode vs. circular motion) and its realization in the application

References

1. Rixen, M.-L., Buyer, S., Heverhagen, T., Wittenberg, C.: Mobile Nutzerschnittstellen in der Automatisierungstechnik (mobile user interfaces in automation). In: Proceedings AALE 2015—Automatisierung im Fokus von Industrie 4.0, Deutscher Industrieverlag Munich, pp. 115–123 (2015)
2. Lai, Y.R., Hwang, T.K.P.: Virtual touchpad for cursor control of touchscreen thumb operation in the mobile context. In: Marcus, A. (ed.) Design, User Experience, and Usability—DUXU 2015, Part II. Lecture Notes on Computer Science, vol. 9187, pp. 563–574. Springer, Cham (2015)

A Multi-modal Interface for Natural Operator Teaming with Autonomous Robots (MINOTAUR)

Stephanie Kane, Kevin McGurgan(✉), Martin Voshell, Camille Monnier,
Stan German, and Andrey Ost

Charles River Analytics Inc, 625 Mt. Auburn Street, Cambridge, MA 02138, USA
kmcgurgan@cra.com

Abstract. Dismounted squads face logistical problems, including the management of physical burdens in complex operating environments. Autonomous unmanned ground vehicles (UGVs) can help transport equipment and supplies, but require active remote control or teleoperation, even for mundane tasks such as long-distance travel. This requires heads down attention, causing fatigue and reducing situational awareness. To address these needs, we designed and prototyped a Multi-modal Interface for Natural Operator Teaming with Autonomous Robots (MINOTAUR). The MINOTAUR human-robot interface (HRI) provides observability and directability of UGV behavior through a multi-modal interface that leverages gesture input, touch/physical input through a watch-based operator control unit (OCU), and voice input. MINOTAUR's multi-modal approach enables operators to leverage the strengths of each modality, while the OCU enables quick control inputs through lightweight interactions and at-a-glance information status summaries. This paper describes the requirements and use case analysis that informed MINOTAUR designs and provides detailed descriptions of design concepts.

Keywords: Human factors · Human-robot interaction · Watch-based interface · Multi-modal interface

1 Introduction

Dismounted squads often face logistical problems, such as the management of physical burdens in complex operating environments. Autonomous unmanned ground vehicles (UGVs) can help transport more equipment and supplies than can be carried by hand or in backpacks. However, these platforms often require active remote control or teleoperation, even for mundane tasks such as long-distance travel. This requires heads down attention from operators, which causes fatigue and reduces situational awareness, making it difficult to maneuver nimbly or watch out for threats. Poorly designed human-robot interfaces (HRIs), which integrate new autonomous capabilities at the expense of good HRI design, further limit the operational benefits of current systems. As a result, users require extensive training for interfaces that do not directly address their needs and only allow them to use a fraction of the available operational capabilities. A successful system should enable UGVs to reliably and autonomously follow a

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_10

dismounted operator and free the Warfighter from control tasks, improving situational awareness and reducing cognitive burden.

To address these needs of dismounted squads, we designed and prototyped a *Multi-modal Interface for Natural Operator Teaming with Autonomous Robots* (MINOTAUR) HRI. The MINOTAUR HRI consists of a UGV (including hardware and software) and a lightweight, wearable operator control unit (OCU), similar in form factor to a wristwatch. MINOTAUR interface designs were informed by a requirements analysis, which identified a broad set of operationally relevant use cases, such as a lead/follow arrangement, changing operational environments, and a range of UGV health and status problems.

MINOTAUR provides observability and directability of UGV behavior through a multi-modal interface that leverages gesture, touch/physical input through a watch-based OCU, and voice input. This approach enables operators to flexibly and opportunistically choose operationally appropriate input modalities and to provide redundant commands across modalities (e.g., a "stop" command simultaneously issued verbally and with a gesture), which promotes robustness in challenging environments and improves command accuracy. This approach also enables operators to leverage the strengths of each modality to provide additional information on base commands, such as giving a verbal command to go to a particular location while providing directional input with a pointing gesture. To minimize the amount of "head down" time, the MINOTAUR watch-based OCU enables quick control inputs through lightweight interactions as well as at-a-glance information status summaries. This enables operators to quickly understand and modify UGV behavior while maintaining focus on the mission at hand.

This paper describes the operational problems faced by MINOTAUR dismounted squad users, as well as the requirements and use case analysis that informed MINOTAUR interface designs. It also provides detailed descriptions of select interface design concepts.

2 Requirements and Use Case Analysis

To inform MINOTAUR design activities, we performed a work domain analysis of an envisioned small team equipped with UGVs conducting operational field maneuvers. This analysis was based on literature reviews and knowledge elicitation interviews with a Marine Corps subject matter expert. As part the analysis, we identified a set of initial support themes to inform design activities, developed a formal abstraction hierarchy [1–3] of a subset of squad and team leader operations, and defined a notional operational scenario.

2.1 Initial Support Themes

Based on the results of our work domain analysis, we defined an initial set of support themes for envisioned squad-based operations. These themes capture a broad set of support needs for human-robot interaction in squad-based contexts, and provide a basis for future analyses (e.g., scenario development, requirements definition), as well as early design activities.

Theme 1: Squad Dynamics. Generally, a platoon consists of three squads, and within each square are three teams. Each team is made up of a team leader, an M249 Squad Automatic Weapon (SAW) gunner, an A-gunner, and a general rifleman. Within a squad, each one of these teams will function as the lead team, scouring the area, while the second team would then serve as the support, with the third team providing overwatch and security for the rear. In any case, each team will require a specific item or resource related to their specific objectives. For example, if a team will be responsible for entering a building, the first team would conduct recon and get to the building first. The second team would then bring the specialized item to assist in breaking down the door. The third team would then provide security. This sort of functional division of labor demonstrates the need for synchronized coordination.

Theme 2: Mission Objectives. A platoon is always going to have a specific mission objective. These objectives could be related to patrol (e.g., securing a perimeter or going out and seeking a target), providing security, conducting search and rescue, or supplying other platoons. Each one of these missions, while functionally similar, will potentially have very different operational tempos and require significantly different forms of support. Generally, a squad would not be sent out if they will be in contact in an overmatch situation (e.g., against a platoon or two platoons); however they must constantly monitor and adapt to the high potential for surprise.

Theme 3: Squad Communication. Coordination and communications depend heavily on the type of mission, the time of day, and the terrain/surroundings the platoon is facing. Communications are primarily conducted through hand-signals and voice communications over digital radios (reliable), as well as verbal commands (e.g., shouting "contact right"). At night, squads will employ night vision goggles (NVGs) that will let them use hand signals, and communicate quietly through radios. Similarly, the definitions of smoke colors would be determined when creating the mission.

For the squads and teams, hand signal communication is heavily dependent on how far away squad members are from one another, with radio comms being the typical fallback. While the most common hand signals are relatively simple, they depend largely on the types of formations used by the team/squad. The types of hand signals used are dependent on the environment – different signals are used in mobile/urban locations than forest or jungle locations.

Theme 4: Managing Spatial Proximity. For a squad, spacing depends largely on the terrain. In an open area, spacing could span 100–200 yards. However, if the terrain is difficult, spacing would likely be closer together (e.g., 75 yards), though the squad and team leader remain mindful of not being too close. Within a team, formation and spacing depends primarily on the mission, terrain, and time of day. In a team of four people, individuals would likely be operating within a 25-yard radius in a smaller area and a 75-yard radius in an open area. Both squad and team spatial organization depend largely on the level of danger of the mission and the potential for receiving enemy fire.

2.2 Envisioned Mission Scenario

Based on the previously described analyses of squad-based operations, we developed a general scenario to explore OCU use cases, HRI, and performance contexts within the setting of this scenario. This highlights coordination opportunities and challenges for an envisioned human-robot ream performing a reconnaissance mission.

Mission Phase 1: Moving Out from the Assembly Area. A platoon receives a mission objective to conduct reconnaissance. The squads will leave from a secure location based on an order within the squad detailing which team would go first and how they would set the perimeter for each team to go out. Once a secure perimeter is established, the other teams with the squad take cover and set security for the team initially leaving the safe line of the assembly area.

After the first team heads out, the team leader determines how best to organize the rifleman, SAW gunner, and A-gunner. As the team progresses toward their assigned area, they move at a pace determined by the team leader. Per their training, each team member takes a few steps (3–4, depending on terrain) before looking back at each other to ensure team members are staying in contact within a particular distance. This relative distance is critical to how the team dictates how and where to give signals. The team leader, at the back of this formation, gives a series of hand signals that are passed to the rifleman (who is unable to see the team leader). Each team will depart the assembly area in a similar manner, and everyone will look back to their squad leader for orders.

For squad-level movements, the squad leader positions himself in the middle of the team leaders. For example, in a wedge position, in which the first team is in the front with a team on either side, the squad leader positions himself in the middle of the formation. The team leaders organize themselves toward the inner parts of their team formations, and frequently look back to receive instructions from the squad leader.

Mission Phase 2: Moving Through Terrain. Although the squad's radios are relatively reliable, visibility decreases as the squad moves through a hilly, heavily wooded area. In these circumstances, it is largely the responsibility of the rifleman to make a good path for the team.

As the squad and teams use different formations to better deal with difficult terrain or guard against potential ambushes, team leaders constantly optimize their position for awareness of the squad leader. The squad and team leaders constantly gauge the location and distance of teams and team members. The squad leader must maintain awareness of all team members and choose the most appropriate formation (e.g., Wedge, Skirmishes Right, or Skirmishes Left). As they move through the terrain, the squad leader relies entirely on the team leaders to communicate positions and key information. The squad is able to use provided intelligence to rapidly locate, assault, envelop, and overcome the enemy.

Mission Phase 3: Additional Support and Return. After the squad has identified, positioned, and engaged with enemy forces, the squad leader decides to drop smoke to identify the need for additional supplies and show their position to indicate where they want additional fire support (either from aircraft or ship-based). The lead team has identified a building for destruction, and the smoke allows the squad to mark their position and provide exact coordinates where they require additional fire support (or extraction). The use of different colored smoke denotes different mission needs, and the definitions of smoke colors are determined during mission planning. Upon successful destruction of the objective, the team identifies the primary path from the objective area and returns safely.

3 OCU Interface Design Concepts

Based on the envisioned scenario, we designed interface concepts that focused on general UGV control, and more specifically on modifying the UGV's following behavior. As part of the MINOTAUR effort, we explored a broad range of control inputs, display devices, and interaction methods. The MINOTAUR multimodal Operator Control Unit (OCU) consists of a watch-based visual display, and accepts physical touch inputs, gesture-based inputs, and voice inputs. This multi-modal approach will enable the operator to flexibly and opportunistically employ input methods based on specific operational needs. It also enables redundant and orthogonal commands across multiple modalities.

We utilized an iterative, work-centered approach grounded in established Cognitive System Engineering (CSE) methods [4] to develop OCU interface concepts that minimize learning time and are well suited for the warfighter in envisioned squad-based operational contexts. The OCU provides quick control inputs through lightweight interactions as well as at-a-glance information status summaries to minimize the amount of operator heads down time. It will also increase the *observability* and *directability* [5, 6] of UGV functions to enable rapid and robust interactions with the robot in dynamic operational environments. Increased observability provides the operator with insight into the current and future activities of automated processes. Observability techniques also include support for operator understanding of the limitations of automation (e.g., speed constraints, connectivity problems). Increased directability enables the operator to efficiently and purposefully direct and re-direct resources, activities, and priorities as situations change and escalate.

3.1 Natural Multimodal Interface Design Concepts

Across the MINOTAUR multimodal interface toolkit, we designed and prototyped interface concepts to enable operators to provide control inputs over multiple modalities, including touch, gesture, and voice inputs [7]. A key advantage of multimodal information display and input methods is their ability to improve the amount of information that can be conveyed to and provided by the operator, as well as the likelihood that the operator will perceive and respond to conveyed information. We will purposefully leverage channels and rendering methods that will be perceptually compelling and successful in squad-based operation contexts. One important aspect of multimodal information design is the consideration of priority of perceptual channels, as information in some channels is harder to ignore, particularly when information conflicts.

The MINOTAUR multi-modal interface approach will enable the operator to *flexibly and opportunistically* determine which interface modality or modalities to employ. For example, the operator could use voice commands in relatively safe operating conditions, and gestures when voice commands are dangerous due to nearby enemies. This approach also enables the operator to *redundantly* provide commands across multiple modalities to improve transmission of information. For example, the operator could simultaneously issue a stop gesture and voice stop command. This promotes robustness in challenging environments (e.g., degraded sound quality or visibility), and improves the accuracy of commands.

Finally, our approach enables orthogonal commands across multiple modalities, such as verbally directing the robot to go to a location and pointing a direction to convey additional, more specific spatial information. For example, MINOTAUR's multi-modal interface approach enables the operator to *orthogonally* provide commands across multiple modalities, such as verbally directing the robot to go to a location and pointing a direction to convey additional spatial information. This approach leverages the strengths of each modality, and enables the operator to provide additional information on base commands. This approach leverages the strengths of each modality, and enables the operator to provide additional information on base commands.

3.2 Watch-Based Control Interface

Initial MINOTAUR design activities focused on the Operator Control Unit (OCU), which enables the operator to interact with the UGV through a watch-based form factor. We designed a Tracker View, which provides observability and directability of current UGV behavior, a Command Status display, which provides observability of the UGV's processing of received commands, and a Command Log, which provides observability of the history of commands provided to the UGV.

Follow Modes. An initial focus of MINOTAUR OCU design efforts was a Tracker View, which provides observability of the UGV's current following behavior and enables the operator to modify the manner in which the robot follows the human operator through lightweight interactions. We explored a range of interaction methods for modifying robot behavior, including toggling and drag-and-drop concepts.

Figure 1 below shows a workflow for changing the robot's follow mode from loose to exact using the Tracker View. In this figure, the pink lines show the operator's path through the interface. The pink circles indicate the location of user interactions or gestures, such as finger presses and drag-and-drop locations. The blue circle with white outline (located at the bottom of the screens) represents the robot and the white circle with the black X-shape (located at the top of the screens) represents the human team leader. The unbroken white and green line represents exact following and the dashed, curved green and white line represents loose following. The first panel of Fig. 1 shows the primary Tracker view display, which provides a high level summary of the robot's current leader/follower mode behavior (e.g., exact/loose) through a vertical orientation to visually reinforce the specified leader/follower configuration. Additional support information on the execution of the behavior, such as distance and speed, appears in the bottom right hand corner. The operators can set distance and speed as constraints or they can be derived from the vehicle itself. User-defined constraints are displayed with a lock icon preceding the data value. This allows the operator to maintain a higher level of control over the robot's task execution.

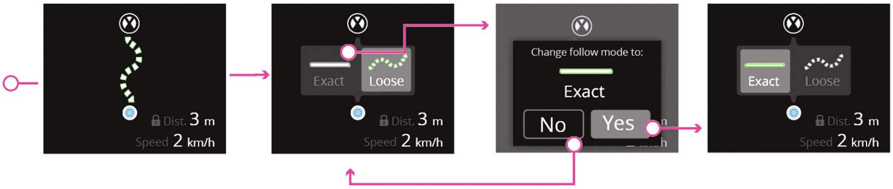


Fig. 1. A notional series of interface concepts for changing the robot's follow mode from loose to exact using the Tracker View. The vertical orientation of operator and robot symbols visually reinforces the specified leader/follower configuration. Salience cueing visually guides the operator through the task of switching the robot's follow mode, while a two-step confirmation prevents operator errors.

Within the Tracker View, the operator can toggle between the robot's two follow modes – exact and loose. In exact mode, the robot follows the operator's exact route. In loose mode, the robot autonomously navigates its own route. The second panel of Fig. 1 depicts the updated touch-based toggle control. In this toggle control, the operator can see visual representations of each mode, with supporting text describing the available modes. This toggle also provides salient cueing to the operator of the robot's current follow mode. Providing this additional context directly within the toggle control itself reinforces the operator's mental model of the current state of the robot and accelerates learning of controls and available options for new users who may be unfamiliar with the interface. To prevent accidental touch-based inputs, this view employs a two-step confirmation approach. The third panel displays a dialog box that shows the operators new along with "yes" and "no" buttons. Finally, the last panel of Fig. 1 displays the updated control for switching from the new exact mode to the original loose mode.

Figure 2 below shows a parallel transition from *exact* follow mode to *loose* follow mode. In the first panel, the robot is currently following the operator and must maintain a distance of 3 meters (denoted by the lock icon). The second panel depicts the updated control depicting the current mode (*exact* follow mode) along with the *loose* follow mode option. As the operator selects the *loose* option, a confirmation dialog pops up to

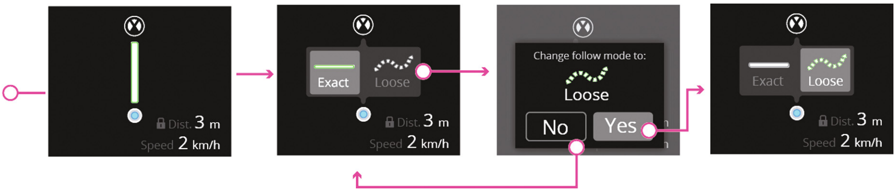


Fig. 2. A notional series of interface concepts for changing the robot's follow mode from exact to loose using the Tracker View.

confirm the change, as seen in the third panel. Finally, the fourth panel depicts the updated control for switching from the new *loose* mode to the original *exact* mode.

Command Status. The Command Status Region appears below the Tracker View and provides observability of the processing status of commands issued to the UGV, including receipt of commands, processing of commands, and acceptance/execution of commands. Figure 3 below shows a series of screens that reflect the transition from exact follow mode to loose follow mode. At the top left of the figure, both the Tracker View and Command Status region show the UGV in "following exact mode." The second display shows that a "follow loosely" command was received. The smaller size and less salient color of text and "Next:" text helps the operator understand that the command has been received, but is not yet being executed. In the next screen, a "refreshing" symbol appears on the right side of the Command Status region to indicate that the UGV is processing the new command. In the tracker view, the solid line between the leader and robot icons is semi-transparent to indicate that a new command is being processed. In the next screen, the follow loosely command is shown in larger white text at the top of the region to indicate that the command has been accepted and is now being executed, while the previous command is shown in smaller, darker text. The "following exact" symbol in the Tracker View (solid white line) has also been replaced with a "following exact" symbol (curved dashed line). Eventually, the text for the previous command disappears, leaving only the current command (as shown in the final screen in Fig. 3).

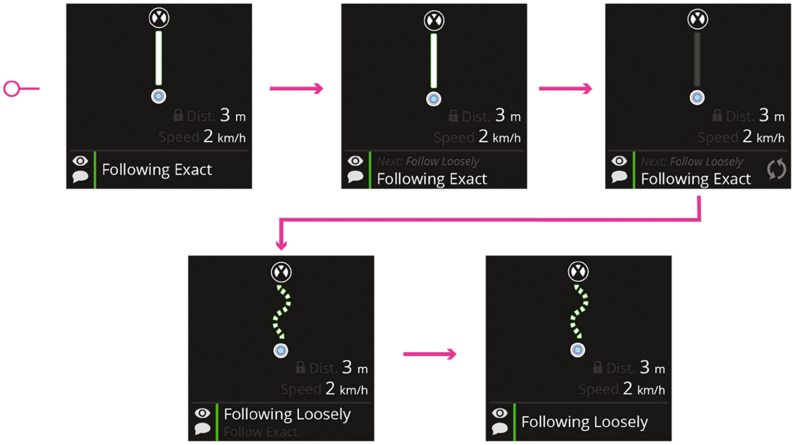
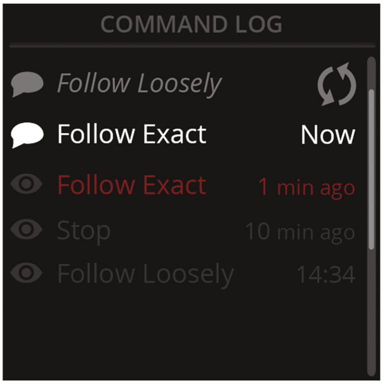


Fig. 3. A series of screens illustrating the Command Status Region, which appears below the Tracker View and provides observability of command status, including receipt of the command, processing, and acceptance/execution. This display region uses variable salience and integrates with the Tracker View to promote operator understanding of command status.

Command Log. The Command Log display provides observability of the history of commands provided by the operator to the UGV. The operator accesses this display by touching the Command Status Region, and enhances operator awareness of robot functioning in context. It can also provide the operator with insight into the effectiveness of different command modalities (e.g., the operator can see that no gesture inputs have been accepted over the course of a mission).

Figure 4 shows the command log display, which shows operator commands ordered by recency, with the most recent command appearing at the top of the display. Each command in the log occupies a row within the display. An icon to the left of each command indicates the modality with which it was provided (e.g., speech bubble icon for a voice command, eye icon for a gesture command). The time since the command was given is shown for commands issued a short time ago (e.g., "1 min ago", "10 min ago) and a timestamp is shown for commands issued longer ago. All past commands are displayed with dark grey text. The word "now" appears next to commands that are currently being executed, which are shown in white text, and a refreshing symbol appears next to commands that are being processed, which appears in light grey text. Finally, failed commands are shown in red text. Salience mapping throughout the Command Log helps the operator quickly understand the various command statuses, and quickly identify commands that failed or that are pending.



COMMAND LOG

Follow Loosely	
Follow Exact	Now
Follow Exact	1 min ago
Stop	10 min ago
Follow Loosely	14:34

Fig. 4. The Command Log display, which provides observability of the history of commands provided by the operator to the UGV, and enhances operator awareness of robot functioning in context. Relevant command properties (e.g., time since command was given, current command, pending commands, failed commands) are provided, and salience mapping helps the operator to quickly understand current, past and future UGV functioning.

4 Conclusion and Future Work

This paper described *Multi-modal Interface for Natural Operator Teaming with Autonomous Robots* (MINOTAUR), a human-robot interface for dismounted squad operations. This effort built upon analyses of squad-based operations to develop OCU concepts that improve observability and directability of UGV functions through lightweight interactions and at-a-glance information summaries. Examples of OCU concepts were presented, including a Tracker View, Command Status display, and Command Log.

This work sets the stage for continued development of OCU display concepts to accommodate additional squad-based use cases, such as waypoint-based navigation. Another focus area for follow-on efforts is "progressive enhancement" of UGV commands (i.e., commanding the UGV to go to a location and later modifying that command so the UGV chooses its own route and goes to the location more quickly). This would allow the operator to update commands based on the current operational context without needing to repeat commands unnecessarily. Future efforts will also explore robust UGV health and status displays. For example, status displays for the various command modalities will explore ways to provides operator cues that enable graceful degradation when one or more modalities are unavailable. By alerting the operator to failures as they occur, health and status displays will also enable the proactive management of UGV issues. The key challenge of these displays will be the balance between showing critical information and alerts while minimizing operator heads down time. Because we developed a broad set of display concepts, future efforts will also focus on user testing to refine our OCU designs.

References

1. Vicente, K.J.: Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer-Based Work. Lawrence Erlbaum Associates, Mahwah (1999). Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)
2. Rasmussen, J., Pejtersen, A., Goodstein, L.P.: Cognitive Systems Engineering. Wiley, New York (1994)
3. Woods, D.D., Roth, E.M.: Cognitive engineering: human problem solving with tools. Hum. Factors **30**(4), 415–430 (1988)
4. Hollnagel, E., Woods, D.D.: Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press, Boca Raton (2005)
5. Woods, D.D., Hollnagel, E.: Joint Cognitive Systems: Patterns in Cognitive Systems Engineering. CRC Press, Boca Raton (2006)
6. Kilgore, R., Voshell, M.: Increasing the transparency of unmanned systems: applications of ecological interface design. In: Shumaker, R., Lackey, S. (eds.) Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality. Springer, Orlando (2014)
7. Oviatt, S.: Multimodal interfaces. In: Jacko, J., Sears, A. (eds.) Handbook of Human–Computer Interaction. Lawrence Erlbaum, Mahwah (2002)

UAS Detect and Avoid – Alert Times and Pilot Performance in Remaining Well Clear

Rania W. Ghatas[1(✉)], James R. Comstock Jr.[1], Michael J. Vincent[1],
Keith D. Hoffler[2], Dimitrios Tsakpinis[3], and Anna M. DeHaven[4]

[1] NASA Langley Research Center, Hampton, VA, USA
{Rania.W.Ghatas, James.R.Comstock,
Michael.J.Vincent}@nasa.gov
[2] Adaptive Aerospace Group, Hampton, VA, USA
KHoffler@adaptiveaero.com
[3] SAIC, Inc., Hampton, VA, USA
Dimitrios.Tsakpinis@nasa.gov
[4] Craig Technologies, Inc., Hampton, VA, USA
Anna.M.DeHaven@nasa.gov

Abstract. With the rapid growth of Unmanned Aircraft Systems (UAS), NASA was called upon to examine crucial operational and safety concerns regarding the integration of UAS into the National Airspace System (NAS) in collaboration with the Federal Aviation Administration (FAA) and industry. Key research efforts paper focused on understanding and developing requirements for Detect and Avoid (DAA) systems and making sure they are interoperable with Collision Avoidance (CA) technologies. These requirements detail necessary performance of a DAA system designed to help the UAS pilot maintain DAA Well Clear (DWC) from intruder aircraft so that safe separation is retained. NASA Langley's Human-in-the-Loop (HITL) simulation study known as Collision Avoidance, Self-Separation, and Alerting Times (CASSAT) addressed these DAA requirements in a two-phase study. The first phase examined eleven active air traffic controllers. The second phase, addressed in this paper, examined twelve pilots' interactions with DAA systems at simulated UAS ground control stations (GCS).

Keywords: Human Factors · Human-Systems Integration · Human-in-the-Loop · Unmanned Aircraft Systems Integration · National Airspace System · Collision Avoidance · Self-Separation · Alert times · Alerting structures · Alerting Logic · Pilot performance · Detect and Avoid · Well Clear

1 Introduction

Unmanned Aircraft Systems (UAS) have become the forefront of aviation technology and will soon be commonplace in the National Airspace System (NAS). As routine access to the NAS becomes a reality, UAS will be required to have new equipage, minimum operational performance standards, rules and regulations, and procedures. Answers to difficult questions concerning these standards, regulations, and procedures

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_11

will be required through many supporting research efforts. With safety being the primary concern, the National Aeronautics and Space Administration (NASA) has established a “UAS Integration in the NAS” project that spans four NASA centers, and in collaboration with the FAA and industry, is examining essential safety concerns regarding the integration of UAS in the NAS. Detect-and-Avoid (DAA) implementations and Collision Avoidance (CA) technologies to remain DAA Well Clear (DWC) of other aircraft are top research priorities in assuring safe integration. Research efforts at NASA Langley Research Center are providing data to help answers to those difficult questions to assure safe and efficient integration of UAS into the NAS.

The DAA technology employed in the Collision Avoidance, Self-Separation, and Alerting Times (CASSAT) study worked much like the algorithms in the Traffic Alert and Collision Avoidance System (TCAS), but with a DWC volume that was large enough to avoid (a) corrective Resolution Advisories (RAs) for TCAS equipped intruders and, (b) undue concern for proximate see-and-avoid pilots. The DWC volume is defined by RTCA SC-228 [1] as having a 35 s Modified Tau threshold, 4,000 feet horizontal miss distance, and 450 feet vertical miss distance. Modified Tau is an approximation of the time to Closest Point of Approach (CPA) used in the TCAS II Resolution Advisory logic. Using a Modified Tau threshold (TAUMOD) of 35 s means that, when Modified Tau is less than 35 s, the aircraft are considered to be in violation of remaining DWC in the horizontal dimension. This DWC volume is combined with an additional parameter known as Time to Co-Altitude (TCOA), which is the time threshold in the vertical dimension [2]. The present series of studies sought to determine operationally acceptable DAA sizes and alert times to inform system designers about required DAA surveillance range. Guidance from the DAA system was provided to the UAS pilot to maintain positions outside the well-clear boundary. Details of the Self-Separation guidance shown to the UAS pilots to maintain well-clear may be found in Chamberlain et al. [3].

2 Approach and Objectives

The primary focus of the pilot-acceptability CASSAT study was to address minimum and maximum acceptable alert times for projected DWC losses from the perspective of Instrument Flight Rules (IFR) rated pilots with and without experience controlling large size UAS, such as a Predator or Global Hawk. Pilots controlled simulated DAA equipped Unmanned Aircraft (UA) and provided ratings on acceptability of distance-threshold (DTHR) values when near traffic encounters occurred, acceptability of Alerting Times, workload ratings during test conditions, and feedback regarding the alerting structure. The DTHR value, along with tau-mod and Alerting Times, is the distance that the Detect and Avoid Alerting Logic for Unmanned Systems (DAIDALUS) algorithms use to provide maneuver guidance bands to the pilot, and was experimentally varied in the study. The top-level functionality of DAIDALUS provides traffic awareness and maneuver guidance supporting UAS operators’ ability to Detect and Avoid other aircraft that have the potential to cause a Loss of Well Clear (LoWC). Horizontal miss distance, used simultaneously with distance threshold (DTHR) in this paper, is a fixed minimum value that constitutes a DWC loss.

3 Method

3.1 Subjects

Twelve pilots from across the country were recruited to perform traffic separation tasks for the scenarios developed. All twelve of the pilots were instrument rated; six of the twelve pilots had experience flying manned aircraft exclusively while the remaining six had additional experience flying Unmanned Aircraft, such as Predators and Global Hawks. Pilots participated for 2 days each, and each data collection session consisted of two subject pilots independently flying a simulated UAS in the Dallas Fort-Worth (DFW) East-side airspace. Subjects were trained on the DFW airspace, the DAA concept, and the simulation environment upon arrival. All subjects had experience communicating with Air Traffic Control (ATC) and held current in their pilot ratings and certifications.

3.2 Procedure

Airspace traffic scenarios for this study were designed so that there were 14 UAS traffic encounters per hour with a total of six one-hour sets split between the two data collection days. Traffic scenarios were split between two ground control stations (GCS-1 and GCS-2) so that each pilot saw seven UAS traffic encounters during each test hour for a total of 42 encounters per pilot, and for all twelve subjects, this meant a set of data with 504 encounters. Both ground control stations displayed the DAA Self-Separation guidance information in real-time along with one of two alerting structures (discussed in Sect. 3.3). Additionally, TCOA for GCS-1 was set to 0 and GCS-2 was set to a TCOA value of 20 – these values remained constant throughout the experiment. TCOA is a parameter considered useful in adding alert time to vertical encounters. On Day 2, the pilots switched seats to allow exposure to each TCOA value. This TCOA value parameter only affected traffic encounters approaching vertically (from above or below). The two alerting structures, which contained different levels of aural and visual cues to alert pilots of oncoming and/or nearby traffic, were presented to the pilot subjects. To maintain a real-world environment and a workload similar to that of actual DFW traffic, the background traffic was controlled by two pseudo-pilots at two additional pilot stations located in a separate simulation room. The DFW East-side controller, with whom the test subjects were communicating, was part of the research team.

The pilot subjects began control of a given UAS with the aircraft already in flight by a handoff process, and when the aircraft completed the traffic encounter, control was handed off to a pilot on the research team. Each subject controlled one aircraft at a time. Sufficient time was allowed between controlling each aircraft so that post-encounter questions could be answered and rating scales completed. Two pilots using separate GCS stations were run at the same time. To increase workload and add some degree of distraction from just looking for traffic encounters to appear, members of the research team asked the pilot subjects questions that required each pilot to conduct a map search task while flying the UAS in the scenarios. Secondary task questions were tailored to each test session so that the questions matched the map area in which the encounters were scripted to occur.

Voice communications channel had a 400 ms two-way delay between the simulated UAS and ATC. Additional manned and unmanned traffic were also on the ATC frequency and, if within range, could be visible on the map display as selected by the subject pilots. Delays in communicating with ATC were possible with the level of traffic simulated, due to voice traffic congestion. Traffic encounters were between the UAS and Visual Flight Rules (VFR) traffic that was transmitting position and altitude information but was not in voice communications with ATC, thus all aircraft maneuvering, if required, was performed by the UAS.

3.3 Independent Variables

The independent variables evaluated in this study are shown in Table 1; however, the focus of this paper will be on discussing results for alert times and alerting structures (Fig. 1).

Table 1. Independent variables evaluated in this study.

Independent values	Number	Values
Horizontal Miss Distance (DTHR)	3 or 4 values	0.7, 1.0, and 1.5 nmi for Crossings and Overtake encounters
		1.0, 1.5, and 2.0 nmi for Head-on encounters
Alert time	3 values	40, 60, and 75 s
Encounter geometry	5 values	Head-on, Overtake, Crossing, Vertical Overtake, and Vertical Crossing
Time to Co-Altitude (TCOA)	2 values	0 and 20 s
Alerting structure	2 types	"A" and "B"

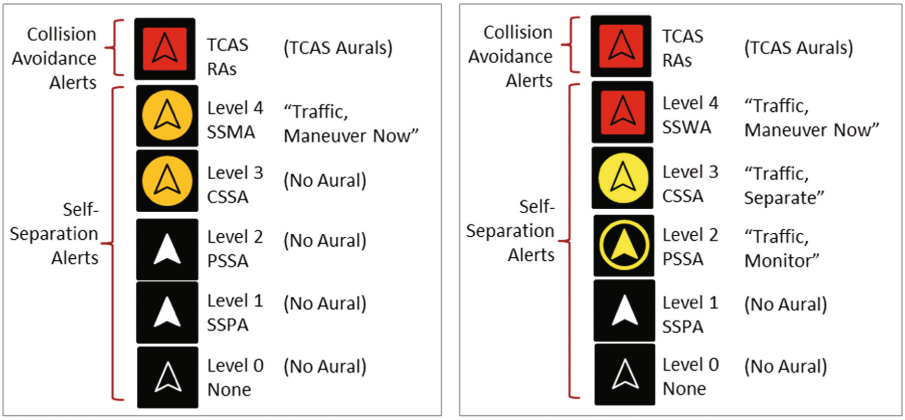


Fig. 1. Two alert structures were shown to the pilot subjects; Alert Structure "A" (left) and Alert Structure "B" (right). One alerting structure was shown to the pilots on Day 1, and on Day 2, the pilots saw the other alerting structure. The order of presentation of alerting structures was counterbalanced across the six pairs of test subjects.

3.4 Scenarios

The airspace modeled for this study was a portion of airspace delegated to the DFW Terminal Radar Approach Control Facility (TRACON) (D10), specifically, Sector DN/AR-7 South Flow. The majority of UAS traffic arrived or departed McKinney National (FAA airport identifier: KTKI), formerly known as Collin County Regional, and is approximately 28 nautical miles (nmi) northeast of DFW. The scenarios were designed and situated in this airspace so as to enable various encounter geometries between the UAS and intruder aircraft set within a realistic background of manned aircraft traffic in order to achieve realistic levels of workload for the pilot subjects. A colored chart of the area was used during the initial training session on Day 1 to familiarize the pilot subjects with the airspace, which is depicted in Fig. 2.

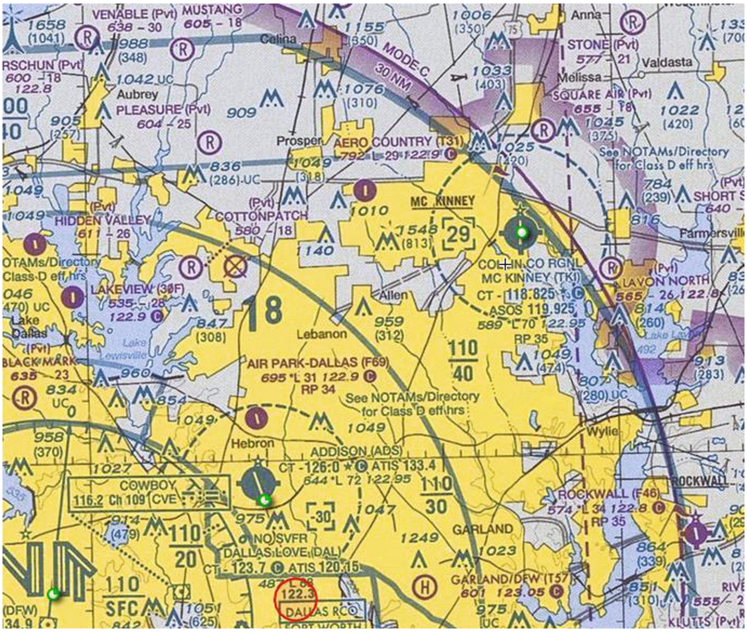


Fig. 2. Chart of area used in initial training. This shows a figure depicting Dallas Fort-Worth (DFW) in the lower left and McKinney National (KTKI) approximately 28 nmi northeast of DFW (upper right).

3.5 Communications, Navigation, and Surveillance Assumptions

The experiment assumed Communication, Navigation, and Surveillance (CNS) architectures and capabilities appropriate for current-day operations in the applicable airspace classes, and that these capabilities were available to all aircraft (manned and unmanned) in the simulation environment. UAS were communicating with ATC in a similar manner to the manned aircraft. The intruders were VFR traffic that were operating with a transponder on but were not in voice communications with ATC. UAS command, control, and communication capability was assumed available between the UA and their respective GCS. The UA was assumed to be capable of receiving/transmitting voice communications to and from ATC facilities and proximate "party line" aircraft via Very High Frequency (VHF) radio in the same manner as manned aircraft in the same airspace and of relaying these voice communications to/from the GCS pilot via one or more UA-GCS links. "Party line" refers to the open radio channel through which all aircraft in a given airspace communicate with ATC; pilots are able to hear their own clearances in addition to those of the other aircraft. It was assumed that, in addition to the relayed voice communications, the UA-GCS link(s) carried all command/control data between the UAS and GCS. The communications delays were for the voice communications channel only and no delay was introduced for UAS control or position reporting.

3.6 Dependent Variables

Horizontal Miss Distance (DTHR). After each traffic encounter, a member of the research team, who was seated next to the pilot subjects, would ask "How was the spacing of that last encounter?" or "How acceptable was the miss distance in the previous encounter?" Subjects had a copy of the rating scale definitions available to them during the test sessions. The rating scale consisted of five assessment definitions, including (1) much too close; unsafe, (2) somewhat close, (3) neither unsafely close nor disruptively large, (4) somewhat wide; might be disruptive, and (5) excessively wide; disruptive. Fractional responses, such as 1.5 or 3.5, were acceptable.

Alert Time Ratings. An alert (at either 40, 60, or 75 s) was provided to the pilot by the guidance algorithms when a maneuver was required to avoid an intruder aircraft. The alert time at which the guidance requested a maneuver was rated according to the scale shown in Table 2. Subjects had a copy of the scale definitions available to them during each test hour. The same researcher who asked questions regarding the horizontal miss distance would also ask the pilot subject "How was the alert time of that last encounter?" or "How acceptable was that alert time?"

Table 2. Rating scale definitions used for assessment of alert time.

Rating Score	Definition
A	Too Early: Request made too early; potentially disruptive if a maneuver not required
A/B	Between A and B
B	Timing Okay: Not too early or too late; timing of request completely acceptable
B/C	Between B and C
C	Too Late: Request made too late; potentially disruptive to adjacent traffic if a large maneuver required

System Performance Metrics. Data concerning encounter aircraft separation distances were recorded throughout the period of the encounter and included aircraft-to-aircraft separation distances and time to the CPA.

Workload and Secondary Task. A workload assessment was requested after each encounter and was filled out by each subject via the end-of-encounter questionnaire on a seven-point scale. Forty-two assessments of workload were collected from each subject. Additionally, in an attempt to increase workload and add some degree of distraction from looking for traffic encounters to appear, members of the research team added a secondary task to the design of the experiment. Questions were asked that required the pilots to conduct a map search task while flying the UAS in the scenarios.

Questionnaires. Two separate questionnaires were utilized in the course of each one-hour session. The first was a post-encounter questionnaire, which was administered after each encounter resulting in seven total post-encounter questionnaires per pilot per one-hour test session. The second was a post-run questionnaire, which was administered after each one-hour test session and was used to record ratings and comments on the preceding test session.

4 Results

4.1 Alert Time Ratings

Three alert time values were tested during the study, which were counterbalanced across the six one-hour test sessions, such that all three alert time values were presented on Day 1 and again on Day 2 of the study. The timing of the alert is represented in seconds prior to breaching DWC. Figure 3 shows the alert time ratings for the crossing geometry encounters. The majority of responses were in the "Timing Okay" category, with no systematic changes attributable to the alert time manipulation.

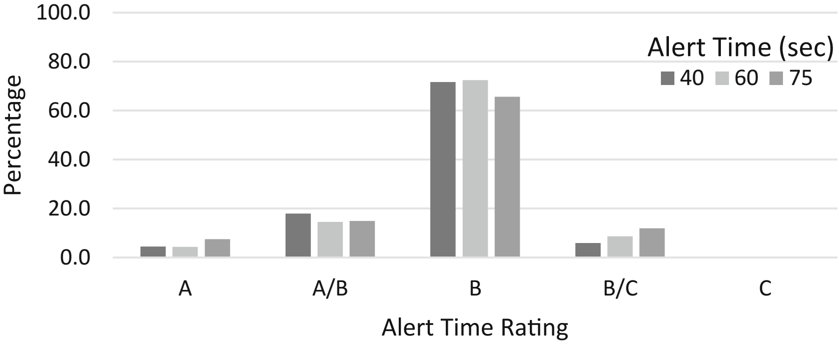


Fig. 3. UAS Pilot ratings of alert times for crossing geometry encounters (A = Too Early, B = Timing Okay, C = Too Late).

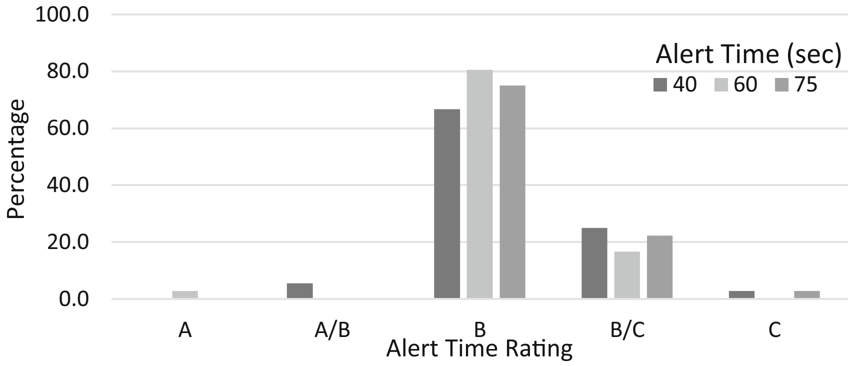


Fig. 4. UAS Pilot ratings of alert times for overtake geometry encounters (A = Too Early, B = Timing Okay, C = Too Late).

Figure 4 shows the alert time ratings for the overtake geometry encounters where the majority of the rating responses were in the "Timing Okay" category, but with about 20% of responses in the "Between B and C" (between Okay and Too Late) category. The speed differential between the UAS and the intruder aircraft was smaller in the overtake geometries than in the other geometries, and therefore allowed the aircraft to get closer in distance before the time-based algorithm indicates that it is time for a maneuver. As in the crossing case, no systematic differences are noted across the three alert times tested.

Figure 5 shows the alert time ratings for the head-on case. As in the other alert time geometries, the majority (>75%) of ratings were in the "Timing Okay" category, regardless of the alert time value.

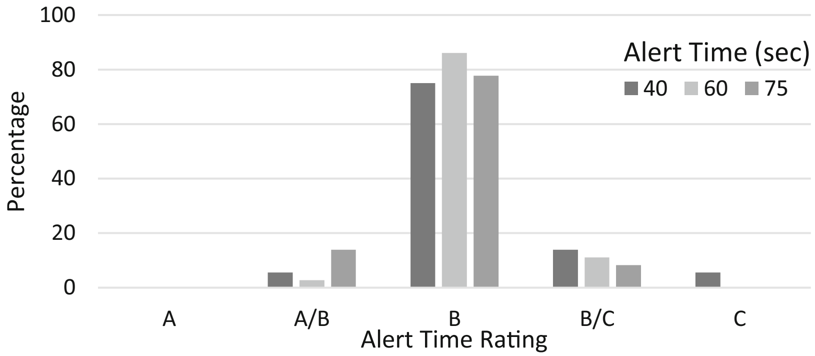


Fig. 5. UAS Pilot ratings of alert times for Head-on Geometry Encounters (A = Too Early, B = Timing Okay, C = Too Late).

4.2 Vertical Encounters

Data were collected from a total of 72 vertical encounters in the study, meaning that each GCS pilot saw one in each hour, or six over the 2 days of testing. All vertical encounters had a DTHR of 0.7 nmi. Figure 6 shows the distance ratings by the GCS pilots and illustrates that a sizable proportion of these were considered "Too close" or "Somewhat close." The alert time variable appeared to have no systematic effect on the distance ratings.

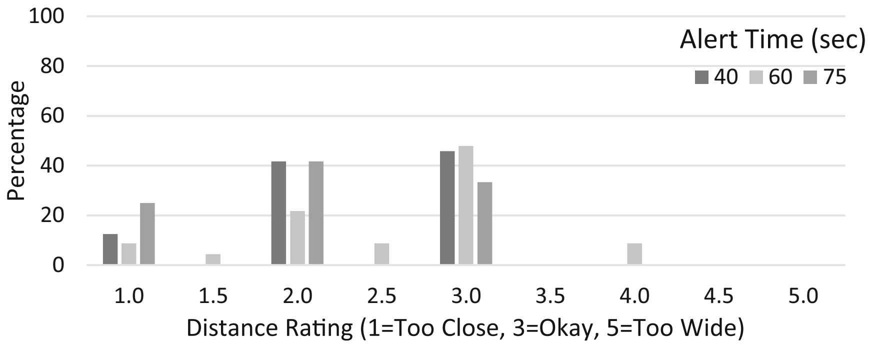


Fig. 6. UAS Pilot ratings of Distance for Vertical Geometry Encounters for each alert time. All DTHR values for these encounters were 0.7 nautical miles.

The GCS pilot alert time ratings, shown in Fig. 7, show a shift in responses towards the "Too Late" end of the scale compared to other encounter geometries. In Fig. 7, the number of responses to the right of "Timing Ok" were 62% for the 40 s alert time, 50% for the 60 s alert time, and 33% for the 75 s alert time. As expected, the 60 s and 75 s alert times allowed the UAS pilots additional time to assess the situation and make a maneuver and diminished the "Too Late" responses.

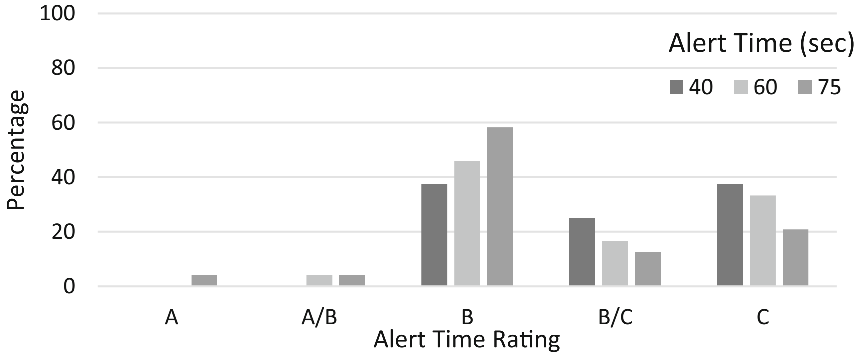


Fig. 7. UAS Pilot ratings of alert time for Vertical Geometry Encounters (A = Too Early, B = Timing Okay, C = Too Late).

Figure 8 shows the distance from DTHR at the CPA for each of the alert times. There were only two losses of DWC among the 72 vertical encounters. This figure also shows a breakdown in the data set for the TCOA parameter in the DAA algorithm. Also worth noting in this figure is that 40 and 60 s alert time encounters were vertical overtakes, while 75 s alert time encounters were of the vertical crossing geometry. In agreement with non-vertical encounters, the vertical crossing geometry encounters have larger mean CPA - DTHR differences.

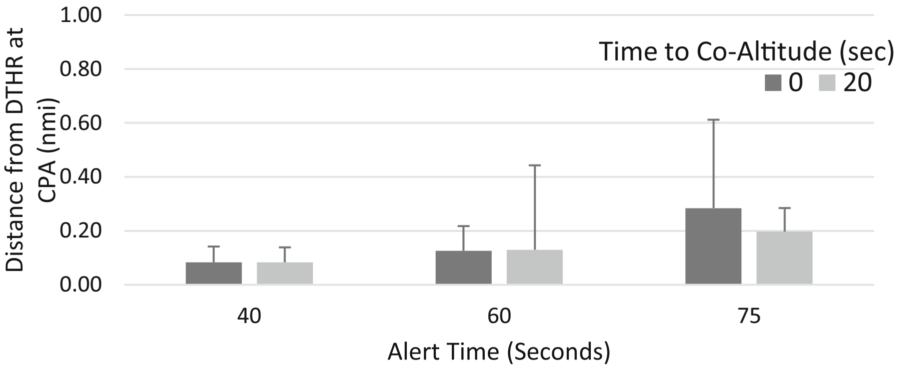


Fig. 8. Mean distance from DTHR at CPA for Vertical Geometry Encounters for 40, 60, and 75 s alert times and Time to Co-Altitude values of 0 and 20 s (72 encounters, 2 losses of DAA Well Clear). Encounters at 40 and 60 s are Vertical Overtake Encounters, 75 s encounters are Vertical Crossing Encounters. Error bars denote standard deviations.

4.3 Alerting Structures

At the end of every one-hour test session, each UAS pilot was asked to answer a series of fifteen questions on an end-of-hour questionnaire. Seven of the fifteen questions focused on the alerting structure that was presented to the pilot subjects on the GCS display and consisted of the following questions: (1) Rate the ease of the display, (2) Rate the ease of the alerting structure, (3) Was the display user-friendly?, (4) Did the display provide the necessary information to predict a potential loss of separation? (5) Rate your trust of the bands on the display, (6) Were the number of alert icons during the past hour acceptable? and (7) Were the number of aural alerts during the past hour acceptable?

Based on responses provided on the end-of-hour questionnaires and during the de-brief session on Day 2 of testing, alerting structure “A” was the preferred schema. Pilot subjects commented that alerting structure “A” was user-friendly and intuitive. Additionally, although some subjects also found alerting structure “B” to be user-friendly, the majority of them did not find having two "red" icons to be appealing. Multiple pilot subjects commented that simpler is better in terms of alerting structures.

5 Discussion

Three alert times were tested in the DAIDALUS algorithm, which included 40, 60, and 75 s. The majority of ratings by the UAS pilots were in the "B = Timing Okay" category regardless of the alert time. As illustrated by Figs. 3, 4, and 5, very few rating responses occurred at either the "A = Too Early" or "C = Too Late" ends of the scale, regardless of the encounter geometry. These results suggest little change in perception of an acceptable alert time across the range of alert times evaluated. Based on these results, none of the values tested can be considered excessive, leading to a nuisance alert for the UAS pilot. Likewise, very few "Too Late" responses indicate that even the shortest alert time tested still provided enough time to contact ATC to negotiate a maneuver.

For the vertical encounters, a different picture emerges from the alert time ratings. Here, a higher percentage of rating responses were found on the "C = Too Late" end of the scale than for lateral maneuvers. Despite this, the greatest frequency still occurred at the "B = Timing Okay" category, as in the lateral maneuver conditions. The CPA-DTHR difference was not affected by the TCOA value (see Fig. 8).

Of the total of 504 encounters, only 17 (3.4%) had distance-based (CPA-DTHR) losses of DWC. These DWC losses occurred over all DTHR values and for all of the geometries under test. The most frequent cause was starting the traffic separation maneuver too late.

Acknowledgements. The support of the many people involved in the conduct of this study is gratefully appreciated. Thanks are extended to Aaron Dutle, George Hagen, Cesar Muñoz, Anthony Narkawicz, and Jason Upchurch of the Safety Critical Avionics Systems Branch at the NASA Langley Research Center for contributions to the development of the DAIDALUS Self-Separation algorithms. Thanks are also extended to contributors participating through the NASA Langley Information Technology Enhanced Services (LITES) contract, including: Pierre Beaudoin, Kevin Chan, Joel Ilboudo, Kristen Mark, Robb Myer, Gaurev Sharma, and Jim Sturdy, as well as Steve Hylinski from Adaptive Aerospace Group, Inc., and other support and management personnel. Thanks are also extended to members of the Air Traffic Operations Laboratory (ATOL) for their support, including: Chad Chapman, Ronald Maddox, and Edward Scearce. Thanks are also extended to the tireless efforts of the persons staffing the background traffic pilot stations, the UAS pilot stations, and the ATC adjacent sector/tower consoles.

References

1. McDuffee, P.: Unmanned Aircraft System (UAS) standards development. RTCA SC-228 status. In: ICAO RPAS Symposium (2015). http://www.icao.int/Meetings/RPAS/RPASSymposiumPresentation/Day%202%20Workshop%205%20Technology%20Paul%20McDuffee%20-%20Unmanned%20Aircraft%20System(UAS)%20Standards%20Development%20RTCA%20SC-228%20Status.pdf
2. Upchurch, J.M., Muñoz, C.A., Narkawicz, A.J., Consiglio, M.C., Chamberlain, J.P.: Characterizing the effects of a vertical time threshold for a class of well-clear definitions. In: Eleventh USA/Europe Air Traffic Management Research and Development Seminar (2015). https://shemesh.larc.nasa.gov/people/cam/publications/ATM-2015-356.pdf
3. Chamberlain, J.P., Consiglio, M.C., Comstock Jr., J.R., Ghatas, R.W., Muñoz, C.: NASA Controller Acceptability Study 1 (CAS-1) Experiment Description and Initial Observations. NASA/TM-2015-218763. Langley Research Center, Hampton (2015)

Advances in Unmanned Aerial Vehicles (UAV) and Drones Research

Drone Relay Stations for Supporting Wireless Communication in Military Operations

Seon Jin Kim[1], Gino J. Lim[1(✉)], and Jaeyoung Cho[2]

[1] Industrial Engineering, University of Houston, Houston, TX 77204, USA
ginolim@uh.edu
[2] Industrial Engineering, Lamar University, Beaumont, TX 77705, USA

Abstract. This paper proposes a concept of drone relay stations for supporting wireless communications in military operations. Communications in military operations are crucial in completing desired missions under highly chaotic situations. Due to dynamic movements of troops, multiple units require wireless communications when they are separated from each other to coordinate the efforts for the mission. Relay stations are often used to enhance wireless signals. Current methods include ground vehicles, helicopters, or satellites to relay wireless signal in military operations. However, these methods have known limitations such as inaccessible areas (mountains, ice roads, etc.), potential exposure to the enemy force, and high operating costs. Hence, this work proposes the concept of drone relay stations to overcome these limitations and explore enabling technologies that have been developed.

Keywords: Drone · Relay station · Wireless communication · Military

1 Introduction

Communication on the battlefield is a vital factor to guarantee the success of military operations [1]. It enables units to conduct successful operations by sharing the location and current state of each unit. Specifically, when a unit is located in enemy area or when wired communication is impossible, wireless (radio) communication is typically used as shown in Fig. 1. In order to improve the range and quality of communication, wireless communication is supported in the field through mobile communication networks such as mobile relay stations that are operated along the movement path of ground forces or supporting units [2].

There are two main types of support for wireless mobile communications: a ground relay station or an aerial relay station.

To ensure smooth communication between units in real time, the ground mobile relay station is equipped with a communication relay device in the vehicle that follows a particular ground unit and supports communication service in a restricted area. However, ground mobile relay station vehicles are limited in their deployment to specific areas (inaccessible areas such as mountains, ice roads, etc.). Also, since it is based on wireless communication, signal can be interfered by surrounding features like tall trees

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_12

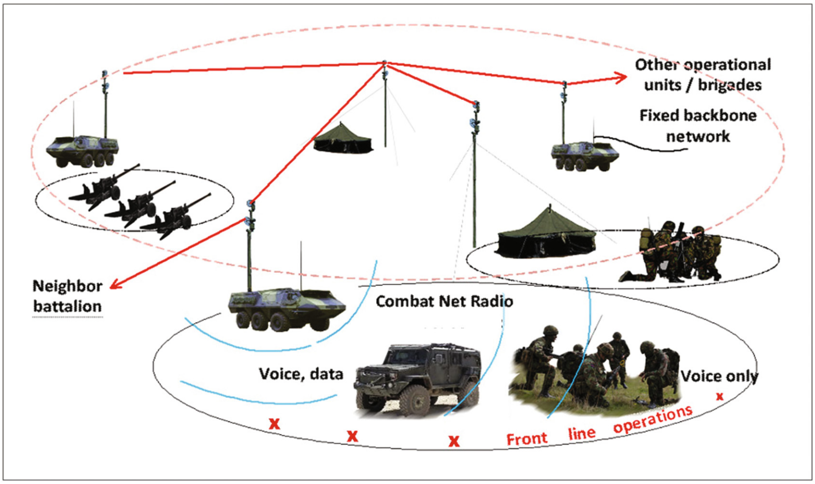


Fig. 1. Wireless communications for supporting military operations [3]

and buildings [4, 5]. Due to these limitations, this is why a ground mobile relay station is unable to be reliable wireless relay station in battlefields.

The aerial relay station can overcome many limitations of the ground relay station. Helicopters or satellites are equipped with wireless relay device [6, 7]. The aerial relay stations using helicopters are unaffected by ground conditions (inaccessible areas, tall trees, buildings, etc.). However, this is not an efficient method considering the cost effectiveness compared to ground vehicles [7, 8]. In addition, a helicopter is a more visible rotorcraft and can be a vulnerable and easy target for the enemy. The satellite communication system supports wireless communication by utilizing satellite signal, an advantage because communication can be supported without being exposed to the enemy. However, it is restricted in the extensive use and the access to satellite [6–8]. Due to its advanced technologies and equipment, few countries can afford to use a satellite to support wireless communications.

In this paper, we propose the concept of a drone relay station (DRS) to overcome the limitations of current mobile wireless relay stations in military operation. Research and application of drones are currently being applied not only in the military field but also commercially such as commercial delivery, medicine delivery, monitoring of infrastructures and power network assessments [9–11]. Various research such as swarm flight, wireless recharging, energy-harvesting flight path and a project, drone-carrying airborne mothership, is being conducted to use drones in the civilian/military field [12–15]. The military has been studying miniaturization and stealth to prevent exposure to the enemy [16]. Using these currently developed and undergoing technologies related to drones, we can operate DRS to support wireless communication in military operations.

This study is composed as follows. In Sect. 2, we discuss the problems of existing communication relay stations, and in Sect. 3, we describe the concept of DRS to overcome the existing problems and explore enabling technologies that have been developed and launched. Section 4 concludes this paper.

2 Limitations of Current Mobile Relay Stations

The military relies heavily on wireless communications when conducting operations in fields. To ensure and enhance the communications between units or a unit and a control station, relay stations are deployed along the units in operation areas using two types: ground relay station and aerial relay station.

2.1 Ground Vehicle Relay Station

Ground vehicles, such as wheel vehicles and track vehicles equipped with wireless relay devices, are used as wireless relay stations. As shown in Fig. 2, ground vehicles are maneuvered along with ground units into operating areas. Also, tactical robots (unmanned ground track vehicles) can be used as a mobile relay system [19]. However, these ground vehicles, regardless of the type, have limitations to conduct a wireless relay mission successfully. First, ground vehicles cannot access some areas due to ground conditions such as an iced road or no passing road, even track vehicles cannot easily move on an iced road. Second, ground vehicles can be easily detected by the enemy due to the size of the vehicles. While conducting as a relay station, antennas and other devices are equipped with the vehicles or deployed, surrounding the vehicles. Even though these devices and the color of the vehicles are camouflaged or concealed to avoid detections by the enemy, the size of the vehicles cannot be hidden. Third, the environment of operating areas affects the quality of wireless communication relay by blocking or interfering wireless signals. Lastly, to deploy or redeploy the devices and vehicles take a long time [3]. The relay vehicles must keep up with moving front line units; otherwise, it can cause a delay in a real-time relay of wireless communications.

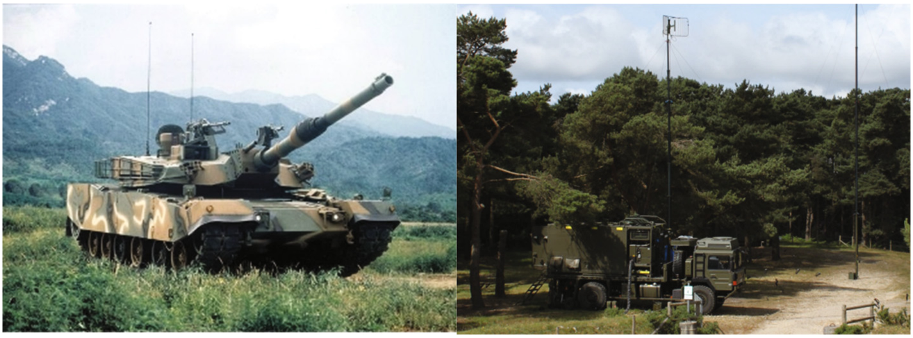


Fig. 2. Examples of ground mobile relay station [17, 18]

2.2 Aerial Vehicle Relay Station

To overcome some limitations of ground relay vehicles, aerial vehicles such as helicopters, air balloons and satellites (as shown in Fig. 3) are also used to link the wireless communications for military operations. Using aerial vehicles, the military can achieve a high mobility of mobile relay stations to support the wireless communications of the front-line units. Through these methods, they can avoid delays or any interference encountered when using ground methods. However, the use of helicopters or air balloons also has limitations such as being more vulnerable to the enemy attack due to larger sizes or louder noises than ground vehicles. The satellite requires a high-performance receiver to receive a relatively weak signal and is not available to relay wireless signals at any time and to any place [7, 22].

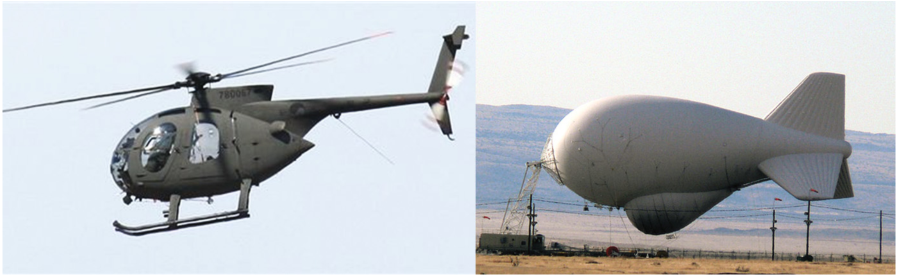


Fig. 3. Examples of aerial vehicle relay station [20, 21]

3 A Concept and Features of Drone Relay Stations

This paper proposes a concept of wireless relay stations using drones for military operations to resolve the limitations of current relay methods. As shown in Fig. 4, we use multiple drones to link wireless communications between a command center and a unit, command centers, or units. Multiple drones equipped with wireless communication relay devices are assigned to each echelon. These assigned drones are launched and form

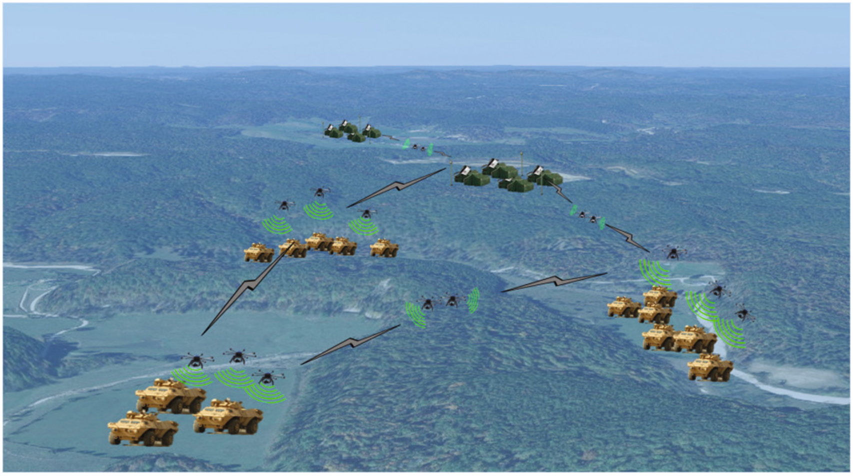


Fig. 4. A concept of drone relay stations (DRS)

aerial relay stations (aerial relay nodes) over operation areas following ground echelons. Without considering the locations of echelons and command centers, surrounding environments and deployment times (launch times), multiple drones can provide ground units with wireless communications.

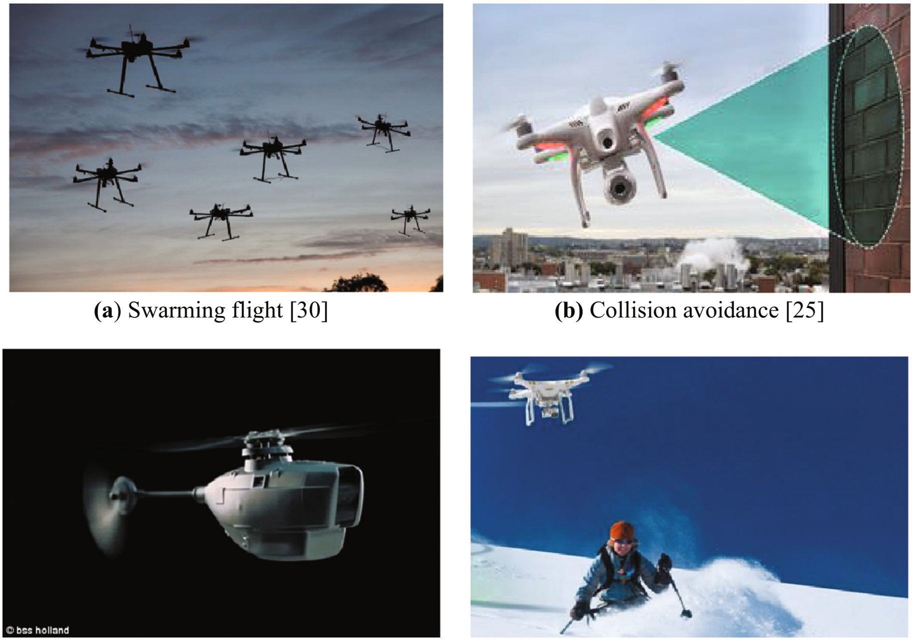
The concept of drone relay stations for providing wireless communications to the military is possible considering features of the drone that are currently developed (or developing) technologies as follows:

3.1 High-Mobility

Drones can fly aerial areas so the flight path of a drone is not affected by ground conditions. The drone also needs less deploying (launching) time than the current methods (ground vehicles and aerial vehicles). The drone can quickly respond to a request for an aerial relay station.

3.2 Swarming Flight

To ensure quality and coverage range of wireless links, drones must fly at a constant distance and/or a formation with other drones and ground vehicles. Drones act as a wireless linked node in aerial relay stations. Hence, swarming flight technology is needed to make a perfect drone flight distance and/or formation [23]. Recently, the US



(a) Swarming flight [30]

(b) Collision avoidance [25]

(c) Black hornet drone [16]

(d) Follow me drone technology [31]

Fig. 5. Developed/developing drone technologies

military tested a swarming flight of drones in California where the drones demonstrated advanced swarm technologies [24] (see Fig. 5(a)).

3.3 Collision Avoidance

Drones can collide with unexpected obstacles when they fly as they are following ground units. They may encounter unexpected barriers such as other drones, tall trees, tall buildings or enemy attacks. In these cases, to ensure real-time wireless relay services for ground units, the drones can detect the unexpected event and avoid them before colliding or hitting said obstacle. Many collision avoidance researches have been conducted for multiple drones using various approaches [25] (see Fig. 5(b)).

3.4 Detection Avoidance

The size of a drone is relatively small compared to the current methods (of ground and aerial vehicles), reducing the probability of exposure to the enemy; but this mini size cannot ensure safety due to noise from its drone rotor(s) [16]. For example, in special forces operations, covertness (without noise) is a critical factor to guarantee the success of the special operation. The Black Hornet Nano, a micro drone, was developed by Prox Dynamics for the military [16] (see Fig. 5(c)).

3.5 Flight Coordinating with Moving Ground Units (Vehicles)

Ground units continue movement corresponding to a progress of operations and timeline. Hence, drones must fly over operation areas keeping a standoff distance to ensure successful wireless relay services. This feature utilizes tracking technology, which has been explored by many researchers [26, 27] (see Fig. 5(d)).

3.6 Long Operating Time

The duration of military operations is varied based on types of operations. It can be less than an hour or longer than a day. Regardless of the length of operation times, drones should be aerial wireless relay stations during the operation duration. Although fuel cells currently have limitations, fuel cells provide drones with longer flight durations compared to the performance of batteries. There are many works that show the use of fuel cells for smaller drones [28, 29].

3.7 Resistance to Harsh Weather Conditions

Drone flights are likely to be affected by weather conditions. First, strong winds cause drones to consume more fuel to keep a flight stable or even making it impossible to fly through strong winds. Second, heavy rain or snow can also interfere with drone flight, and lastly, significantly cold temperatures may result in a reduction in the performance of drone's batteries [32]. Hence, resistance to the harsh weather factors is one of the features to ensure wireless communication relay in the real-time and real world. Recently, however, a test showed how well a drone could fly in 15 m per second level wind [33].

4 Conclusion

This paper proposed the concept of drone relay stations for wireless communication relay during military operations. To resolve the current wireless relay method limitations such as vulnerability to road conditions, lower mobility and longer deployment time, drones were suggested. This concept will be made possible by developed and/or developing technologies for drones.

References

1. Joint Communication System (2015). http://www.dtic.mil/doctrine/new_pubs/jp6_0.pdf
2. U.S. Department of the Army: Communications in a "Come as You Are" War, FM 24-12. Department of the Army, Washington, D.C. (1990)
3. EB Defense Newsletter December 2013. http://www.bittium.com/newsletters/defense
4. U.S. Department of the Army: Tactical Single-Channel Radio Communications Techniques, FM 24-18. Department of the Army, Washington, D.C. (1987)
5. Kozono, S., Watanabe, K.: Influence of environmental buildings on UHF land mobile radio propagation. IEEE Trans. Commun. **25**(10), 1133–1143 (1977)
6. Bonds, T., Mattock, M.G., Hamilton, T., Rhodes, C., Scheiern, M.: Employing Commercial Satellite Communications: Wideband Investment Options for the Department of Defense. Rand Corp., Santa Monica (2000)
7. U.S. Department of the Army: Space Support to Army Operations, FM 100-18. Department of the Army, Washington, D.C. (1995)
8. National Defense Magazine. http://www.nationaldefensemagazine.org/archive/2014
9. Thiels, C.A., Aho, J.M., Zietlow, S.P., Jenkins, D.H.: Use of unmanned aerial vehicles for medical product transport. Air Med. J. **34**(2), 104–108 (2015)
10. Cho, J., Lim, G., Biobaku, T., Kim, S.J., Parsaei, H.: Safety and security management with Unmanned Aerial Vehicle (UAV) in oil and gas industry. Procedia Manuf. **3**, 1343–1349 (2015)
11. Lim, G.J., Kim, S.J., Cho, J., Gong, Y., Khodaei, A.: Multi-UAV pre-positioning and routing for power network damage assessment. IEEE Trans. Smart Grid (2016). doi:10.1109/TSG.2016.2637408. (First online article)
12. Madey, A.G., Madey, G.R.: Design and evaluation of UAV swarm command and control strategies. In: Proceedings of the Agent-Directed Simulation Symposium, p. 7. Society for Computer Simulation International (2013)
13. Wang, C., Ma, Z.: Design of wireless power transfer device for UAV. In: 2016 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 2449–2454. IEEE (2016)
14. Bonnin, V., Bénard, E., Moschetta, J.-M., Toomer, C.A.: Energy-harvesting mechanisms for UAV flight by dynamic soaring. Int. J. Micro Air Veh. **7**(3), 213–229 (2015)
15. Hoerr, E., Mahan, D., Vallejo, A., Driscoll, J.: Multiple UAV coordination (2014). http://cegt201.bradley.edu/projects/proj2015/ballmer/Team%20Ballmer%20Project%20Proposal.pdf. Accessed 29 May 2017

16. Nano UAS PD-100 Black Hornet. http://www.bssholland.com/product
17. Tanks. http://www.military-today.com/tanks/k1a1.jpg
18. UK Armed Forces Commentary: SDSR 2015: solving the problem of hollow army 2020, June 2015. http://ukarmedforcescommentary.blogspot.com
19. Pezeshkian, N., Nguyen, H.G., Burmeister, A.: Unmanned Ground Vehicle Radio Relay Deployment System for Non-line-of-sight Operations. Space and Naval Warfare Systems Center, San Diego (2007)
20. Helicopter aging: UH-1H. http://military.asiae.co.kr
21. Tethered aerostat radar system. https://en.wikipedia.org/wiki/Aerostat
22. Satellite Limitations. http://electriciantraining.tpub.com/14189/css
23. Ryan, A., Zennaro, M., Howell, A., Sengupta, R., Hedrick, J.K.: An overview of emerging results in cooperative UAV control. In: 43rd IEEE Conference on Decision and Control, CDC, 2004, vol. 1, pp. 602–607. IEEE (2004)
24. DOD successfully tests terrifying swarm of 104 micro-drones. https://arstechnica.com/information-technology/2017/01
25. Howard, C.: Panoptes enters consumer drone market with eBumper4, obstacle avoidance technology. http://www.intelligent-aerospace.com/articles/2015/07l
26. Kim, S., Oh, H., Tsourdos, A.: Nonlinear model predictive coordinated standoff tracking of a moving ground vehicle. J. Guid. Control Dyn. **36**(2), 557–566 (2013)
27. Yu, H., Meier, K., Argyle, M., Beard, R.W.: Cooperative path planning for target tracking in urban environments using unmanned air and ground vehicles. IEEE/ASME Trans. Mechatron. **20**(2), 541–552 (2015)
28. Kumar, G., Sepat, S., Bansal, S.: Review paper of solar-powered UAV. Int. J. Sci. Eng. Res. **6**(2), 41–44 (2015)
29. Canis, B.: Unmanned Aircraft Systems (UAS): Commercial Outlook for a New Industry. Congressional Research Service, Washington, D.C. (2015)
30. Russon, M.-A.: Google robot army and military drone swarms: UAVs may replace people in the theatre of war. http://www.ibtimes.co.uk/google-robot-army-military-drone-swarms-uavs-may-replace-people-theatre-war-1496615
31. Corrigan, F.: 12 best follow me drones and follow me technology reviewed. https://www.dronezon.com/drone-reviews
32. Jaguemont, J., Boulon, L., Dubé, Y.: A comprehensive review of lithium-ion batteries used in hybrid and electric vehicles at cold temperatures. Appl. Energy **164**, 99–114 (2016)
33. Drone wind-resistance test. https://www.youtube.com/watch?v=CmbuuZdc5_s

Human-Systems Integration Challenges in Resilient Multi-UAV Operation

Edwin Ordoukhanian(✉) and Azad M. Madni

Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA
{ordoukha,azad.madni}@usc.edu

Abstract. Operating multiple unmanned aerial vehicles (UAVs) simultaneously gives superior mission coverage but creates human-systems integration challenges as well as joint human-system performance challenges. For example, when disrupting events happen, one can use multiple resilience techniques to recover from them. Some require human-in-the-loop response while others require humans to act in a backup capacity. Often these recovery mechanisms require context switching – something that humans are not good at because they have cognitive limitations. This paper examines multi-UAV operations from the perspective of different human roles and identifies challenges that exist when human operators are in the loop and have to adapt and context switch in response to the system's adaptive behavior.

Keywords: Resilient systems · Human-systems integration · Multi-UAV systems

1 Introduction

Unmanned Aerial Vehicles (UAVs) are used in a variety of missions such as search and rescue, reconnaissance and surveillance, law enforcement, agriculture, and payload delivery [1, 2]. It is customary today for multiple UAVs to be simultaneously deployed to achieve superior mission coverage [1, 2]. Multi-UAVs operate in open environments; in which they are susceptible to disruptions of various kinds. Uncertainties and unexpected factors in the environment disrupt multi-UAV operation and adversely impact overall system performance [1, 2].

The multi-UAV complex is essentially a system-of-systems (SoS), and therefore can exploit the methods used to model and analyze a SoS [2]. In particular, SoS methods can be used to analyze system flexibility, adaptability and resilience in the face of disruptions. A system-of-systems (SoS) with the requisite flexibility, adaptability, and resilience should be able to deal with disruptions, and continue to operate at acceptable performance levels in dynamic environments [1–4].

Disruptions can be conveniently categorized into three types [4]. *External disruptions* are associated with environmental obstacles and incidents that happen outside of system's boundary [4]. They are often random with unknown severity and duration. *Systemic disruptions* happen when internal component's functionality, capability or capacity causes performance degradation [4]. They are easily detectable in technological

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_13

systems. *Human-triggered disruptions* are associated with human operators inside or outside of the system boundary impact system performance. In general, these disruptions can be predictable or random [4].

This paper discusses the key human-systems integration challenges in resilient multi-UAV systems-of-systems design. The paper is organized as follows. Section 2 discusses single and multi-UAV operations and different command and control typologies. Section 3 discusses human-UAV interactions and different control and management schemes. Section 4 discusses the challenges that exist when brining human into the loop. Section 5 summaries the paper and recommend future research direction.

2 Multi-UAV Operations

Multi-UAV SoS operations pose an unprecedented HSI challenge. First, UAV operations require adaptable responses to disruptions – the cornerstone of a resilient SoS, Second, the multi-faceted role of the human (i.e. system operator, supervisor, agent) imposes cognitive demands arising from frequent context switching and multi-tasking. The solution is not simply having sufficient training for the human. [5–7] There are other considerations such as situational awareness and finite human cognitive capacity need to be considered [5–8].

Situational awareness is one of the key deriving factor of design process [8]. Since human operator is not physically on board, the operator's lack of situational awareness is a key factor to mission failure [9]. Often time, data connection between vehicle and the ground station includes a time delay (lag), jeopardizing the ability to give pilots the required information in timely fashion. As information come in from multiple sensors, data aggregation and presenting those data to the operator in timely fashion becomes a key challenge [9]. Advanced automation has reduced the role of traditional piloting skills and brought greater emphasis on monitoring and collaborative decision making between humans and machines [9, 10].

For example, in a search and rescue scenario multi-UAV system-of-systems operates in an open environment to search an area for missing personnel, and then communicates the coordinates to the ground station [1]. The multi-UAV system-of-system faces many external disruptions such as extreme weather and wind gusts, and physical obstacle. In this case, the system-of-system must reallocate tasks and functions to different vehicles to deal with these disruptions. Vehicles send their status to the ground station continuously for monitoring purposes [1]. Figure 1 depicts multi-UAV SoS for search and rescue scenario.

In general, there are three command and control (C2) typologies. Human-centered control, decentralized control with human as supervisor, and fully autonomous command and control. Each of these C^2 are discussed next [1, 2, 9].

In a human-centered control, human is the central part of the system and gives all the commands and controls. In this case system has less autonomy and this is a relatively big disadvantage for long missions or when the area is big. This is mostly because of humans suffering from loss of attention when monitoring displays for long period of time [1, 2, 9, 11].

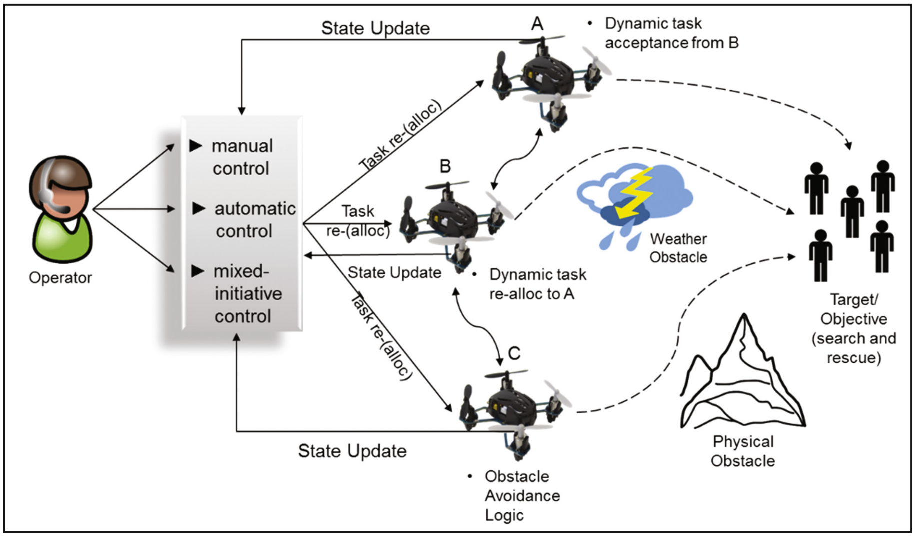


Fig. 1. Multi-UAV SoS rescue mission example [1]

In a decentralized control with human as a supervisor each vehicle has control over its own mission and share data with its neighbor vehicles. This requires some basic level autonomy. Human operator is brought into the loop in case of emergency. In the decentralized control typology, the system aims to defer human intervention as much as possible and let the systems operate autonomously. The system can employ different resilience approaches such as function reallocation and functional redundancy, physical redundancy at both system level (i.e. having two similar components within a vehicle) as well as system-of-system level (i.e. having two or more similar vehicles in the system-of-system to perform the same functionality [1, 2, 9, 11].

In fully autonomous command and control, human has almost no role in the system. System distributes the task and operates autonomously. This put extra emphasizes on attributes such as resilience and self-regulation. This is usually costly and poses several risks since system validation and verification requires substantial effort. It is mostly suitable for long missions and requires no personnel training, which reduces operations cost [1, 2, 9, 11].

3 Human-UAV Interaction

UAVs today are capable of performing multiple tasks without exposing humans to harm [9]. However, human operators are still needed for supervisory control. Human supervisory control (HSC) is essentially the shift from lower-level skill-based behaviors to higher-level knowledge-based behaviors [9]. The lower level tasks require non-stop attention and require multiple simultaneous actions. Hence, these tasks can be delegated to automation.

In HSC, human operator interacts with a computer, receives feedback from and provides commands [9]. Figure 2 depicts human supervisory control. Both civilian and military applications of UAVs require supervisory control for complex operations. These operations can be border patrol, agriculture monitoring, package delivery, or disaster response. UAVs may require human guidance to varying degrees. The combination of human and UAVs essentially can be defined as Unmanned Aerial System (UAS) [1, 2, 9].

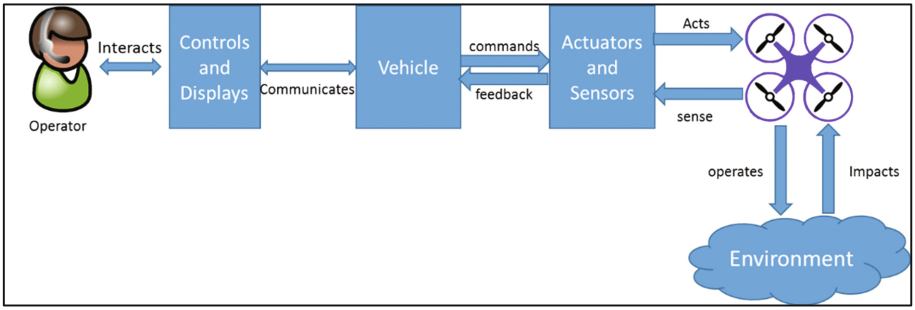


Fig. 2. Human supervisory control (adapted from [9])

HSC in single and multi-UAV operation is hierarchical process and has four major loops: (a) control; (b) navigation; (c) mission and payload management; and (d) system health and status monitoring. In multi-UAV control mission and payload management and system health and status monitoring loops are shared among all UAVs participating the SoS. Each of these loops are discussed next in details [1, 2, 9].

Control: This step consists of pilot (if remotely piloted) or autopilot (if autonomous) and flight controls. This loop is the most critical loop that obeys physical laws of nature such as aerodynamics constraints. In this loop pilots or autopilot are focused only on the short term and local control (i.e. keeping the vehicle stable) [1, 2, 9].

Navigation is the step that an agent (human or autopilot) must execute to meet mission constrains such as rendezvous, passing through certain waypoints or targets, or avoiding obstacles [1, 2, 9].

Mission and payload management: represents the highest level of control. On this level, sensors must monitor and decisions must be made based on the incoming information to meet overall mission requirements. This is where humans play a key role since this loop requires knowledge-based reasoning (e.g. judgement, experience, and abstract reasoning) that cannot be performed by automation [1, 2, 9].

System health and status monitoring is continual supervision performed by either human or automation, or both to ensure that systems are operating within normal limits [1, 2, 9].

From human-in-the-loop perspective, if control loop fails other loops fail as well. If humans are required to do most of the tasks in 4 loops depicted in Fig. 3 manually, then the design and operation cost grows exponentially. Performing multiple tasks simultaneously is also hard due to human cognitive limitations [1, 2, 9].

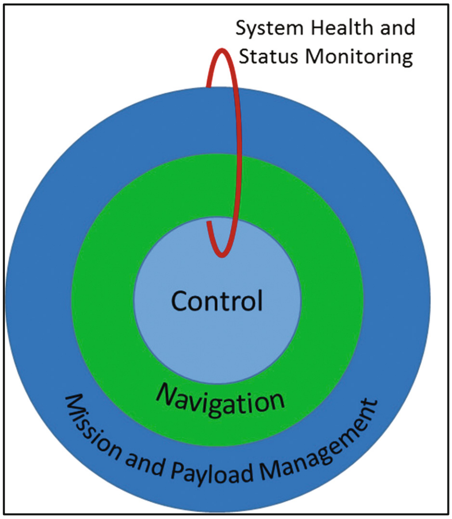


Fig. 3. HSC hierarchical loops (adapted from [9])

In controlling multiple-UAVs human perform tasks related to overall mission and payload management and delegate navigation and motion control tasks to automation [2, 9]. However, if due to disruptions any of these functions are interrupted, then human operator must intervene and assist the system [1, 2]. This is where human operator switch contexts and changes its role. In this situation, human operator must perform navigation and control tasks on top of mission and payload management tasks. If there are multiple vehicles, then operator must perform these tasks for multiple vehicles which increases the probability of human operator going under cognitive overload.

In human-UAV interactions, there are two management schemes: management by consent and management by exception [9]. In management by consent human operator must approve an automated solution before execution. In management by exception the automation gives the operator a period of time to reject the solution. Depending on the number of UAVs participating in the SoS, different management schemes can be used. For example, Ruff et al. [12] showed that management by consent provided best situation awareness ratings, the best performance scores, and the most trust for controlling up to four UAVs [9]. On the other hand, one might think that delegating most of the work to automation and ask human operator to reject an action proposed by the system would work better if number of participating UAVs increases. However, studies show that in this case, operators tend to not reject system's action since they are not engaged enough to know what is happening in the system [13].

4 Challenges of Human-in-the Loop Resilience Approach

Human-in-the-loop and human-as-a-backup are both means to introduce resilience within SoS that have to deal with disruptions [1, 2, 4]. During normal operation, multi-UAV SoS operate mostly autonomously. When disruptions occur, humans can intervene at any level in multi-UAV operations and interact with multiple autonomous or semi-autonomous UAVs.

The key challenge lies in the human's inability to adapt quickly to the situation. During normal operation, human operator plays the role of an observing agent, however, when dealing with a disruption, the human has to play the role of a command agent or simply an agent performing a specific task to satisfy a mission requirement [9]. This transition between human roles within a SoS has significant impact on system performance, and the ability of the system in dealing with disrupting events [5–8].

Due to human cognitive limitations, humans are poor at monitoring systems for extended time periods [5–8]. Yerkes-Dobson law states that human vigilance tends to drop when monitoring for low probability event. However, often time the consequence of these low probability events can be devastating.

Human operators cannot multi-task well [5–8]. They are unable to recall all the information most of the time when multi-tasking. Their task performance decreases significantly when they fail to pay full attention to each task. This consideration is important when humans are brought into the loop to deal with disruptions. With attention is divided on multiple tasks, the likelihood of human operators being able to help a system recover from disruptions decreases [5–8].

Human operators adaptivity rate varies with information channels, language, command and control patterns, and architectural configurations [5–8]. There is a trade-off between adaptation rate and the likelihood of human error. If human operator adapts fast, the likelihood of error increases [5–8].

The challenges can be divided into the two main categories. First, the challenges that are associated with human operator limitations and second the challenges that are associated with the methodologies and processes that can be used to bring the human operator into the equation to restore smooth operation of the system.

The main challenge is to keep the human engaged with the system enough amount of time so when the time comes for the human to change roles, the human has enough knowledge about the system to be effective in achieving desired system behavior. The main problem is that if humans are engaged all the time, they will go under cognitive overload and won't be able to perform well when the time comes. On the other hand, when the human operator is not engaged enough then when brought into the system, the operator will know little about the system [14–16].

Another challenge involves the process by which the human operator should be introduced into the system. If the human is put into a situation with no knowledge of the system's current status, the human may end up taking actions that could further harm the system. On the other hand, if the human operator is given all the necessary information about current system status simultaneously, the human will experience cognitive overload, lose track of the system state, and be prone to errors. If human is given the information systematically, then the challenge is what are the key information that human operator should know first before it takes some action. The challenge is to identify high priority information that has to be communicated instantaneously to the human to ensure maximum situational awareness on the part of the human to deal with disruptions [14–16].

5 Conclusion

When disrupting events that require human intervention occur, human-system integration considerations and human-UAV interaction become central to joint human-system performance [15, 16]. In this paper, we have identified multiple challenges that need to be addressed when human operators have to be brought into the loop. What is needed is a systematic way to engage humans, and bring them into the loop with full contextual awareness to successfully deal with the disruption. Recent advances in Model Based Systems Engineering are beginning to address this challenge [17, 18].

References

1. Ordoukhanian, E., Madni, A.M.: Toward development of resilient multi-UAV system-of-systems. In: AIAA SPACE 2016, vol. 5414 (2016)
2. Ordoukhanian, E., Madni, A.M.: Introducing resilience into multi-UAV SoS. In: Proceedings of Conference on Systems Engineering Research, Redondo Beach, CA, March 2017
3. Neches, R., Madni, A.M.: Towards affordably adaptable and effective systems. Syst. Eng. **16**(2), 224–234 (2013)
4. Madni, A.M., Jackson, S.: Towards a conceptual framework for resilience engineering. IEEE Syst. J. **3**(2), 181–191 (2009)
5. Madni, A.M.: Integrating humans with and within software and systems: challenges and opportunities (Invited Paper). CrossTalk J. Def. Softw. Eng. (2011)
6. Madni, A.M.: Towards a generalizable aiding-training continuum for human performance enhancement. Syst. Eng. **14**(2), 129–140 (2011)
7. Madni, A.M.: Integrating humans with software and systems: technical challenges and a research agenda. Syst. Eng. **13**(3), 232–245 (2010)
8. Lyman, J., Madni, A.M.: Operator roles in robotics. Robot. Age (1984)
9. Goodrich, M.A, Cummings, M.L.: Human factors perspective on next generation unmanned aerial systems. In: Handbook of Unmanned Aerial Vehicles, pp. 2405–2423. Springer, Netherlands (2015)
10. Madni, A.M., Freedy, A.: Intelligent interfaces for human control of advanced automation and smart systems. In: Zeidner, J. (ed.) Human Productivity Enhancement, Training and Human Factors in Systems Design, vol. 1, pp. 318–331. Praeger, Santa Barbara (1986)
11. Freedy, A., Madni, A., Samet, M.: Adaptive user models: methodology and applications. In: Rouse, W.B. (ed.) Man-Computer Systems, vol. 2, pp. 249–293. Jai Press Inc., Greenwich (1985)
12. Ruff, H.A., Narayanan, S., Draper, M.H.: Human interaction with levels of automation and decision aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. Presence **11**, 335–351 (2002)

13. Mosier, K.L., Skitka, L.J.: Human decision makers and automated decision aids: made for each other? In: Parasuraman, R., Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, Human Factors in Transportation, pp. 201–220. Lawrence Erlbaum, Mahwah (1996)
14. Madni, A.M., Sage, A., Madni, C.C.: Infusion of cognitive engineering into systems engineering processes and practices. In: Proceedings of the 2005 IEEE International Conference on Systems, Man, and Cybernetics, Hawaii, 10–12 October 2005
15. Madni, A.M.: HUMANE: a knowledge-based simulation environment for human-machine function allocation. In: Proceedings of IEEE National Aerospace and Electronics Conference, Dayton, Ohio, May, 1988
16. Madni, A.M.: HUMANE: a designer's assistant for modeling and evaluating function allocation options. In: Proceedings of Ergonomics of Advanced Manufacturing and Automated Systems Conference, Louisville, KY, pp. 291–302, 16–18 August 1988
17. Madni, A.M., Ross, A.: Exploring concept trade-offs. In: Parnell, G. (ed.) Trade-off Analytics. Wiley, Hoboken (2016)
18. Madni, A.M., Sievers, M.: Model-based systems engineering: motivation, current status, and needed advances. In: Proceedings of Conference on Systems Engineering Research, Redondo Beach, CA, March 2017

Human Factors in Military Maritime and Expeditionary Settings: Opportunity for Autonomous Systems?

Jacob N. Norris(✉)

Space and Naval Warfare Center Pacific, 53560 Hull Street, San Diego, CA 92152, USA
jacob.norris@navy.mil

Abstract. In military settings, the maritime and expeditionary environments present unique operational, environmental, and logistical considerations for the human which are not present within land-based and established military locations (i.e. training or garrison settings). Advances in autonomous and unmanned systems may play key roles in addressing these issues. This paper will: briefly provide an overview of some human factors considerations in the maritime and expeditionary environment; address how autonomous systems may help overcome human factors challenges; and discuss how human-machine teaming and science & technology informed policy may be effective means to overcome some of autonomy's unintended risks (i.e. accidental escalation) presented to military leaders and facilitate stability.

Keywords: Defense · Military · Human-machine teaming · Maritime · Expeditionary · Autonomous systems · Policy · Humanitarian relief

1 Introduction

The defense establishment's response to the ever-changing global environment has been to turn to science & technology (S&T) and S&T informed policy. In the past, spread of nuclear technology to the Soviet Union, led to the doctrine of Mutually Assured Destruction (MAD). As an offset strategy[1], nuclear arms were intended to mitigate any conventional weapons advantage held by the Soviet Union. Their use was not necessarily intended to reduce risk to US personnel, but rather their value was as a deterrent. They were intended to reduce risk to US personnel and US civilian populations, by dissuading Soviet actions. MAD was seen as helping to prevent any full-scale conflicts between the US and the Soviet Union [3, 4]. Paradoxically, a weapon whose existence was to deter aggression against conventional forces, increased risk of an accidental escalation against

[1] Nuclear weapons and precision-guided munitions have been recognized as the 1st and 2nd Offset strategies by DoD experts. Autonomous systems arranged with the human operator as a human-machine team has been characterized as the newest, 3rd Offset strategy [1, 2].

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_14

the populaces the military was protecting.[2] Another example of S&T helping to maintain an edge is the precision guided weapon. In the 1990s, precision guided weapons, allowed the US to project power and precision globally with relative safety to US personnel. They have proven to be highly effective [2].

Future S&T, like autonomous and unmanned systems, may play key roles in future maritime and expeditionary settings. There are those who ask compelling legal and ethical questions about autonomous and unmanned systems (for review see [5]), but the real concern needs to be centered on operational utility. Leadership must ask if autonomy changes the risk assessment. Does it improve a country's ability to meet growing challenges and provide assistance to other countries in this changing environment? Do autonomous systems improve the overall global situation and deter aggression from nontraditional actors like pirates or terrorists? This paper argues that advances in the area of the autonomous unmanned systems hold the potential to improve safety, improve decision-making, and maintain global stability in complex maritime and expeditionary settings, but autonomy must be adopted prudently in order to avoid negative effects.

One must understand: what constitutes an autonomous unmanned system; some human factors in the military maritime and expeditionary settings; arguments how use of autonomous systems may benefit operations in the maritime and expeditionary settings; and some counterarguments that risk to personnel may decrease at the tactical level, but autonomous systems may generate new strategic risks to militaries and their populaces. Finally, this paper will conclude with a brief discussion of how the human element, through human-machine teaming and thoughtful policy can preserve the benefit of autonomous unmanned systems in such a way that mitigates potential negatives.

2 Autonomous Systems

Militaries have increasingly incorporated robotics and unmanned vehicle systems in the last decade; small explosive ordnance disposal robots and large aerial assets for intelligence gathering being most prominent.[3] The vast majority are remotely operated rather

[2] In 1995 for a brief time, Russia thought it might be under attack from the US, and Russian President Yeltsin opened his nuclear briefcase for the first time in history (other than as part of an exercise). These incidents highlight the vulnerability to human error present in MAD doctrine and its corollaries [6].

[3] Approximately 8,000 unmanned ground vehicles of various types have seen action in Iraq and Afghanistan. As of September 2010, unmanned ground vehicles have been used in over 125,000 missions, including suspected object identification and route clearance, to locate and defuse improvised explosive devices (IEDs). During these counter-IED missions, Army, Navy, and USMC explosive ordnance teams detected and defeated over 11,000 IEDs using UGVs. In 2009, US DoD completed almost 500,000 UAS flight hours just in support of operations in Afghanistan and Iraq. In May 2010, unmanned systems surpassed one million flight hours and in November 2010 achieved one million combat hours. The use of unmanned maritime system is not new. After World War II, unmanned surface vessels (USVs) conducted minesweeping missions and tested the radioactivity of water after atomic bomb tests. During the Vietnam War in an area south of Saigon, remotely controlled USVs conducted minesweeping operations [8].

than truly autonomous. Military research and development has made slow, but consistent headway, transitioning unmanned platforms from pre-programmed and remotely-controlled systems towards autonomous functionality. Unmanned systems with elements of autonomy are beginning to appear in the military [4].

The foundation for autonomy springs from the capability of computer systems to perform tasks that traditionally require human intelligence, using a *sense-think-act* loop.[4] Intelligent systems aim to apply artificial intelligence to a particular problem or domain; the system is programmed or trained to operate within the bounds of a defined knowledge base. There are considered to be two categories of intelligent systems. There are systems that are disembodied or *autonomy at rest*. These include applications like data compilation, data analysis, web search, recommendation engines, forecasting, and cybersecurity. These are in part, driven by limitations of humans to rapidly process the vast amounts of data available today; disembodied autonomous systems are now required to find trends and analyze patterns within the timeframe needed by human-users [4]. The second type of autonomous systems are embodied or those employing *autonomy in motion* that have a presence in the physical world and include robotics and autonomous vehicles [4]. The more traditionally imagined intelligent system, embodied autonomy such as robots or autonomous vehicles typically possess additional kinds of sensors, actuators, and mobility.

3 Human Factors in Maritime and Expeditionary Settings

In military settings, the maritime and expeditionary environments present unique operational, environmental, and logistical considerations that are not present within land-based and established military locations. Despite many similarities which exist between maritime and deployed land based settings as well as many similarities between military and civilian maritime settings there remain unique elements in the military maritime and expeditionary environments (for an exhaustive review of civilian maritime human factors, see [7]). Core differences in mission areas between types of military forces lead to different types of challenges. Within the US military, the maritime and expeditionary domains are predominantly that of the Navy and Marine Corps.

These present both challenges and opportunities for the human, the machine, and the human-machine interaction. While there are many ways to frame military maritime and expeditionary settings from human factors perspective, this manuscript will use the Socio-technical system model [7]. This model provides terminology that is "expansive" enough to accommodate the political, legal, and tactical considerations present within defense maritime and expeditionary settings without being inconsistent with other frameworks. Within the socio-technical system all aspects of the maritime and expeditionary setting can be conceptualized as falling into or across one of seven domains.

[4] To be truly autonomous, a system must be able to compose independently and select among various courses of action to accomplish goals based on its knowledge and understanding of the environment, itself, and the situation [4]. From an engineering perspective, building true autonomy is a series of technological challenges akin to developing understanding of and then building a human-being from scratch.

These include: Individual, Group, Technology, Physical Environment, Organizational Environment, Society and Culture, and Practice.

The organizational scale, technological speed, and complexity of maritime and expeditionary operations influence execution of Command and Control (C2). One challenge of the military maritime and expeditionary setting is to successfully integrate large numbers of personnel and organizations tasked with complex missions using effective Command and Control (C2) processes. Integration into a larger complex mission is not something unfamiliar to land-based and established military locations, like bases or posts. However as military units go, few entities like a naval vessel exist that must operate with a unity of purpose beyond all others. One key difference between the civilian and military naval vessel is the combat information center (CIC). The CIC serves as the "nerve center" of the military naval vessel. The CIC during combat operations renders many decisions, becomes the defacto ship's bridge and is responsible for integrated defensive and offensive activities. On the individual level, personnel assigned to the CIC must maintain high proficiency and vigilance [9, 10]. On the broader social and organizational level, naval vessels must be capable of sustained integrated operations with many other vessels and platforms within an expeditionary strike group or carrier strike group; these consist of several thousand personnel spread across a half-dozen of more ships with accompanying squadrons and Marines elements. Finally, these operations must be coordinated, at times, away from land basing. Even relatively small navies may still participate in coordinated exercises or interdiction of smugglers with allies and partner nations.

As an expeditionary example of complex C2, an amphibious landing or ship-to-shore movement under hostile conditions is a large scale, fast-paced, complex operations requiring planning and execution. An amphibious assault involves the establishment of a landing force on a hostile or potentially hostile shore. Organic capabilities of amphibious forces, including fire support, logistics, and mobility must be integrated. Forcible entry operations can be accomplished through amphibious operations, airborne operations, air assault operations, or a combination of any or all of these forcible entry techniques. How these operations are conducted depends on the Combatant Commander's/ Joint Task Force Commander's assessment of maritime factors and mission, enemy, terrain and weather, troops and support available, time available, and whether to conduct the forcible entries as concurrent or integrated [11]. Presence of an enemy minefield may preclude or slow down operations. Clearing minefields is a hazardous and time-consuming task that may influence the commander's assessment. Adverse weather, while on land, may help with the tactical element of surprise, may lead to delays in amphibious operations. Poor operations by any one naval vessel (each of which is made up of integrated departments with non-overlapping functions) within the operation can impair the timing or execution of the overall affair.

The austere and variable environment impacts technology, culture, practice, and the individual. Poor proximity to maintenance depots and vendor-level maintenance compels simplicity and ruggedness in design. By situation and by definition, personnel in maritime and expeditionary settings respectively have limited access to maintenance and logistics. This also impacts the type of technology and practices that have traditionally been adopted. Naval forces may utilize older more mature technologies, like the "1MC" or ship-made damage control devices, which are easier to maintain and make when newer less tested alternatives may exist [9]. The US Marine Corps emphasizes portability, ruggedness in design, and redundancy for all platforms [11]. Marine Corps units are organized to be self-sustaining for periods of time without logistics tails. The variable austere environments have an impact on training and skills sustainment. For example, Navies, by tradition and experience, emphasize on-the-job training, better known as qualification. This stands in stark opposition to army traditions of "schoolhouses." Qualification as Officer of the Deck (OOD) is noted as the cornerstone accomplishment for a surface warfare officer [10]. The tradition of qualifying has endured due to the variable nature of the sea and time-tested lesson that proficiency in naval warfare is premised on technical acumen in ship handling, which can only be gained at sea. Expeditionary forces serve as the forces of readiness. This dictates that preparedness must be sustained at or near deployment levels at all times.

Finally, the physical operational environment and social environment creates physiological and psychological stress. Aboard modern ships, many personnel spend long hours staring at computer screens performing difficult mental work, such as monitoring equipment or communications with sometimes limited opportunity for physical exercise. Austere deployment environment also presents factors that impact the individual and organization. Environmental stressors like heavy seas and variable operational tempo (i.e. general quarters) in very dynamic situations are common. Organizationally, watch standing schedules and stimulation poor environments like those in submarines also present challenges. Some other prominent examples of physiological issues include motion sickness, motion induced fatigue, and combat & operational stress.

4 Benefits of Autonomous Systems in Maritime and Expeditionary Settings

Autonomy can reduce human workload and enables the optimization of the human role in the system. Human decision-making can focus on points where it is most needed or focus decision-making in scenarios where the human is fatigued, stressed, or inexperienced. Autonomy at rest may be able to provide integrated information to OODs and planners of operations more efficiently than previously allowed. Autonomy can also enable operations where C2 is limited or unachievable such as in caves, under water, or in areas with degraded communications as well as increase stand-off from hazardous or operationally risky areas [8, 12]. Unmanned systems equipped with autonomous capabilities are projected to enhance defense and law enforcement capacities; swarms of UAVs will be capable of operating in difficult environments over the horizon. Similarly, non-warfare applications for unmanned systems, such as humanitarian and disaster relief, support the military's commitment to partnership with allies and finding common ground with other countries. These include search and rescue, real-time data gathering to inform decision making, mapping, damage assessments, asset tracking, and creation of ad-hoc communication networks during large scale disasters [8, 12]. Some less apparent examples include: Chemical, Biological, Radiological, Nuclear (CBRN) Response like the Fukushima reactor incident; Protection of allies' economic infrastructure against non-state actors; automated cyber-response to hackers [4]; and in kinetic environments, automated monitoring and management of multiple patients during prolonged aero-medical evacuation (complex morphine and fluid resuscitation protocols).

From a defense perspective increasing use of autonomous systems should yield benefits by improving safety and performance at the tactical level. Safety improves by reducing the lethality of conflict, anti-terrorism, or anti-piracy operations. Leaders can adopt more daring tactics or missions because a system is unmanned. For example, some project that unmanned vehicles will be the foundation for current and future mine counter-measure (MCM) operations.[5] Performance improves through enhanced accuracy of systems with more endurance, range, and speed in comparison to manned vehicles. Advanced unmanned surface assets, like the Sea Hunter will be able to persist in the maritime setting longer than manned platforms and execute long duration missions.[6] Autonomy enables the execution of new missions, such as cyber and electronic warfare, in which decision speed beyond human capability is critical to success against criminal hackers [4].

5 Considerations for the Military Commander

The advantages of autonomy for the commander must be weighed against nuanced considerations and potential scenarios. Aside from the legal and ethical arguments, one must concern themselves with how autonomous systems may re-calibrate risk assessments for the military commander at the strategic as well as tactical level. Poorly considered use of autonomous unmanned systems may lead to escalation of conflict and proliferation of destabilizing autonomous systems.

Governments currently state no intent to field lethal autonomous systems, but availability of technologies or their tangible development may change that. It is argued that

[5] One example of a rapidly emerging area for autonomous systems to benefit the military is the mine countermeasures (MCM) domain. Current manned and unmanned MCM platforms all require personnel in the minefield. MCM-1 class ships can detect, classify, and neutralize all known types of mines, but require large manned crews. Increased utilization of autonomy-enabled UUVs can significantly reduce personnel risk during MCM operations. Personnel can conduct MCM operations remotely rather than entering the minefield. The UUV program has demonstrated significant progress in utilizing UUVs for MCM. Further gains are possible. Development of both autonomy in motion and autonomy at rest will reduce tactical timelines with intensive operator involvement. Increased autonomy in the areas of automated target recognition and mine disposal could reduce risk and decrease the tactical timeline [4].

[6] Another example is the Anti-Submarine Warfare (ASW) Continuous Trail Unmanned Vessel (ACTUV). This is an unmanned vessel optimized to track quiet diesel electric submarines at a fraction of their size and cost. This system will be able independently deploy on missions spanning thousands of kilometers of range and months of endurance under a sparse remote supervisory control model. This includes autonomous compliance with maritime laws and conventions for safe navigation, autonomous system management for operational reliability, and autonomous interactions with an intelligent adversary. While the ACTUV program is focused on demonstrating the ASW tracking capability in this configuration, the core platform and autonomy technologies will be broadly extendable to a wide range of missions and configurations for future unmanned naval vessels [13, 14].

lethal autonomous systems would lessen constraints on war surrounding death and killing [5]. One could extend this argument to include not just lethal autonomous systems but to autonomous systems being placed in "harms" way in lieu of personnel. It is easy to assess that unmanned underwater vehicles and surface vehicles will reduce harm to personnel and expand capacity in anti-terrorism, anti-piracy, and disaster relief. However, they may indirectly increase probability of tension and conflict between potential adversaries because of deployment in hostile or tension riddled areas. Actors with competing national or economic interests may be more likely to shoot at each other's autonomous systems in scenarios where they might not shoot a manned system. A country might react aggressively towards unmanned surface vehicles conducting routine navigation operations, in international waters, but near to territorial waters. They could make the argument that no one is being killed when an unmanned vehicle is destroyed. Similarly, adversaries might be more likely to conduct escalatory activities against each other if they both possessed unmanned or autonomous unmanned systems. For example, a rogue actor might be emboldened to conduct mine-laying operations or aggressive low-flying intelligence gathering activities if they believe that incursions by unmanned systems won't be treated as acts of war. Finally, as a decision aid, machine autonomy must be balanced with elements that are difficult to model. Autonomy at rest, may render guidance or a firing solution that produces tactical advantage and minimizes risk to friendly personnel, but doesn't account for the nuances of human feeling or project out to the ramifications of a "cold-hearted" act during peace talks or stability operations.

Unlike atomic weapons and precision guided weapons, robotics and autonomy research has flourished in academic and corporate realms; technological advances by one country will be harder to maintain as computing costs continue to go down. An arms race between many countries should be expected.[7] Intelligence assessments highlight advances in artificial intelligence and data science are threats to national infrastructures and are improving globally in both the commercial and military sectors. Many actors are developing autonomous and cyberattack systems. Advances made by one country will eventually be met and countered [15]. Countries may not improve their situations by pushing wide-scale introduction of autonomous systems. Other countries or their populaces may view this as "upping the ante," and use the opportunity to assert their priorities visa vie sabre rattling. Autonomy may produce scenarios akin to MAD, but without potential for widespread death; the strategic calculation is drastically different. Adversaries may unleash autonomous cyber-attack systems on each other's power grids.

[7] In contrast to the U.S., China, South Korea, Japan, and India are investing heavily in higher education and research. India and China are systematically luring back their scientists and engineers after they are trained in the U.S. This contrast in investment is evident in the specific areas related to robotics and manufacturing. Korea has been investing $100M per year for 10 years (2002–2012) into robotics research and education as part of their 21st Century Frontier Program. European Commission investments total over $1.5 billion into robotics and cognitive systems, and manufacturing. Japan is investing $350M over the next 10 years in humanoid robotics, service robotics, and intelligent environments. In 2016, Japan has also announced a major push to become leader in robotics with a 5 year investment of $1B for industrial robotics. The non-defense U.S. federal investment in robotics and automation is small by most measures compared to these investments [12].

Attacks while crippling, may be deemed allowable because they don't expressly kill. Autonomous systems fielded by competing countries, with simple rule-based decision-making algorithms could interfere with maritime commerce at relatively low-cost. Unlike remotely controlled vehicles, fully autonomous vehicles would not require communications links to operate and therefore could act analogously to a swarm of cockroaches.

6 Conclusions

Advances in the area of autonomous unmanned systems hold the potential to improve safety, improve decision-making, and maintain global stability in complex maritime and expeditionary settings, but autonomy must be adopted prudently in order to avoid negative effects. With autonomous unmanned systems, military commanders will be able to more effectively execute a wide range of mission across vast distances with reduced manpower. Autonomy could potentially give the military commander stand-off distance and precision. Autonomy may enable countries to conduct more daring missions and keep personnel safe. However, autonomy may also produce scenarios that lead to reckless escalation because human risk has been reduced; a cornerstone of previous deterrence approaches like MAD. Tactical advantages in safety may produce strategic dilemmas. In many ways, the strategic-tactical risk tradeoff isn't new; Marines have called this the "strategic corporal problem."[8] A framework does exist to think about these issues. As part of prudent implement, military leaders need to push for policy and technical approaches that mitigate potential strategic risks. Future work should center on human-in/on-the-loop designs, operator informed rapid retraining of machine learning based algorithms, explainable autonomy, improvements in Live, Virtual, and Constructive (LVC) Training, policy development for use of autonomy across the spectrum of military operations, social-technical integration, and interagency efforts to develop strong international roadmaps for use of autonomy.

Acknowledgements. Approved for public release: distribution unlimited. The contents of this paper reflect the personal views of the author and are not necessarily endorsed by the SPAWAR Systems Center Pacific, Department of the Navy, or the US Government. The author would like to thank Gregory Lafave for feedback on an earlier version of this manuscript.

[8] The strategic corporal concept was coined in 1999 by USMC General Charles Krulak but popularized during OEF/OIF and refers to the concept that low-level operational personnel must be capable of not just thinking through immediate tactical situations, but must also be capable of anticipating the operational and strategic consequences.

References

1. Work, R.: Deputy Secretary of Defense Speech: The Third U.S. Offset Strategy and Its Implications for Partners and Allies. Willard Hotel, Washington, DC, 28 January 2015. http://www.defense.gov/News/Speeches/Speech-View/Article/606641/the-third-us-offset-strategy-and-its-implications-for-partners-and-allies. Accessed 26 Nov 2016
2. O'Hanlon, M., Petraeus, D.: America's awesome military: and how to make it even better. Foreign Aff. **95**, 10–17 (2016)
3. Parrington, A.: Mutually assured destruction revisited: strategic doctrine in question. (1997) http://www.airpower.maxwell.af.mil/airchronicles/apj/apj97/win97/parrin.html. Accessed 26 Nov 2016
4. Fields, C.: Report of the Defense Science Board Summer Study on Autonomy. Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, Washington, DC (2016)
5. Pilling, M.: Issues Regarding the Future Application of Autonomous Systems to Command and Control (C2). DSTO Defence Science and Technology Organisation, Australian Department of Defence (2015)
6. Frankel, M.J., Scouras, J., Ullrich, G.W.: The Uncertain Consequences of Nuclear Weapons Use. The Johns Hopkins University Applied Physics Laboratory LLC, Laurel (2015)
7. Grech, M., Horberry, T., Koester, T.: Human Factors in the Maritime Domain. CRC Press, Taylor & Francis Group , Boca Raton (2008)
8. US Department of Defense: Unmanned Systems Integrated Roadmap FY2011-2036, Washington, DC (2011)
9. Cutler, T.J.: Bluejackets Manual, 24th edn. US Naval Institute, Annapolis (2009)
10. Stavridis, J., Girrier, R.: Watch Officers Guide, 15th edn. US Naval Institute, Annapolis (2007)
11. Fleet Marine Force Qualified Officer Study Guide. http://www.imef.marines.mil/Portals/68/Docs/IMEF/Surgeon/FMFQO_Study_Guide.pdf. Accessed 8 Mar 2017
12. Christensen, H.: A Roadmap for US Robotics from Internet to Robotics, 2016 edn. Sponsored by National Science Foundation & University of California, San Diego (2016)
13. Littlefield, S.: Anti-Submarine Warfare (ASW) Continuous Trail Unmanned Vessel (ACTUV). http://www.darpa.mil/program/anti-submarine-warfare-continuous-trail-unmanned-vessel (undated). Accessed 26 Nov 2016
14. Anti-Submarine Warfare (ASW) Continuous Trail Unmanned Vessel (ACTUV) "Sea Hunter". http://www.navaldrones.com/ACTUV.html. Accessed 26 Nov 2016
15. Clapper, J.R.: DNI: Worldwide Threat Assessment of the US Intelligence Community. Statement for the Record, Senate Armed Services Committee, pp. 1–17, 9 February 2016

A Motivation for Co-adaptive Human-Robot Interaction

Caroline E. Harriott(✉), Sara Garver, and Meredith Cunha

The Charles Stark Draper Laboratory, Inc., Cambridge, MA, USA
{charriott,sgarver,mcunha}@draper.com

Abstract. As robotic technology advances, we experience a shift from seeing robots behind a barrier to working with robots as teammates. Building co-adaptive human-robot relationships will enable better teamwork between a human and a robot. Co-adaptive agents adapt their behavior over time in response to a dynamic understanding of the individual human operator. Creating a co-adaptive human-robot team requires bi-directional and non-invasive communication between the human and the robot. The presented study investigates the impact of individual traits on performance with three styles of adapting software. Results from this study indicate a need to tailor a co-adaptive robot's behavior for the individual, including operator traits – specifically dispositional trust and extraversion.

Keywords: Human-robot interaction · Human-computer interaction · Adaptive software · Personality traits · Human performance

1 Introduction

As robot technology advances, human-robot teams are accomplishing increasingly complex tasks (e.g., [1, 2]). Humans no longer confine robots to factories, separated by cages; rather, people are shifting to work with robots as teammates in their daily tasks. Robots are becoming increasingly nuanced in how they adapt, taking in more information and comprehension of operator intention through negotiation, and breaking down tasks [3].

In human-human relationships, as the team performs a task, each individual adapts to the other person over time, as well as adapting to the task (e.g., [4, 5]). Individuals work together collaboratively in team settings, learning each other's strengths and skills over time. Human teamwork does not require each person to have an exhaustive knowledge of other individual contributors' skills; instead, iterative adaptation on both sides improves performance and rapport. In human-robot relationships, the human may not fully understand or need to know about each individual capability or feature of the robot; likewise, a robot will not understand every aspect of a human's behavior. Previous research has shown that in fact perceptions are often misaligned, with human collaborators not fully understanding robotic capabilities (e.g., [6]).

Work in the human-robot interaction literature has identified the need for, and methods of, adapting the robot's behavior to the human collaborator (e.g., [7–9]). Co-adaptive robots will enable more balanced teamwork between a human and a robot; the robot will change its own behavior to support the needs of the individual human

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_15

teammate. An individual human's performance needs are shaped by personality traits, experience, current state (e.g., workload), or current task [11]. Co-adaptive agents adapt their behavior over time in response to a dynamic understanding of the individual human operator. Such an agent may be able to scope and prioritize the information presented to the operator by adapting to the operator's changing needs.

It has been demonstrated that graphical user interface design, such as color choice or location of features on a screen, impacts a user's attention, focus, and behavior. This paper focuses on adaptation styles in software, in order to understand the specific impacts of three adaptation styles, which change the level of user control, on behavior, preference, and performance. The goal is to understand how the role an adaptation takes (e.g., suggesting a subset of options as compared to preventing the selection of a number of options) may affect human behavior along specific human traits.

This paper presents motivation for creating a co-adaptive agent, for human-robot relationships, which considers operator traits to dictate the type or style of adaptation. Even if the adaptation is providing the same information, the method by which the information is presented can positively influence team performance if it matches the style that works best for the human teammate. Section 2 provides a brief overview of related work in collaboration and teamwork, adapting robot behavior, and individual traits. Section 3 describes the method used in our study. Section 4 provides the results, and Sect. 5 offers a discussion of the results.

2 Related Work

Collaboration and Teamwork. Team adaptation is a complex phenomenon. A input-throughput-output model was created to capture the complexity of human team adaptation [4]. Inputs to team adaptation include individual characteristics and the task description. During the adaptive cycle, the team assesses the situation, forms a plan, executes the plan, and learns from the results. Psychological safety, situation awareness, and shared mental models affect the way a team adapts and evolves. While modeling human-robot interaction has been explored [10], this type of model has not been adapted to specifically capture human-robot collaborative teams, leaving a gap in the literature.

Team training and shared cognition also were determined to be crucial aspects of teamwork [5]. The impact of team composition on team effectiveness is an active research area. The present study focuses only on one-human, one-agent teams, but future work will investigate how adaptation styles affect team adaptation.

Adapting Robot Behavior. Robots that adapt to the individual can provide benefits to the performance of the human-robot team. Prior work has demonstrated higher preference for robots that adapt their movements in response to the human partner's behavior [6].

Adapting the role of the robot in a human-robot collaborative task was also shown to be an improvement over complete robot control or human leadership [8]. Additionally, changing the human's task allocation in response to changing workload conditions (e.g., reducing the human's tasking) was shown to improve performance [9].

Individual Traits. Traits are unique qualities that differentiate individuals. A trait, such as openness to new experience, is a largely invariant (or very slowly variant) observable human characteristic [11]. The Big Five Inventory is the most widely accepted taxonomy for personality traits that describes personality in five categories: extraversion, agreeableness, conscientiousness, neuroticism, and openness [12].

The effect of some stable user traits on the usage of software adaptations has been studied [13], showing an increase in adoption of adaptation by individuals with low levels of extraversion and higher levels of Need For Cognition.

Individual human traits are an important and influential component of human-robot teaming and must be considered while developing robot teammates. The design implications of human responses to different adaptation styles, associated with individual traits, has not been investigated in human-robot interaction or in adaptive software, to the authors' knowledge.

3 Method

3.1 Experimental Design

The experiment leveraged a within-subjects design and consisted of eight image-editing tasks, performed in a custom graphical user interface based on the open-source image-editing software ImageMagick [14]. The tasks involved between 7 and 12 steps to complete – involving filters placed on the image and adding annotations, including captions, arrows, and boxes.

Task 4, for example, instructed the participants to "Place a compass in the center of the image. Find the items: Elephant, Frog prince, Giraffe. For each object: draw a box around the object and place a caption below the object describing the position of the object according to the compass you placed (ex: North, Northeast, South, Southwest, etc.)." Figure 1 demonstrates the starting image and an example of a completed task by one participant. The starting images were borrowed from the children's book, ISPY [15] as existing examples of visually complex images without requiring any background information to search.

The four conditions provided one base interface and three adaptation styles: "big buttons," "highlighting," and "macro" (see Fig. 2 for screenshots of the four interfaces from each condition). The big buttons adaptation was designed to limit participant choices, providing only the features needed for the task. The highlighting adaptation provided suggestions in-place, highlighting the suggested features without changing the interface presentation. The macro adaptation style provided feature chains to automate the execution of three features. This paper will focus on measured task efficiency, as measured by time taken to complete each task step. The tasks were paired with condition, and the condition ordering was counterbalanced. Each task required between 7 and 12 steps of the same order of magnitude of difficulty to complete; thus, task completion time was normalized by the number of steps required in order to more directly compare results from the four conditions.

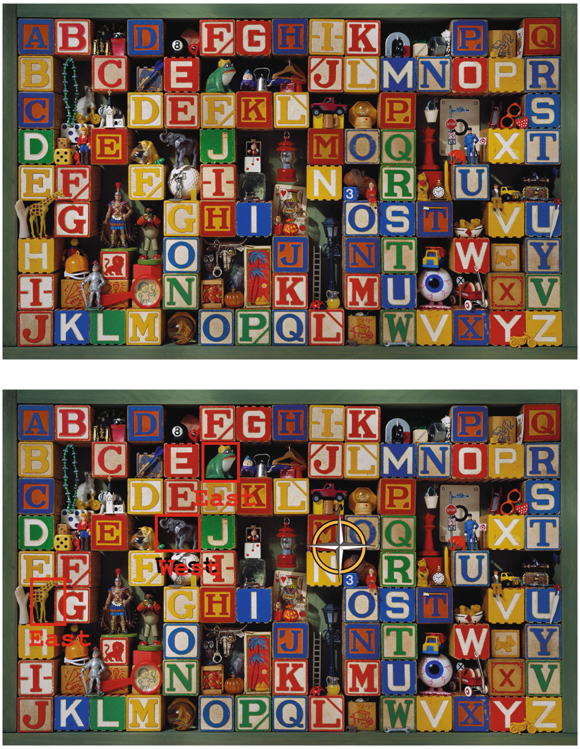
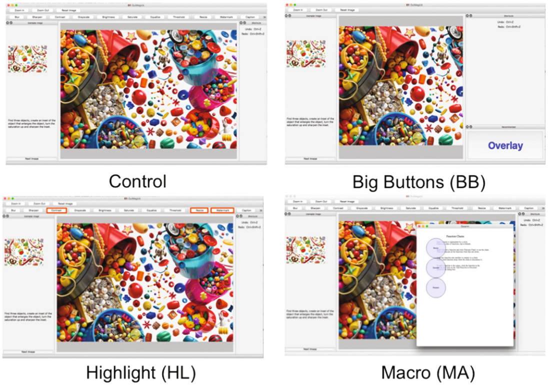


Fig. 1. Starting image (top) and example of resulting image (bottom). Participants manipulated given images in order to meet given task instructions. The given interface provided suggestions for specific interface features needed to complete the tasks.



Control

Big Buttons (BB)

Highlight (HL)

Macro (MA)

Fig. 2. Screenshots of the interfaces corresponding to four experimental conditions: control, big buttons (limiting features), highlighting (suggestions in-place), and macro (partial automation).

3.2 Participants

The experiment included 28 participants (11 male and 17 female). Their ages ranged between 18 and 55 years old. Participants were recruited via flyers and online postings in the Boston, Massachusetts area and were compensated $20.

3.3 Procedure

Each participant completed an initial intake survey that included basic demographic questions, and a series of personality inventories [16] (i.e., need for cognition [17], locus of control [18], dispositional trust, neuroticism, and openness to new experiences), and a Buy-In Assessment [19], assessing willingness to pay for an interface.

The participants were trained using a set of introductory slides with screenshots and explanations of the interface's primary features. Explanations of the mechanics of each adaptation also were included. The participants were told that the interface was attempting to help each user complete the tasks. Participants were able to consult the training material at any time throughout the experiment, but were not permitted to ask the experimenter questions regarding the interface or tasks. Participants completed blocks of two tasks within each condition. Condition order was counterbalanced.

Following the completion of each condition, participants completed the System Usability Scale (SUS) questionnaire [20] and secondary Buy-In Assessment, with a total of four blocks of the two surveys. After completing all four conditions – eight total tasks – the participants completed a preference ranking survey, asking them to compare the four conditions in terms of preference.

4 Results

The presented results include analysis of efficiency, as measured by time per required task step (see Sect. 3.1 for further description) and satisfaction, as measured by the SUS rating results. Each of these dependent measures was analyzed by condition and participant traits (see Sect. 3.3 for all captured traits). The two traits of focus in this analysis are extraversion and dispositional trust. Participants were grouped into tertiles based on their reported level of dispositional trust and extraversion. The breakdown (i.e., trait score range and number of participants per tertile) of the tertiles for extraversion and dispositional trust are reported in Tables 1 and 2, respectively.

Table 1. Extraversion ranges for each of the low, medium, and high tertiles used for analysis of efficiency and satisfaction.

Tertile	Extraversion range	n
Low	19–31	11
Medium	32–36	12
High	37–47	5

Table 2. Dispositional trust ranges for each of the low, medium, and high tertiles used for analysis of efficiency and satisfaction.

Tertile	Dispositional trust range	n
Low	22–36	10
Medium	37–41	10
High	42–50	8

Efficiency. The time per required task step descriptive statistics by condition are summarized in Table 3. A one-way ANOVA indicated a significant main effect of condition on time per required task step (F(3, 220) = 12.25, $p < 0.0001$). A Tukey Honestly Significant Difference (HSD) test indicated that the macro adaptation condition produced significantly longer time per task step than the control condition ($p < 0.0001$), big buttons adaptation ($p < 0.001$), and the highlight adaptation ($p < 0.0001$). The highlight adaptation condition resulted in the shortest task step completion times, and the control condition resulted in very similar time. The macro adaptation resulted in the slowest task completion times, almost twice the time taken to complete each task step in the highlight condition.

Table 3. Reported means and standard deviations for time per required task step by condition.

Condition	Mean time per required task step	Standard deviation (SD)
Control	7.73	3.34
Big buttons	8.21	5.46
Highlight	7.52	8.11
Macro	13.29	5.62

The time per required task step by trait also was analyzed – by dispositional trust and by extraversion. The descriptive statistics for time per required task step by trait is provided in Table 4. A one-way ANOVA indicated no significant main effect of extraversion tertile on time per required task step. A second one-way ANOVA indicated a significant main effect of dispositional trust tertile on time per required task step (F(2, 221) = 7.44, $p < 0.001$). A Tukey HSD post-hoc test indicated that high dispositional trust participants were significantly faster than low ($p = 0.01$) and medium $p < 0.001$) dispositional trust participants.

Table 4. Reported means and standard deviations for time per required task step by trait.

Tertile	Mean time per required task step	SD
Low extraversion	8.48	5.99
Medium extraversion	9.87	6.58
High extraversion	9.12	6.34
Low dispositional trust	9.78	7.02
Medium dispositional trust	10.56	6.86
High dispositional trust	6.74	3.39

A two-way ANOVA was conducted measuring the effect on time per task step by extraversion and condition (see Fig. 3). There was a significant main effect of condition (F(3, 212) = 12.56, p < 0.0001), but no significant main effect of extraversion tertile; there was a nearly significant interaction between extraversion and condition on time per task step (F(6, 312) = 1.83, p = 0.096). A Tukey HSD post-hoc test indicated that the macro adaptation condition resulted in significantly longer times per required task step than the control condition (p = 0.020) and the highlight condition (p = 0.003) for low extraversion participants. The macro adaptation took significantly longer than the big buttons adaptation (p = 0.002) and nearly significantly longer than the control (p = 0.067) for medium extraversion participants. No significant differences in time per task step were found for high extraversion participants.

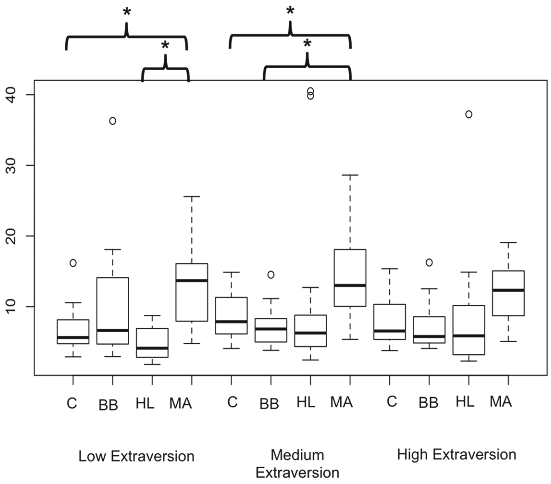


Fig. 3. Time per required task step by extraversion tertiles and condition: Control (C), Big Buttons (BB), Highlighting (HL) and Macro (MA). The asterisk in all figures denotes a statistically significant difference (p < 0.05).

A second two-way ANOVA was conducted to measure time per task step by dispositional trust tertile and condition (see Fig. 4). There was a significant main effect of condition (F(3, 212) = 13.09, p < 0.0001) and a significant main effect of dispositional trust tertile (F(2, 212) = 8.65, p < 0.001); there was no significant interaction between dispositional trust and condition on time per task step. A Tukey HSD post-hoc test indicated that using the macro adaptation resulted in significantly longer time per task step than the control (p < 0.001) and the highlight condition (p < 0.001) and nearly significantly longer than the big buttons adaptation (p = 0.056) for low dispositional trust participants. No significant differences in time per task step were found for medium and high dispositional trust participants.

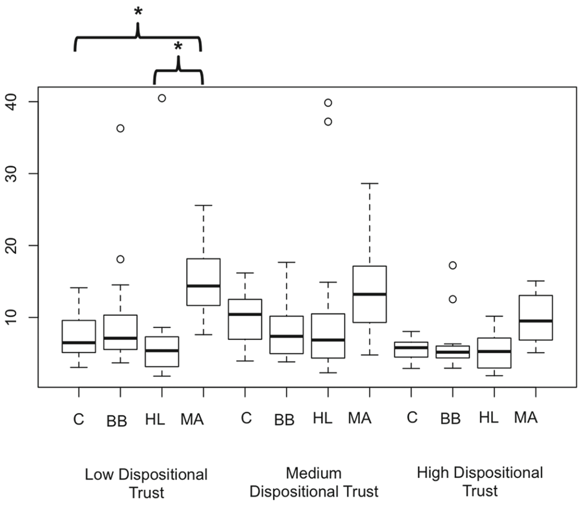


Fig. 4. Time per required task step by dispositional trust tertiles and condition.

Satisfaction. The SUS ratings descriptive statistics by condition are summarized in Table 5 SUS ratings have a possible range between 0 and 100, with 100 representing a perfect score. Scores below 68 are considered below average [20]. Overall the macro condition fared the worst, scoring below average. The control condition and big buttons scored very similarly, and the highlight condition had the highest score. A one-way ANOVA found a significant main effect of condition on SUS ratings (F(3, 108) = 10.60, p < 0.0001). A Tukey HSD post hoc test indicated significantly lower SUS ratings for the macro adaptation than the control condition (p < 0.001), the big buttons adaptation condition (p < 0.001), and the highlight adaptation condition (p < 0.0001).

Table 5. Reported means and standard deviations for SUS ratings by condition.

Condition	Mean SUS ratings	SD
Control	73.75	13.57
Big buttons	73.66	16.07
Highlight	75.80	15.31
Macro	53.57	21.93

SUS ratings by trait also were analyzed – by dispositional trust and by extraversion. The descriptive statistics for time per required task step by trait is provided in Table 6. A one-way ANOVA indicated a significant main effect of extraversion tertile on SUS ratings (F(2, 109) = 5.37, p = 0.006. Higher extraversion participants tended to provide higher SUS ratings overall. A Tukey HSD post hoc test indicated high extraversion participants provided higher SUS ratings than low (p = 0.03) and medium (p = 0.004) extraversion participants. A second one-way ANOVA indicated a significant main effect of dispositional trust tertile on SUS ratings (F(2, 109) = 4.34, p = 0.015). Participants with higher dispositional trust tended to provide higher SUS ratings. A Tukey HSD post hoc test indicated high dispositional trust participants provided higher SUS ratings than medium (p = 0.012) dispositional trust participants.

Table 6. Reported means and standard deviations for SUS ratings by trait.

Tertile	Mean SUS ratings	SD
Low extraversion	68.30	19.89
Medium extraversion	65.10	20.02
High extraversion	81.00	7.14
Low dispositional trust	68.19	16.20
Medium dispositional trust	64.06	21.75
High dispositional trust	76.88	16.82

A two-way ANOVA was conducted to measure SUS ratings by extraversion tertile and condition (see Fig. 5). There was a significant main effect of condition (F(3, 100) = 11.62, p < 0.001) and a significant main effect of extraversion tertile (F(2, 100) = 6.89, p = 0.002); there was no significant interaction effect between extraversion and condition on SUS ratings. A Tukey HSD post-hoc test indicated the macro adaptation was rated significantly lower than the highlight adaptation (p = 0.002) and the big buttons adaptation (p < 0.01) for low extraversion participants. Medium extraversion participants rated the macro adaptation significantly lower than the control (p = 0.037) and nearly significantly lower than highlight (p = 0.067).

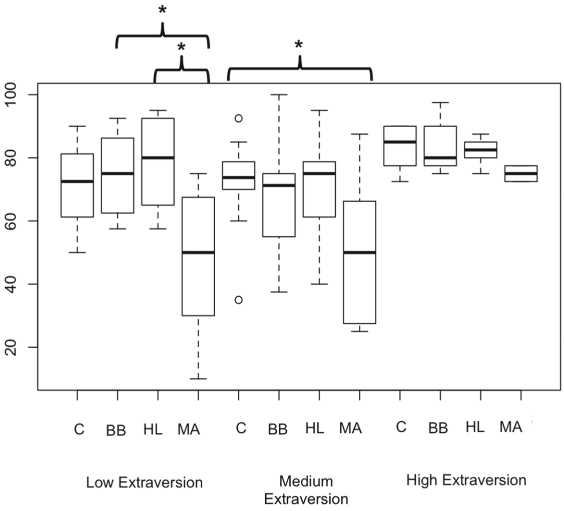


Fig. 5. SUS ratings by extraversion tertiles and condition.

A second two-way ANOVA analyzed the effect of dispositional trust and condition on SUS ratings (see Fig. 6). There was a significant main effect of condition (F(3, 100) = 11.22, p < 0.001) and a significant main effect of extraversion tertile (F(2, 100) = 5.46, p = 0.006); there was no significant interaction effect between extraversion and condition on SUS ratings. A Tukey HSD post-hoc analysis indicated that the macro adaptation resulted in significantly lower SUS ratings than the big buttons adaptation (p = 0.043). Low dispositional trust participants rated the macro adaptation significantly lower than the highlighting adaption (p = 0.011) and nearly significantly lower than the control (p = 0.072).

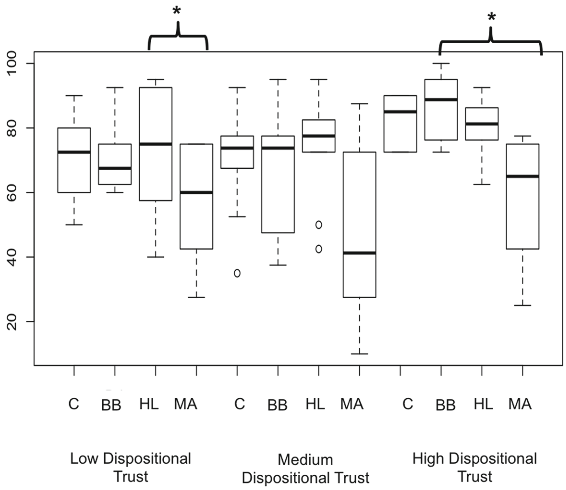


Fig. 6. SUS ratings by dispositional trust tertiles and condition: C, BB, HL, MA.

5 Discussion

Overall, this evaluation focused on two primary topics: (1) the effect of adaptation style on performance and (2) the effect of user trait on performance with adaptation styles. The results of this study indicate significant effects of adaptation style on performance. Additionally, the results indicate that two specific traits, dispositional trust and extraversion, affected the relationship between performance and adaptation style.

The outcome of this evaluation suggests that adaptation style generally affects performance; the adaptation styles evaluated in this study provided ways for users to access the true “right answer” regarding features to use for each task. It is important to note that providing the right answer to participants did not always help; while the highlight adaptation generally allowed participants to perform faster, the big buttons adaptation provided fewer choices to the participant and resulted in a slower average time per required task step. These results are counterintuitive when considering a typical additive model of choice reaction time – with fewer choices available, a faster decision time is expected; however, this was not the case in this study.

This research also demonstrates that traits are an important consideration when designing adaptation. These results suggest that general adaptation may not be successful for each participant and offering an individualized adaptation style based on trait will be more successful. An interesting finding from the efficiency analysis is that low levels of extraversion related to faster performance with the highlight adaptation and slower with the macro adaptation, while high levels of extraversion did not produce this difference – extroverts did not experience the same slow down with the macro adaptation nor a clear benefit from the highlight adaptation. Extraverts also did not rate any condition as significantly different from any other. These results may point to the flexibility of extraverts to accept and perform well with a variety of adaptations, while introverts may have stronger reactions. Introverts performed best with and rated highest the highlight adaptation. Participants with higher dispositional trust tended to execute task steps faster and rate satisfaction higher. This may indicate that those with lower dispositional trust may take extra time to evaluate the adaptation, without quickly accepting the adaptations' suggestions.

The interface evaluated in this study was custom-built and very simple; it offered fewer than twenty high level features (e.g., brightness). The assigned tasks were also designed to be easy, short-term, and abstract. Future work aims to evaluate co-adaptation using complex interfaces with a large and potentially overwhelming number of features, as well as an increase in task realism. It is important to evaluate the generalizability of the presented results to a wider set of applications.

Future work is focused on evaluating the generalizability of these results by assessing adaptations in an online data collection environment and recruiting participants via Amazon Mechanical Turk. An additional online data collection effort is ongoing, collecting interaction with a baseline interface – a robot navigation logic game. This data will contribute to the evaluation of the feasibility of detecting user traits via non-intrusive naturalistic behavior collection, using Draper's Software as a Sensor capability [21]. Determining non-intrusive methods of measuring human behavior in mobile human-robot teams is non-trivial, and future work will investigate the expansion of naturalistic behavior capture and modeling for creating co-adaptive human-robot relationships.

The presented results offer a comparison between three styles of software adaptation based on degree of user freedom; highlight offered users the most freedom, allowing users to choose any feature. Introverts tended to benefit from the highlight adaptation most, preferring to keep control. The macro adaptation was shown widely to be the least successful adaptation – partial automation of the task was not helpful, and did not consider the usage patterns and ordering of task steps that may be most intuitive for the user. Extroverts tended to handle the macro adaptation best, perhaps indicating a willingness of extroverts to hand over control to a co-adaptive teammate. The level of choice and autonomy given to each teammate in co-adaptive human-robot teams should consider the human traits.

Acknowledgments. This research is supported by Draper Internal Research & Development. The authors thank Ryan Brill, Brian Kimerer, and Krysta Chauncey for their contributions to this work.

References

1. Hayes, B., Scassellati, B.: Autonomously constructing hierarchical task networks for planning and human-robot collaboration. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 5469–5476 (2016)
2. Evans, A., Marge, M., Stump, E., Warnell, G., Conroy, J., Summers-Stay, D., Baran, D.: The future of human robot teams in the army: factors affecting a model of human-system dialogue towards greater team collaboration. In: Advances in Human Factors in Robots and Unmanned Systems, p. 197 (2017)
3. Nikolaidis, S., Nath, S., Procaccia, A.D., Srinivasa, S.: Game-theoretic modeling of human adaptation in human-robot collaboration. In: Proceedings of Human Robot Interaction, Vienna, Austria, March 2017
4. Burke, C.S., Stagl, K.C., Salas, E., Pierce, L., Kendall, D.: Understanding team adaptation: a conceptual analysis and model. Appl. Psychol. **91**(6), 1189 (2006)
5. Salas, E., Cooke, N.J., Rosen, M.A.: On teams, teamwork, and team performance: discoveries and developments. Hum. Factors **50**(3), 540–547 (2008)
6. Cha, E., Dragan, A.D., Srinivasa, S.S.: Perceived robot capability. In: Proceedings of Robot and Human Interactive Communication (RO-MAN), pp. 541–548 (2015)
7. Shen, Q., Dautenhahn, K., Saunders, J., Kose, H.: Can real-time, adaptive human – robot motor coordination improve humans' overall perception of a robot? IEEE Trans. Auton. Ment. Dev. **7**(1), 52–64 (2015)
8. Li, Y., Tee, K.P., Chan, W.L., Yan, R., Chua, Y., Limbu, D.K.: Role adaptation of human and robot in collaborative tasks. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 5602–5607 (2015)
9. Shannon, C.J., Horney, D.C., Jackson, K.F., How, J.P.: Human-autonomy teaming using flexible human performance models: an initial pilot study. In: Advances in Human Factors in Robots and Unmanned Systems, p. 211 (2016)
10. Harriott, C.E., Adams, J.A.: Modeling human performance for human–robot systems. Rev. Hum. Factors Ergon. **9**(1), 94–130 (2013)
11. Winter, D.G., John, O.P., Stewart, A.J., Klohnen, E.C., Duncan, L.E.: Traits and motives: toward an integration of two traditions in personality research. Psychol. Rev. **105**(2), 230 (1998)
12. John, O.: The 'Big Five' factor taxonomy: dimensions of personality in the natural language and in questionnaires. In: Pervin, L. (ed.) Handbook of Personality: Theory and Research. Guilford, New York (1990)
13. Gajos, K.Z., Chauncey, K.: The influence of personality traits and cognitive load on the use of adaptive user interfaces. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces, pp. 301–306 (2017)
14. Cristy, J., Bergougnoux, K., Bogart, R., Peterson, J.W., Brown, N., Chiarappa, M., Crimmins, T.R., Edwards, T., Fojtik, J., Franklin, F.J., Friedl, M.: ImageMagick (1994). http://www.imagemagick.org
15. Marzollo, J., Wick, W., Carson, C.D.: I SPY: A Book of Picture Riddles. Cartwheel Books, New York (1992)
16. Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., Gough, H.C.: The international personality item pool and the future of public-domain personality measures. J. Res. Pers. **40**, 84–96 (2006)
17. Cacioppo, J.T., Petty, R.E.: The need for cognition. J. Pers. Soc. Psychol. **42**(1), 116 (1982)

18. Levenson, H.: Differentiating among internality, powerful others, and chance. In: Lefcourt, H.M. (ed.) Research with the Locus of Control Construct, vol. 1, pp. 15–63. Academic Press, New York (1981)
19. Chauncey, K., Harriott, C.E., Prasov, Z., Cunha, M.: A framework for co-adaptive human-robot interaction metrics. In: Proceedings of the Workshop on Human-Robot Collaboration at IEEE/RSJ International Conference on Intelligent Robots and Systems, Daejeon, Korea, October 2016
20. Brooke, J.: SUS-A quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
21. Mariano, L.J., Poore, J.C., Krum, D.M., Schwartz, J.L., Coskren, W.D., Jones, E.M.: Modeling strategic use of human computer interfaces with novel hidden Markov models. Front. Psychol. **6**, 919 (2015)

An Integration of Cognitive Task Analysis Results for Situation Awareness-Focused Training Program Development

David B. Kaber[1], Rebecca S. Green[1], and Manida Swangnetr[2](✉)

[1] Edward P. Fitts Department of Industrial and Systems Engineering,
North Carolina State University, Raleigh, NC 27695-7906, USA
dbkaber@ncsu.edu, rebecca.green@cerner.com

[2] Department of Production Technology, Research Center in Back, Neck,
Other Joint Pain and Human Performance, Khon Kaen University,
Khon Kaen 40002, Thailand
manida@kku.ac.th

Abstract. The objective of this study was to develop an approach for translation of cognitive task analysis (CTA) results to training program content for operators of high throughput screening (HTS) systems. Currently, no standardized methods exist for translating outcomes of multiple CTA methods to support training program design and promote situation awareness (SA). We combined information requirements from a goal-directed task analysis (GDTA) and system resource requirements from abstraction hierarchy (AH) models to establish content on HTS processes and to address three levels of operator SA. The two electronic forms of training were compared with on-the-job (OTJ) training. Results provided preliminary evidence that CTA-based training increased operator knowledge beyond OTJ training and increased SA-related knowledge beyond operator initial system knowledge. A set of general guidelines was developed for design of CTA-based training programs, including methods for structuring components of the training program to support SA.

Keywords: Cognitive task analysis · Life science automation · Training design · Situation awareness · Goal-directed task analysis · Abstraction hierarchy models

1 Introduction

Over the past decade, high throughput screening (HTS) has increased the pace at which biological compounds can be tested for potential uses in development of new drugs or industrial products. Currently, a combination of modern robotic systems, data processing and control software, liquid handling devices and sensor technology allows researchers to effectively conduct thousands of biochemical, genetic or pharmacological tests in a short period of time [1].

Generally, manual biochemical assays can be transferred to automated robotic systems with limited changes [1]; however, many process details need to be reviewed to ensure consistency of results across automated runs and with manual assay

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_16

performance at a bench or individual workstations. Consequently, the role of human operators in this domain has shifted from manual material handling tasks to planning, controlling, and analyzing results of automated screening line outputs. This shift has led to an increase in systems monitoring workload and a higher requirement for accuracy in operation performance, due to the need for greater attention to decision-making and error handling tasks as part of HTS processes.

The emergence of new automated technologies in the domain of HTS has challenged traditional approaches to the education and training of laboratory technicians. The backgrounds of HTS operators are primarily in biological sciences often with little or no experience in automated systems development and control. Unlike typical automated manufacturing processes, tasks and procedures associated with life science automation are not well defined and often involve trial and error before achieving successful system integration. HTS operations are akin to conducting research and development combined with systems troubleshooting. These functions occur under tight time constraints with rare and costly test materials creating high stress for operators and the potential for substantial losses with errors.

Most current training practices involve a combination of review of operation manuals and documents, on-the-job (OTJ) training, formal classroom training, and software tutorials that have been structured based on hierarchical task analysis (HTA). The OTJ training of new HTS operators usually consists of several weeks of assisting a lead biochemist in order to become familiar with methods and automation in use [2]. This process is typically unstructured, which means there is usually no written documentation of training procedures to follow and there are few objective means by which to measure task learning in order to ensure all operators are trained to similar levels of skill. As another drawback, during actual HTS operations, technicians may encounter high workload conditions or automated system errors (e.g., in programming or process execution) that may not be addressed during OTJ training.

Since learning involves the process of (memory) schema construction and skill automation for a range of operating conditions, it is important for a learner to effectively link information gathered when practicing a task to appropriate methods in memory (e.g., error recovery procedures). Practice relative to specific task demands in the use of HTS systems can provide learners with the opportunity to develop problem-solving schema, potentially reducing working memory demands during actual operations and leading to improved performance. Related to this, operator capability to achieve situation awareness (SA; [3]) is important for successful HTS performance, as SA requirements focus not only on what information operators need for task performance (Level 1 SA), but also how information is integrated or combined (Level 2 SA) to address each decision within a process under both nominal and off-nominal conditions. Moreover, the ability to project the future status (Level 3 SA) of a HTS process is required in order to prevent potential errors. Practice of such SA requirements has been facilitated using computer-based training systems with content based on HTAs for various operations. However, the identification of learner information processing requirements and demands posed by a HTS task and automation may be more effectively achieved by using cognitive task analysis (CTA) methods and results as a basis for training program development. The use of CTA methods as a basis for training program development has previously been documented by other research [4] but not in the context of HTS and resulting programs have not been experimentally compared with traditional training methods (e.g., OTJ).

The objective of this study was to develop an approach for translation of CTA results to training program content for operators of HTS systems. Currently, no standardized methods exist for translating outcomes of multiple CTA methods to support operator training program design and promote SA during task performance. Details of methodology of the proposed CTA-based training development and preliminary experimental results are presented in subsequent sections. Finally, a set of general guidelines for design of CTA-based training programs were developed, including methods for structuring components of the training program to support SA.

2 CTA Methods Towards Developing SA-Specific Training

Several CTA methods have previously been identified as resources for training program development (see [4] for a comprehensive review). In the case of HTS system operations, the identification of task demands and information processing requirements for SA-specific training for operators may be best achieved through goal-directed task analysis (GDTA; [5]). The method focuses on identifying operator perception, comprehension and projection requirements for performing complex systems control. Prior studies have identified GDTA as a tool for revealing specific SA requirements and to provide a basis for SA training courses [6]. One major limitation of GDTA models is that they do not make reference to existing automated systems or software. Therefore, another CTA method is needed to represent the purpose and function of the software and devices that operators interact with in task performance.

Among the available methods, abstraction hierarchy (AH) models can be used to describe process automation and interfaces as part of training. AH modeling is a representation framework used to describe human-machine interaction through a hierarchy of functional relationships of a complex working environment in an event-independent manner in order to inform operators of approaches to recovery from unanticipated error conditions [7]. Such modeling has been found to be an effective tool for revealing how automated system processes and operator functions are facilitated through specific system components and to provide explanations of why certain components are needed to achieve system purposes. However, one major concern in AH-based training program design is that important cognitive skills may be neglected.

Previous research by Kaber et al. [8] developed AH models for automated devices and proprietary software integrated in a HTS line. This knowledge can also be considered fundamental to any training program for HTS systems. Kaber et al. [8] also indicated that use of the GDTA methodology may provide a method for identifying skills required to use automated process systems in terms of SA requirements, and such SA skills may be embedded in appropriate training techniques. However, they did not address the potential of GDTA for HTS training program development.

The approach of the present research was to combine information requirements identified through GDTA with specification of current interface action sequences, based on AH models, in order to establish training content for HTS processes. This approach allowed for operator goals in use of HTS automation (based on the GDTA) to be described in terms of simple interface actions (determined from the AH models). In this way, specific interface behaviors could be related to high-level cognitive task goals or learning outcomes. Figure 1 summarizes the approach of integrating components of the GDTA and AH models to develop CTA-based training program content. Operator sub-goal information from the GDTA was linked to the functional purposes of HTS systems represented in AH models (connecting dashed lines in figure) as a basis for identifying specific tasks to be trained. The decisions under each sub-goal in the GDTA were linked to generic functions in the AH system models to develop training content for each process step. This information was used as a basis to further specify training content to address the three levels of operator SA. Level 1 SA requirements were associated with physical device interfaces to identify critical cues an operator should learn for accurate perception of the HTS environment. Level 2 and 3 SA requirements represented information that could be used as a basis for assessing operator learning in any task step. All of this content was presented through a part-task simulation trainer in which certain information appeared through different displays.

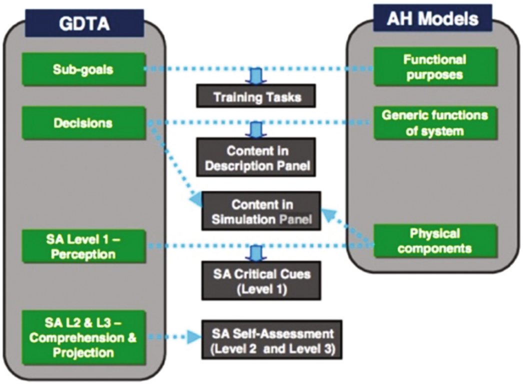

GDTA
Sub-goals
Decisions
SA Level 1 – Perception
SA L2 & L3 – Comprehension & Projection
AH Models
Functional purposes
Generic functions of system
Physical components
Training Tasks
Content in Description Panel
Content in Simulation Panel
SA Critical Cues (Level 1)
SA Self-Assessment (Level 2 and Level 3)

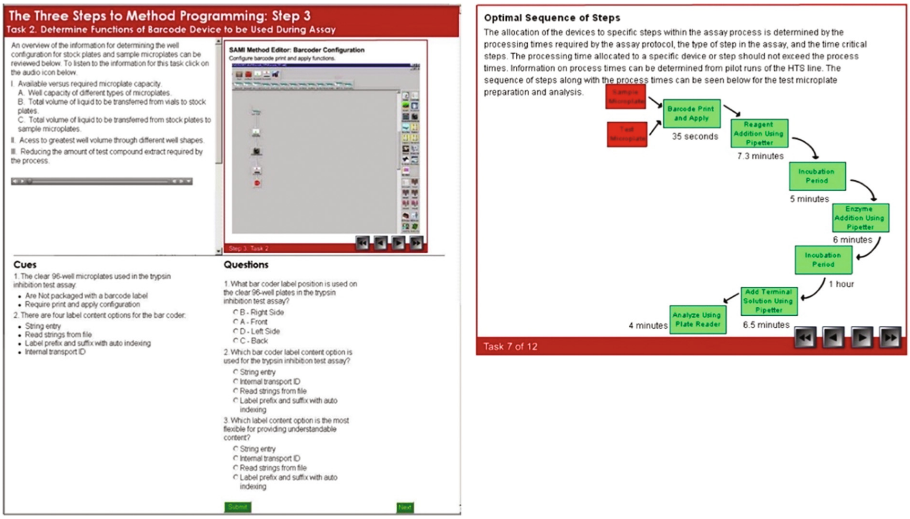
Fig. 1. Integration of CTA method content for training program design.

3 Training System Prototyping

3.1 CTA-Based Training System

Existing learning models were identified through training literature as a basis for structuring and formatting system content. Gagne et al.'s [9] classification of learning models provides a methodology for also classifying instructional content for a target task and developing a hierarchy of learning objectives. A training program can exploit such hierarchies in terms of identifying concepts that must be learned by an operator at a particular stage of training (or in combination with each other) and planning instructional sequences. The multimedia learning systems model [10] suggests that any computer-based training (CBT) program with the goals of motivating users to learning and facilitating encoding and retention should do so through effective and efficient presentation of information with multimedia features. Consequently, the training system presented in this research included multimedia animation, narration as text, and step-by-step simulations of each task. Finally, the learner-centered design (LCD) approach [11] to training systems addresses the importance and methods of usability and evaluation of trainee knowledge in CBT implementations. The present training program design: (1) followed Gagne et al. approach by using the GDTA and AH models (or HTA for the traditional training system design) to systematically identify user knowledge requirements; (2) supported user motivation, encoding and retention of information through application of multimedia learning systems approach; and (3) made use of a usability evaluation technique to assess the effectiveness, efficiency and motivational aspects of the training systems design as in the LCD approach.

The interface of the prototype training system was arranged in four panels (see an example in Fig. 2(a)). One panel included a text description of the status of equipment and processes at a given point in a HTS task sequence. Another included a part-task simulation of the process (which was also used to show previous and subsequent points in the task sequence). A third panel included a list of SA requirements, focusing on perception of key elements in the task environment (Level 1 SA). The final panel presented a set of self-assessment questions to promote operator understanding and projection of future states of the process (Level 2 and Level 3 SA). Training prerequisites for HTS system operation provided a basis for the content of the "critical cues" panel and the "embedded self-assessment" panel of the interface. Immediate feedback was also presented through the self-assessment panel to provide operators with knowledge of results.



The Three Steps to Method Programming: Step 3

Task 2. Determine Functions of Barcode Device to be Used During Assay

An overview of the information for determining the well configuration for stock plates and sample microplates can be reviewed below. To listen to the information for this task click on the audio icon below.

I. Available versus required microplate capacity.
A. Well capacity of different types of microplates.
B. Total volume of liquid to be transferred from vials to stock plates.
C. Total volume of liquid to be transferred from stock plates to sample microplates.
II. Acess to greatest well volume through different well shapes.
III. Reducing the amount of test compound extract required by the process.

SAMI Method Editor: Barcoder Configuration

Configure barcode print and apply functions.

Cues

1. The clear 96-well microplates used in the trypsin inhibition test assay.
- Are Not packaged with a barcode label
- Require print and apply configuration

2. There are four label content options for the bar coder:
- String entry
- Read strings from file
- Label prefix and suffix with auto indexing
- Internal transport ID

Questions

1. What bar coder label position is used on the clear 96-well plates in the trypsin inhibition test assay?
○ B - Right Side
○ A - Front
○ D - Left Side
○ C - Back

2. Which bar coder label content option is used for the trypsin inhibition test assay?
○ String entry
○ Internal transport ID
○ Read strings from file
○ Label prefix and suffix with auto indexing

3. Which label content option is the most flexible for providing understandable content?
○ String entry
○ Internal transport ID
○ Read strings from file
○ Label prefix and suffix with auto indexing

Optimal Sequence of Steps

The allocation of the devices to specific steps within the assay process is determined by the processing times required by the assay protocol, the type of step in the assay, and the time critical steps. The processing time allocated to a specific device or step should not exceed the process times. Information on process times can be determined from pilot runs of the HTS line. The sequence of steps along with the process times can be seen below for the test microplate preparation and analysis.

(a) (b)

Fig. 2. Example of content screen as part of HTS training system for: (a) CTA-based; and (b) traditional HTA-based training system.

3.2 Traditional HTA-Based Training System

The traditional training system interface was based on a standard format for technical training programs, in which each screen presents a task objective and detailed information relating to the objective. For content development, a semi-structured interview, as described by Jonassen et al. [12], was conducted with an expert biopharmacologist to determine the sequence of task objectives in HTS processing and to capture the detailed information relating to each objective for assay optimization. Given the use of traditional task analytic (HTA) methods, the training system does not capture system operator SA requirements. An example of the interface for the traditional HTA-based training system for a specific HTS task objective can be seen in Fig. 2(b). Each screen included: (1) the sub-task objective for each point in the process (at the top of the screen); (2) a bulleted list of information about the equipment or environment related to the process covered by the current objective; and (3) images of devices or process steps to illustrate the task information.

4 Preliminary Experimental Study

For this preliminary study, a small sample of four expert scientists were recruited from the University of Rostock (URO, Germany) Center for Life Science Automation for evaluation of the training system program. The participants ranged in age from 37 to 41 years (M = 39, SD = 1.8) and three were male. Their educational background was either in chemistry, life science automation, or process measurement and control. The level of education or degrees of the participants ranged from a diploma (5 yrs. in the German university system) to PhD (4–6 additional yrs.). With respect to work experience, participants ranged from 2–6 yrs. on the job as a biochemist or/engineer. All experts had prior experience with life science process automation.

The experimental training program was delivered through two experimental sessions. The first session included: instructions of how to access and use the HTS training programs, the background survey, initial knowledge tests, training programs, usability evaluation, and training effectiveness survey. A randomized within-subjects experiment design was applied for presentation of the two training programs (CTA-based and traditional training). The two programs were targeted at different tasks in order to prevent condition carryover effects in the experiment. Related to this, the relative cognitive and physical work complexity of the tasks was comparable, as indicated by the content of the GDTA diagrams and relevant AH models for each task. After a one-week retention period, the second session was conducted by replication of above steps in the other type of training program. During the one-week retention interval, all participants continued with their regular job tasks at the Center, which did not include exposure to the exact task information presented in either training program. A final debriefing was conducted to cover the objectives of the study.

The response measures used in this study included: (1) the operator experience and education level; (2) initial knowledge test scores (to measure current OTJ training knowledge); (3) an SA assessment for the CTA-based training program using the situation awareness global assessment technique (SAGAT; [3]); and (4) final knowledge test scores. Two separate tests and performance criteria were created to measure required operator skills for each task. The knowledge test was created as multiple-choice, four-answer questions, including 38 questions related to the CTA-based training program and 17 questions for the standard training program. The training was structured in this way due to the CTA-based approach being broader in terms of steps. In general, each task goal or sub-goal was translated into a selected-response assessment item using templates described by Haladyna [13] in order to ensure "good" assessment item construction. The SA test, included a set of 72 SA queries, which were formulated based on the information requirements identified in the GDTA. SAGAT queries were presented to trainees in parallel with the training content. At random points in a training program, an additional interface was presented with a subset of SA queries targeting those sub-tasks recently completed by a participant. The queries were based on the content of the critical-cues panel and the embedded-questions panel for each sub-task. All participants responded to SA queries in the use of the CTA-based training programs.

In general, we expected (E1) that the integrated GDTA and AH modeling approach to training program design would significantly improve biopharmacologist knowledge structures beyond prior OTJ training as well as traditional training programs. It was also expected (E2) that the CTA-based training program identifying the information elements of the HTS domain, related to each of the levels of SA in Endsley's [3] theory, would lead to a more complete learning of SA requirements and support achievement of an expert level of performance (i.e., 80% correct responses to queries) in HTS processes for trainees as compared to OTJ training.

5 Results

Due to the limited sample size of this preliminary assessment, non-parametric statistical tests were used to make initial inferences based on the test data.

5.1 Effects of Training Program on Knowledge Test

A Wilcoxon rank sum test procedure indicated that participants significantly improved in knowledge of the HTS tasks as a result of the CTA-based training program (median = 0.26, SD = 0.10, W = 24.5, p = 0.03) as well as the traditional training program (median = 0.25, SD = 0.12, W = 25, p = 0.04). Related to this, results indicated no ceiling effects for performance on the initial knowledge assessment test; i.e., participants entered the study with some learning potential. These results were partially in support of our initial expectation (E1). The CTA-based training program did add to operator OTJ training but did not appear superior to the traditional (HTA-based) training program.

The degree of performance improvement on the knowledge tests (i.e., the change in the scores from the initial to the final knowledge test) were also compared across the two training systems. Counter to our initial expectation (E1), there was not sufficient evidence to indicate that participants improved to a greater extent with the CTA-based training program than for the traditional training program (median = 0.09, SD = 0.21, W = 19, p = 0.44). However, it is important to recall that the tasks trained with the two systems were of similar technical complexity but the topic for the CTA-based training was broader.

5.2 Effects of Operator Background on Knowledge Test

Since the background of operators may be critical in developing process knowledge structures, we conducted correlation analyses on performance on the initial and final knowledge assessment tests for both training programs with the number of years of operator experience and level of education. The analyses are presented in Table 1. There was evidence of a significant positive association between level of education (Education) and the initial knowledge assessment test scores relating to the traditional training program (TRAD-I), p = 0.05. There was also evidence of a significant positive association between number of years of experience (Experience) and the final knowledge assessment test scores for the CTA-based training program (CTA-F), p = 0.05. In general, results indicated that operator development of knowledge through the CTA-based training built-on prior OTJ training; whereas, development of additional knowledge through the traditional training program was not dependent upon prior experience. On the basis of these results, it is possible that the content and structure of the CTA-based training program may have been easier for operators to integrate into existing schemata (memory) structures on HTS operations than the content of the traditional training program.

Table 1. Correlations for number of year experience and level of education for knowledge assessment test scores.

Variable	Spearman rho	
	Experience	Education
Education	−0.83	
CTA-I	0.32	0.21
TRAD-I	−0.63	0.95*
CTA-F	0.95*	−0.63
TRAD-F	−0.83	0.89

(*p = .05)

Based on these relationships between the type of training program and the overall experience measures, an additional detailed comparison of knowledge test performance for “high” and “low” education levels was conducted. Due to the small sample size, only descriptive statistics are provided (see Table 2). The changes in performance indicating that the CTA-based approach was, on average, more effective for less educated operators than for highly educated operators. Comparison across the types of training programs revealed the operator (No. 2) demonstrating the least improvement in performance with CTA-based training had the greatest performance with the traditional training program. Another operator (No. 3), who demonstrated the greatest improvement with the CTA-based training program, had the least improvement with the traditional training program.

These differences suggest that there might have also been a preference in learning style or training method for some operators recruited for the study.

Table 2. Training program knowledge test performance (proportion of correct responses to questions) based on education (I = Initial test; F = Final test).

Operator no.	Education	CTA-I	CTA-F	TRAD-I	TRAD-F
1	High	0.68	0.82	0.59	0.76
2	High	0.53	0.66	0.29	0.65
3	Low	0.58	0.95	0.47	0.65
4	Low	0.47	0.79	0.53	0.88

For analysis of the dependence of the task training based on the CTA-based training program and prior technician experience, an additional set of correlations were determined for operator performance on the final knowledge test with operator ratings of the frequency with which they had performed target tasks as well as ratings of their current level of knowledge of related skills. Results revealed a significant negative association between the frequency of planning automated devices and performance on related knowledge test questions, $p < .0001$. It is possible that the CTA-based training included a method for this particular task that was representative of specific HTS operator expertise. There was also significant evidence of a positive association between operator ratings of the level of knowledge required to program automated devices and performance on related knowledge test questions, $p < .0001$. Unlike with the automation planning step, greater knowledge of programming was associated with large changes in knowledge assessment scores for the CTA-based training program.

5.3 Effects of Training Program on SA Test

To examine the impact of the CTA-based training on operator SA, the percentage of correct responses to SAGAT queries targeting each level of SA and overall SA were calculated for the CTA-based training program. Wilcoxon sign rank tests on all responses were significantly different from chance, $p < 0.05$. That is, the CTA-based training program led to improvements in operator SA for HTS operations. Furthermore, additional tests indicated that the level of performance achieved by the training participants was comparable to (i.e., not significantly different from) the level of performance expected of an expert operator (i.e., 80% correct responses to queries), $p > 0.05$. This finding was in line with our expectation (E2).

5.4 Effects of Operator Background on SA Test

Results of the correlation analyses revealed significant evidence of a positive association between number of years of experience (Experience) and level of education (Education) for each level of SA and overall SA, $p < 0.0001$(Table 3). There was also evidence of a positive association between initial knowledge score and Level 3 SA scores (projection), $p = 0.05$.

Table 3. Correlations among years of experience, levels of education, initial knowledge test scores, and SA score for each level of SA.

SA level	Variable	Spearman rho		
		Education	Experience	CTA-I
1	Education			
	Experience	1.00**		
	CTA-I	0.21	0.21	
	SA	–0.32	–0.32	0.80
2	Education			
	Experience	1.00**		
	CTA-I	0.21	0.21	
	SA	–0.83	–0.83	0.32
3	Education			
	Experience	1.00**		
	CTA-I	0.21	0.21	
	SA	0.06	0.06	0.95*
Overall	Education			
	Experience	1.00**		
	CTA-I	0.21	0.21	
	SA	–0.50	–0.50	0.74

(*$p < .10$, **$p < .0001$)

6 Discussions and Conclusions

This study involved the systematic use of contemporary CTA methods, including GDTA and AH modeling, as an approach to identifying HTS operator knowledge requirements, factors in operator SA, and system components and resources required for programming, control and optimization of automated process systems. The approach also involved comparison of GDTA results and AH models to formulate the content of a CTA-based training program.

Results of a preliminary experiment revealed partial support for the CTA-based training design for improving HTS operator knowledge structures for planning and comprehension. This finding was attributed to the CTA-based program presenting operators with information requirements for each task organized according to levels of SA. Knowledge test scores indicated significantly greater performance for both training programs compared to OTJ training; however, use of the CTA-based training program did not produce significantly greater improvements in operator knowledge of HTS processes than the traditional training program. This finding may be attributable to the breadth of the task trained using the CTA-based approach as compared to a task with fewer steps as trained using the traditional program. This conservative approach for assessment of the new CTA-based training program might have put the prototype at a disadvantage to the traditional training program.

Correlation analysis results indicated the CTA-based training program content to be significantly related to what operators learn in actual HTS operations. Furthermore, relationships between specific types of operator experience and knowledge assessment test performance were found for the CTA-based training program, but not for the traditional training program. These findings were attributed to the potential compatibility of the CTA-based training program content and structure with operator existing schemata (memory structures).

Performance on the SA questionnaires indicated operators were not guessing at answers and achieved improvements in knowledge from the CTA-based training. The CTA-based training program appeared to be particularly effective in improving operator perceptual skills (Level 1 SA). Results also indicated a general improvement in Level 2 and 3 SA scores and overall SA scores for HTS operators (with both high and low levels of education) beyond their initial knowledge of automated processes. These results serve to identify where the training program was most effective in providing Level 1 SA. They also indicate where operators with greater experience were able to improve their higher-level SA beyond initial knowledge of a HTS system.

On the basis of these results, there are several lessons learned that can be identified for training program development using the methods applied in this research:

(1) Either CTA or traditional (HTA)-based training programs can be used for developing fundamental process knowledge with selection based on individual learning-styles. In this preliminary experiment, both approaches were found to lead to significant improvements in HTS operator knowledge, as compared with OTJ, but with no difference between approaches.
(2) CTA-based training programs should provide trainees with a method for relating task goals to process states as part of specific sub-task performance as well as with the overall skill being trained.
(3) In CTA-based training program design, training information should be provided in order of relevance, as required by the task being trained.
(4) Operators of complex automated systems should be provided with lists of perceptual requirements necessary to develop Level 1 SA on sub-tasks to be trained and performed during operations. The CTA-based training program investigated here was most effective in providing Level 1 SA.
(5) Subsequent to an initial knowledge assessment and use of a CTA- or HTA-based training program, trainees should be posed with additional requirements, including a set of knowledge-assessment questions in order to promote target task/process comprehension and the ability to project future process states (Level 2 and Level 3 SA). Trainees should be provided with immediate feedback on their performance in such assessments. Focus should be placed on supporting trainees with fewer prerequisites in achieving higher SA.
(6) In presenting any CBT on system components, content should be included in both text and audio format.

These basic guidelines may be used to enhance CBT of complex human-automation interaction tasks, in general, and as bases for design of CTA-based training programs.

The guidelines may promote usable delivery of information for trainee development of SA and process knowledge structures.

Future research should evaluate the use of the structured approach for SA-based training program design using CTA results in the development of training programs for other domains. Larger sample sizes of expert operators will be used to promote the statistical reliability of any comparative results among training programs. Another future line of research might include an evaluation of transfer of learning with CTA-based training within the HTS domain from historical to updated or similar systems.

Acknowledgements. A National Science Foundation (NSF) Information Technology Research Grant (no. 046852) supported this research. Ephraim Glinert was the technical monitor for the NSF. The views and opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF. We thank the University of Rostock and the Center for Life Science Automation for providing us with access to high-throughput biological screening systems and supporting the research through allocation of biopharmacologist time.

References

1. Cohen, S., Trinka, K.: High-throughput screening. In: Janzen, W.P. (ed.) High Throughput Screening: Methods and Protocols, pp. 213–228. Humana Press, Totowa (2002)
2. Hamilton, S.: Introduction to screening automation. In: Janzen, W.P. (ed.) High Throughput Screening: Methods and Protocols, pp. 169–193. Humana Press, Totowa (2002)
3. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**(1), 32–64 (1995)
4. Clark, R.E., Feldon, D., van Merriënboer, J., Yates, K., Early, S.: Cognitive task analysis. In: Spector, J.M., Merrill, M.D., van Merriënboer, J.J.G., Driscoll, M.P. (eds.) Handbook of Research on Educational Communications and Technology, 3rd edn., pp. 577–593. Lawrence Erlbaum Associates, Mahwah (2007)
5. Endsley, M.R.: A survey of situation awareness requirements in air-to-air combat fighters. Int. J. Aviat. Psychol. **3**(2), 157–168 (1993)
6. Endsley, M.R., Robertson, M.M.: Training for situation awareness in individuals and teams. In: Endsley, M.R., Garland, D.J. (eds.) Situation Awareness: Analysis and Measurement, pp. 349–366. Lawrence Erlbaum Associates, Mahwah (2000)
7. Bisantz, A.M., Vicente, K.J.: Making the abstraction hierarchy concrete. Int. J. Hum. Comput. Stud. **40**, 83–117 (1994)
8. Kaber, D.B., Segall, N., Green, R.S., Entzian, K., Junginger, S.: Using multiple cognitive task analysis methods for supervisory control interface design in high-throughput biological screening processes. Int. J. Cogn. Technol. Work **8**, 237–252 (2006)
9. Gagne, R.M., Briggs, L.J., Wager, W.W. (eds.): Analysis of the learning task. In: Principles of Instructional Design, pp. 145–164. Harcort Brace College Publishers, New York (1992)
10. Meyer, R.E., Moreno, R.: Aids to computer-based multimedia learning. Learn. Instr. **12**, 107–119 (2002)
11. Lohr, L.L.: Designing the instructional interface. Comput. Hum. Behav. **16**, 161–182 (2000)
12. Jonassen, D.H., Tessmer, M., Hannum, W.H.: Task Analysis Methods for Instructional Design. Lawrence Erlbaum, Mahwah (1999)
13. Haladyna, T.M.: Developing and Validating Multiple-Choice Test Items, pp. 125–141. Lawrence Erlbaum Associates, Mahwah (1999)

Control Using Multimodal Input

Challenges of Using Gestures in Multimodal HMI for Unmanned Mission Planning

Meghan Chandarana[1](✉), Erica L. Meszaros[2], Anna Trujillo[3], and B. Danette Allen[3]

[1] Carnegie Mellon University, Pittsburgh, PA, USA
mchandar@cmu.edu
[2] Eastern Michigan University, Ypsilanti, MI, USA
ewicks@emich.edu
[3] NASA Langley Research Center, Hampton, VA, USA
{a.c.trujillo,danette.allen}@nasa.gov

Abstract. As the use of autonomous systems continues to proliferate, their user base has transitioned from one primarily comprised of pilots and engineers with knowledge of the low level systems and algorithms to non-expert UAV users like scientists. This shift has highlighted the need to develop more intuitive and easy-to-use interfaces such that the strengths of the autonomous system can still be utilized without requiring any prior knowledge about the complexities of running such a system. Gesture-based natural language interfaces have emerged as a promising new alternative input modality. While on their own gesture-based interfaces can build general descriptions of desired inputs (e.g., flight path shapes), it is difficult to define more specific information (e.g., lengths, radii, height) while simultaneously preserving the intuitiveness of the interface. In order to assuage this issue, multimodal interfaces that integrate both gesture and speech can be used. These interfaces are intended to model typical human-human communication patterns which supplement gestures with speech. However, challenges arise when integrating gestures into a multimodal HMI architecture such as user perception of their ability vs. actual performance, system feedback, synchronization between input modalities, and the bounds on gesture execution requirements. We discuss these challenges, their possible causes and provide suggestions for mitigating these issues in the design of future multimodal interfaces. Although this paper discusses these challenges in the context of unmanned aerial vehicle mission planning, similar issues and solutions can be extended to unmanned ground and underwater missions.

Keywords: Gesture · Multimodal · Unmanned

1 Introduction

As new applications for unmanned aerial vehicles (UAV) systems continue to be developed, the user base shifts from one of highly trained pilots and engineers to one of unskilled, non-expert users [1]. Current interaction schemes require operators to have domain knowledge of low-level system requirements and parameters. As such only

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_17

highly trained operators are able to effectively and easily use these interfaces [2]. Therefore, new – more intuitive – interaction methods must be developed.

Interfaces that make use of natural human-human communication modalities like speech and gesture are more effective and extensible to a broader user base [3–5]. Gestures are the most widely explored natural language input modality used in next generation human-robot interfaces. This can be attributed to their extensive use by humans to relay complex concepts to one another [6]. Dynamic hand gestures have been used to define general shapes [1] and relative position of vehicles [7]. Static hand gestures have been used to program by demonstration [8, 9] and define complex indirect movement [10, 11].

This paper provides an overview of several UAV applications and discusses the benefits of using a UAV system in the context of each application (Sect. 2). Section 3 then examines the use of natural language based multimodal interfaces for human-UAV interaction. The challenges associated with integrating commonly used gesture-based inputs are explored and possible methods for resolving these issues are identified (Sect. 3.1). Although these challenges and mitigation techniques explored in the context of UAV missions they can be generalized to other unmanned vehicle missions such as ground and underwater. Section 4 provides concluding remarks.

2 Unmanned Mission Planning

Over time the capabilities of UAV systems have expanded and evolved. These capabilities are leveraged to develop methods and procedures for various applications.

Although the vehicle platform itself can be similar between applications, each mission requires a unique sensor payload and planning algorithm to achieve the intended goal. The user base's expertise and background knowledge of the UAV system may also vary between applications. Future UAV interfaces must be accessible to this wide variety of users and robust enough to bridge the gap between given inputs and required system level parameters. The following subsections will discuss four of the applications that have emerged: (1) atmospheric science missions, (2) disaster relief, (3) search and rescue, and (4) intelligence, surveillance and reconnaissance.

2.1 Atmospheric Science Missions

Traditionally atmospheric scientists attach a sensor payload like an ozone sensor (Fig. 1) to a weather balloon, satellite or manned aircraft. These missions require long planning periods and necessitate trained engineers and/or pilots to launch or fly the collection apparatus. Due to the uncertainty in factors such as weather conditions, data collection can be unpredictable. In many instances the sensor or even the apparatus itself is lost or damaged [13]. The current collection methods limit the scale and ability to take more complex data (e.g., correlative measurements at different locations).

Fig. 1. Ozone sensor used by scientists to collect atmospheric data.

Recently, scientists have been exploring the use of multiple coordinated UAVs for data collection [14]. These vehicles will allow for faster and more complex data collection and diminish the risk of lost or damaged sensors. By outfitting each UAV with a different sensor payload or deploying each UAV to a different location, various types of data can be collected simultaneously in situ. This correlative data collection scheme provides a method for scientists to design and conduct more sophisticated environmental studies.

2.2 Disaster Relief

Directly following a catastrophic natural or man-made disaster, first responders and emergency aid organizations must efficiently distribute a large number of common commodities to populations in need. These commodities range from clothes and food to medicine (Fig. 2) and emergency personnel themselves [15]. Often environmental conditions and transportation and resource limits make it difficult for aid organizations to distribute supplies quickly and efficiently to those in need. Aid organizations typically rely heavily on the availability of transportation resources from local governments and military bases [16].

Fig. 2. Distribution of resources given by disaster relief aid workers in the aftermath of Typhoon Haiyan in 2013 in the Philippines [12].

Current disaster relief and emergency response vehicles are manned and include ground vehicles and conventional fixed-wing and rotorcraft aerial vehicles [17]. Autonomous UAV systems are a viable alternative that reduce the reliance on human operators and pilots [12]. The diversity in size of UAVs available provides a way to increase the efficiency in distribution by allowing the flexibility to tailor the payloads to the actual need of the destination area. This diminishes wasted space, resources and time.

2.3 Search and Rescue

Search and rescue personnel are tasked with looking for people as a result of a variety of causes. These include and are not limited to getting trapped due to natural disasters such as earthquakes and hurricanes or getting lost or injured. The search area may include naturally occurring environments (e.g., forests or oceans) and/or urban and crumbling infrastructure. This is often time sensitive, dangerous, and requires search personnel to make their way through areas with limited access. Often times dim lighting, fallen debris, and dust make it difficult to for search personnel to search [18].

In the context of a search and rescue mission the use of a UAV system can mitigate some of the common issues that arise due to a conventional human-based search. One of the primary benefits for using a UAV system is its ability to carry a sensor payload. These payloads along with the vehicle's long hovering capabilities are able to inspect search areas more thoroughly than humans [19]. In addition, UAVs are capable of operating in environments which may be unreachable or challenging to reach by human counterparts.

2.4 Intelligence, Surveillance and Reconnaissance

Intelligence, surveillance and reconnaissance (ISR) systems must be able to locate and track targets for extended periods of time in challenging environments. These environments can range anywhere from indoor to outdoor settings. As such, these missions require highly trained personnel to accomplish the task either on foot or by operating a vehicle which often times is equipped with additional sensors [20]. These current methods are costly in terms of human time and are often times dangerous. As in the case of search and rescue, UAVs' ability to carry multi-sensor payloads and reach traditionally inaccessible to difficult to access areas makes them a viable alternative for ISR missions [21].

3 Multimodal Interfaces

Conventional human-UAV interaction schemes employ the user of a mouse and keyboard system. Recently, researchers have explored methods for more intuitive human-UAV interaction via natural language based interfaces. These methods attempt to mimic the communication modalities seen in typical human-human communication – such as speech and gesture – thereby making the system more accessible to non-expert operators [7].

By utilizing multiple input modalities simultaneously, multimodal interfaces leverage the strengths of each individual input modality. This allows operators the ability to easily and naturally develop more complex UAV mission plans without being hindered with low-level parameter and system definitions. For example, in the case of trajectory generation planning a multimodal interface which includes both speech and gesture input modalities gives operators a method for defining the shape (gesture component) and geometric information (speech component) – distance, radius, height, etc.

3.1 Gesture Integration Challenges

Although humans use multiple modalities to communication with each other, mimicking the natural integration of various modalities is often non-trivial. Without identifying and mitigating these issues, multimodal interfaces may see poor overall performance. In addition, human operators may feel increased workloads even though all input modalities used are commonly seen in human-human communication. The remainder of this section will explore four challenges seen when integrating gestures into a multimodal framework: (1) differences between user perceived performance versus their actual performance, (2) developing appropriate system feedback for the operator, (3) synchronization between the input modalities, (4) and identifying the requirements and bounds of gesture inputs to human operators.

User Perceived Performance. Despite the frequency of gestures seen in human-human interactions, an obvious learning curve is observed for human operators using artificially constructed gesture-based human-machine interfaces. Even in the case of interfaces where the gesture inputs themselves are deemed "natural" feeling by operators, operators' interactions with the system are often seen as robotic and rigid. Given their high familiarity with gestures, human operators are generally very critical of any misinterpreted inputs made when learning multimodal human-machine interfaces that utilize gesture inputs. This overly critical attitude towards any mistake made produces a gap between operators' perceived measure of their performance versus the actual measure of their performance. Specifically, our previous work found that although subjects got an average defined 74.36% of trajectory segments correctly using a gesture-based interface [2], they rated their performance as 5.62 (where 0 was good performance and 10 was poor). In most cases, the overall accuracy is fairly high compared to the amount of time they have spent training on the multimodal system.

In order to shrink the gap seen between the perceived performance and actual performance, more feedback must be given to human operators when training on the multimodal system. A more accurate representation of their accuracy throughout the training process will diminish frustrations by visually demonstrating that their overall accuracies are relatively high compared with the amount of time spent training. Since this added feedback may overload the user after they are familiar with the interface, future multimodal interfaces can benefit from developing a separate training system.

System State Feedback. Humans easily outperform computers/machines when interpreting gestures. While using multimodal interfaces this translates into human operators expecting the system to immediately recognize their gesture inputs. This often leads to operators performing additional gestures before the system has finished interpreting their previous input. For example, subjects added an extra trajectory segment 11.11% of the time when using our gesture-based interface [2], which may have been due to the lack of feedback from the system. Therefore, more system feedback to indicate the current state of the system is required when using the multimodal interface.

More specifically, visual and/or auditory feedback can be used to reflect the system's current state to the operator. This will allow human operators to better pace themselves when giving gesture inputs. By learning the overall pace of the system, operators will be able to increase their overall accuracy in defining intended mission plans.

Synchronization. One of the most common challenges seen when integrating gestures with other input modalities is the synchronization between the two inputs. In order to develop a general interface that is usable by a broad range of operators, multimodal interfaces must make assumptions about the overall structure of system inputs. This structure includes the order of input and length of time required for each input. As these assumptions are developed for the largest amount of operators to effectively use the interface, they may differ slightly from each individual human operator's assumption.

Differences between the operator's assumed input synchronization and the actual system's can be mitigated by developing a method for the system to learn a model for each operator's assumed synchronization. Simple machine learning techniques can be applied to create a library of customized synchronization parameters. Although this will require an additional setup step before each operator can effectively use the multimodal interface, it may lead to lower overall operator workload and higher accuracy.

Gesture Execution Bounds. The sensing range and assumptions made about given inputs for each sensor is unique. In most cases operators will not be familiar with the specific sensor used in a given multimodal interface. Therefore, a significant aspect of the training is operators' attempt to learn the sensing range and general characteristics of a successful gesture input for the given sensor. Human operators typically use a generic trial and error method to develop a mental model of the sensor's sensing range and assumptions.

If the execution bounds can be accurately represented and presented to the human operator, a higher percentage of their training time can be utilized for improving their overall performance. Various simple methods can be employed such as (1) creating a physical representation of the bounds on the surface the sensor is mounted or placed or (2) providing visual feedback on the screen for acceptable hand position versus those outside the sensing boundary. Future multimodal interfaces can even include this visual feedback during general use in addition to during training.

4 Conclusion

Natural language based multimodal interfaces are a viable alternative for non-expert users to effectively define unmanned mission plans. These interfaces allow operators to define complex mission objectives without being required to understand the specific low-level system requirements. Given their broad use in human-human communications, a typical input modality used in multimodal interfaces is gestures. Although the gestures themselves may be intuitive, additional integration challenges can reduce users' overall performance and workload. Several of these challenges were identified and possible mitigation strategies were given for use in future multimodal interface iterations.

References

1. Chandarana, M., Trujillo, A., Shimada, K., Allen, B.D.: A natural interaction interface for UAVs using intuitive gesture recognition. In: Savage-Knepshield, P., Chen, J. (eds.) Advances in Human Factors in Robots and Unmanned Systems, pp. 387–398. Springer, Berlin (2017)
2. Chandarana, M., Meszaros, E., Trujillo, A., Allen, B.: Fly like this: Natural language interfaces for UAV mission planning. In: Proceedings of the 10th International Conference on Advances in Computer-Human Interaction. ThinkMind (2017)
3. Reitsema, J., Chun, W., Fong, T., Stiles, R.: Team-centered virtual interactive presence for adjustable autonomy. In: AIAA conference, Space 2005, p. 6606 (2005)
4. Wachs, J.P., Kölsch, M., Stern, H., Edan, Y.: Vision-based hand-gesture applications. Commun. ACM **54**(2), 60–71 (2011)
5. Perzanowski, D., Schultz, A.C., Adams, W., Marsh, E., Bugajska, M.: Building a multimodal human-robot interface. IEEE Intell. Syst. **16**(1), 16–21 (2001)

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. Fernández, R.A.S., Sanchez-Lopez, J.L., Sampedro, C., Bavle, H., Molina, M., Campoy, P.: Natural user interfaces for human-drone multi-modal interaction. In: 2016 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 1013–1022. IEEE (2016)
8. Becker, M., Kefalea, E., Maël, E., Von Der Malsburg, C., Pagel, M., Triesch, J., Vorbrüggen, J.C., Würtz, R.P., Zadel, S.: Gripsee: a gesture-controlled robot for object perception and manipulation. Auton. Robots **6**(2), 203–221 (1999)
9. Lambrecht, J., Kleinsorge, M., Krüger, J.: Markerless gesture-based motion control and programming of industrial robots. In: 2011 IEEE 16th Conference on Emerging Technologies & Factory Automation (ETFA), pp. 1–4. IEEE (2011)
10. Ende, T., et al.: A human-centered approach to robot gesture based communication within collaborative working processes. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3367–3374. IEEE (2011)
11. Raheja, J.L., Shyam, R., Kumar, U., Prasad, P.B.: Real-time robotic hand control using hand gestures. In: 2010 Second International Conference on Machine Learning and Computing (ICMLC), pp. 12–16. IEEE (2010)
12. Marks, P.: Smart software uses drones to plot disaster relief. https://www.newscientist.com/article/mg22029455-100-smart-softwareusesdrones-to-plot-disaster-relief/ (2013). Accessed Feb 2017
13. Wegener, S., Schoenung, S., Totah, J., Sullivan, D., Frank, J., Enomoto, F., Frost, C., Theodore, C.: UAV autonomous operations for airborne science missions. In: AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop and Exhibit, p. 6416 (2004)
14. Wegener, S., Schoenung, S.: Lessons learned from NASA UAV science demonstration program missions. In: 2nd AIAA" Unmanned Unlimited" Conference and Workshop and Exhibit, p. 6616 (2003)
15. Afshar, A., Haghani, A.: Modeling integrated supply chain logistics in real-time large-scale disaster relief operations. Soc. Econ. Plan. Sci. **46**(4), 327–338 (2012)
16. Long, D.: Logistics for disaster relief: engineering on the run. IIE Solut. **29**(6), 26–30 (1997)
17. DeBusk, W.: Unmanned aerial vehicle systems for disaster relief: Tornado alley. In: AIAA Infotech @ Aerospace 2010, p. 3506 (2010)
18. Waharte, S., Trigoni, N.: Supporting search and rescue operations with UAVs. In: 2010 International Conference on Emerging Security Technologies (EST), pp. 142–147. IEEE (2010)
19. Naidoo, Y., Stopforth, R., Bright, G.: Development of an UAV for search & rescue applications. In: AFRICON, 2011, pp. 1–6. IEEE (2011)
20. Semsch, E., Jakob, M., Pavlicek, D., Pechoucek, M.: Autonomous UAV surveillance in complex urban environments. In: IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09, vol. 2, pp. 82–85. IEEE (2009)
21. Nigam, N., Kroo, I.: Persistent surveillance using multiple unmanned air vehicles. In: Aerospace Conference, 2008 IEEE, pp. 1–14. IEEE (2008)

Compensating for Limitations in Speech-Based Natural Language Processing with Multimodal Interfaces in UAV Operation

Erica L. Meszaros[1](✉), Meghan Chandarana[2], Anna Trujillo[3], and B. Danette Allen[3]

[1] University of Chicago, Chicago, IL, USA
elmeszaros@uchicago.edu
[2] Carnegie Mellon University, Pittsburgh, PA, USA
mchandar@cmu.edu
[3] NASA Langley Research Center, Hampton, VA, USA
{a.c.trujillo,danette.allen}@nasa.gov

Abstract. Natural language interfaces are becoming more ubiquitous. By allowing for more natural communication, reducing the complexity of interacting with machines, and enabling non-expert users, these interfaces have found homes in numerous common products. However, these natural language interfaces still have great room for growth and development in order to better reflect human speech patterns. Intuitive speech communication is often accompanied by gestural information that is currently lacking from most speech interfaces. Exclusion of gestural data reduces a machine's ability to interpret deictic information and understand some semantic intent. To allow for truly intuitive communication between humans and machines, a natural language interface must understand not only speech but also gestural data. This paper will outline the limitations and restrictions of some of the most popular and common speech-only natural language processing algorithms and systems in use today. Focus will be given to extra-linguistic communication aspects, including gestural information. Current research trends will then be presented that have been designed to compensate for these gaps by incorporating extra-linguistic information. The success of each of these trends will then be evaluated, as well as the hopefulness of continued investigative efforts. Additionally, a model multimodal interface will be presented that incorporates language and gesture data in order to demonstrate the effectiveness of such an interface. The gestural portion of this interface is included to compensate for some of the limitations of speech-only natural language interfaces. Combining these two types of natural language interfaces thereby works to reduce the limitations of natural language interfaces and increase their success. This presentation will discuss how the two interfaces work together and will specify how the speech interface limitations are addressed through the inclusion of a gestural system.

Keywords: Natural language processing · Speech interface · Multimodal

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_18

1 Introduction

Natural language interfaces have existed for decades, but with the increasing ubiquity of computers and the continuous search to innovate in interface design they have recently reached a new level of familiarity. Unfortunately, their familiarity is often the result of their failure, with incorrect results and interpretations leading to viral news stories. Despite bad publicity, the algorithms that run these interfaces are often powerful, intelligent, and intricate systems that have consistently improved over the years. Why do natural language processing (NLP) algorithms still produce sub-optimal results? To answer this question, the limitations of some of the most common and widely used NLP algorithms are examined within this paper. Focus is given to the deep learning algorithms that are most commonly in use today.

My analysis will focus on current limitations of these algorithms, including speed and algorithm runtime, hardware restrictions, and input data limitations. Current interfaces, such as the iPhone touchscreen or the traditional mouse and keyboard computer interface, rely on near instantaneous feedback to the user, with limited time between offering input to the system (i.e., a tap on the screen, a click of the mouse button) and a reaction from the system (i.e., selecting an app) [1]. Natural language interfaces, however, rely on hearing and interpreting speech before the computer can take any action. This adds additional time between giving input to the computer and receiving any response [2]. While none of this is inherently incorrect, and in fact mimics the same process humans go through when giving and receiving natural language input, the addition of lag time into the system reduces the usability of the systems as a whole and often prevents such interfaces from being used. Additionally, a natural language interface can only function as well as it is able to hear the natural language input, relying on a microphone to pick up speech signals. Distance from microphone, clarity of speech, and how well the microphone function all impact the quality of the signal and how well the algorithm is able to interpret the input.

However, the overall goal is to determine whether there are any intrinsic restrictions to this type of interface that cannot be overcome with changes to algorithm structure, available hardware, and improved data. For example, natural language interfaces typically expect and accept only speech information. Humans, however, send a lot of information in addition to speech data when communicating. This is one of the reasons that talking on the phone can feel so different from talking face-to-face—a lack of information from gestural and body language signals often leaves speech data incomplete (see, for example, deictic pronouns) [3]. Furthermore, NLP algorithms may not be able to incorporate information on intonation into an analysis, whereas humans rely on this information to interpret questions, sarcasm, mood, veracity of information, etc. Can NLP algorithms parse this suprasegmental information? In a sense, this paper will attempt to tackle the idea of language as a human product and tool, and address whether there are aspects of language that remain outside the mechanical understanding of even increasingly autonomous, learning algorithms.

This paper presents interviews with experts from academia and industry as well as students in the field of computational linguistics that have worked with such algorithms to determine what limitations and restrictions they have met. These conversations will provide the information for the bulk of this analysis, allowing for a determination of the capabilities and limitations of NLP algorithms based on their use in real world research and application.

2 Overview of NLP Research

Understanding how the field of NLP has changed over the decades provides a basic introduction to some of the key algorithm groups available to researchers today. NLP lies at the "intersection of artificial intelligence and linguistics" [4], a field dedicated to allowing machines to interpret and produce human language. Because NLP research has been carried out for more than half a century now, many different types of algorithms have been developed and popularized to accomplish this understanding. Early NLP algorithms progressed the field through the creation of rules based on the grammatical structure of language to guide machine understanding of language, providing a type of top-down guidance for understanding language [5]. Statistical Linguistics has since played a significant role in the history of NLP to enable a bottom-up type of understanding that imitates the learning process encountered within the human brain [5]. Statistical NLP uses machine learning algorithms, such as neural nets or Hidden Markov Models, to allow systems to develop their own rules from correctly tagged and identified human language [6]. Without imposing rules from the outside, machines can generate statistically-based rules on their own that generally out-perform rule-based systems. Current trends have attempted to combine the success of statistical language processing with the earlier rule-based understanding to generate a more thorough and successful language understanding system.

While many different algorithms have been used in NLP, recent focus in the field has almost unanimously landed on deep learning. Originating from advancements in computer vision research, deep learning has been adopted by NLP researchers only within the last few years. Deep learning makes use of non-linear information learning structured hierarchically in many layers in order to understand several levels of representation and abstraction to make sense of language [7]. Frequently, deep learning algorithms are built around neural networks, a type of machine learning algorithm modelled on human brain function and architecture to enable systems to learn through observation [8]. In addition to neural networks, Hidden Markov Models have been popularized as a type of deep learning algorithm. Hidden Markov Models have frequently been used for part of speech taggers, content classification, and sentiment analysis.

All of the experts interviewed for this paper identified deep learning and big data processing as the current focus of research in NLP, including David McAllester, Professor and Chief Academic Officer at the Toyota Technological Institute at Chicago. McAllester says that this focus has changed significantly from the popular algorithms in use just five years ago and that the degree to which deep learning has become involved in linguistics research has been unexpected [9]. This sentiment has been echoed by Richard Sproat, a Research Scientist with Google. Sproat describes the field before deep learning as utilizing "a much more varied set of machine learning tools," but describes everyone in the field as committed to the use of deep learning now [10]. Christopher Brew, a Computational

Scientist with Digital Operatives, says that the “deep learning trend is here to stay and will continue to thrive” [11], while Kristy Hollingshead Seitz, a Research Scientist at the Florida Institute for Human & Machine Cognition, attributes recent significant progress in most aspects of NLP to deep learning [12].

3 Low-Level Limitations

In prior research, focus has been given towards only low-level limitations of these NLP algorithms [13]. Low level issues include mechanical restrictions, such as available data or hardware, and tend to have clearer paths towards overcoming or compensating. Low-level restrictions are not easier to tackle than higher level limitations intrinsically, and in fact pose some of the most complex issues in NLP research today. As such, these low-level limitations must be understood in order to progress in the field. In order to update understanding of these, limitations most frequently referenced in interviews with experts in the field are presented in this section.

Because language is a complex phenomenon, it causes problems in machine interpretation of language. Most NLP algorithms focus on processing speech or text data by using words as a unit of analysis. Some algorithms used for speech-to-text analysis focus on phonemes, but these are ultimately used to interpret full words. It cannot be denied that words are important to language, and this focus has paved the way for the progress that the field has seen in the recent decade: Systems, like Siri and Alexa, that are competent and capable. However, focusing on words reduces the complexity of language substantially and leaves important meaningful information out of the analysis. One such unexamined aspect is intonation, used frequently in human conversation to convey questions, emotions such as anger, excitement, and even joy [14], as well as sarcasm and irony [15]. Similar to intonation, syllabic or word-based stress can be used to emphasize important semantic aspects of a sentence [16]. Such information is also currently unexamined by NLP algorithms. A further unexamined aspect is gestural information; frequently visual information like pointing or head-nodding is included with deictic pronouns like “this” and “that” in order to help specify the referenced object. Divorced from this additional extra-linguistic information, the pronouns alone cannot necessarily interpret which object is being referenced.

Intonation, stress, and gestural information form an important part of language but are often overlooked in NLP algorithms. Intuitive human-to-human communication relies on this type of information [17]. Damir Cavar, Associate Professor of Computational Linguistics at Indiana University Bloomington, further specifies that “intonation is the only way for you to detect deception, irony, sarcasm,” underlining the importance of machine interpretation of this information [18]. However, this data cannot be determined from a textual analysis and has to be obtained from the acoustic signal, in the case of intonation and stress, or from video, in the case of gestures. Bringing together acoustic and video data with speech information, Cavar describes, will allow linguists “to come up with an analysis of pragmatics and real semantics,” further enhancing the ability of machines to understanding and interpret language [18]. This type of analysis is difficult because, as Brew describes, “it requires a lot of different resources to work together across the whole spectrum of linguistics and engineering" [11]. Researchers in linguistics and computer science have to work together with mechanical engineers and even cognitive scientists to correctly model and implement audio and video signal processing and interpret the combined information correctly. However, he further specifies that despite this difficulty, including analysis of intonation, stress, and gesture is terribly important.

Perhaps it is not only the active collaboration of many different fields of linguistics and inclusion of text, acoustic, and video signals that has so far prevented machine interpretation of this suprasegmental information. Linguists seem to agree that this information is important for human communication, that it contributes valuable data to our understanding and that limiting communication only to the words being said is, therefore, necessarily limited. Without full understanding of these intonations, stress, and gestures, however, full interpretation of this important information may not be possible.

Another limitation of NLP algorithms identified by experts is their lack of ability to adapt. While collecting data on multimodal natural language interfaces, researchers noted how the speech behavior of participants changed when they were interacting with machines instead of humans [19]. Participants would engage in conversation with the human researchers in colloquial speech patterns, including sentence fragments, self-corrections, conversational fillers, etc. [20]. When participants began to give verbal instructions to the machine, however, these colloquial patterns fell away. While the grammar and vocabulary understood by the machine were limited and dictionaries were provided for the participant, users' pronunciation patterns changed to include over enunciation of consonants increased substantially (e.g., the "t" at the end of "right" was hit harder and held longer than in normal colloquial speech patterns). This is interesting for at least two separate reasons: Initially, this suggests that humans intuitively want to interact verbally with machines in a different manner than humans, using different speech patterns and pronunciations. Critically, however, the system's NLP algorithms had trained on colloquial speech patterns, so the over-enunciation was unexpected and reduced the overall success of the system.

This behavior is a result of one recurrent limitation encountered by many NLP algorithms: human adaptability. One aspect of this adaptability is visible when subjects adapt their speaking patterns to what they think the machine is most likely to understand, as in the over-enunciation described above. Another aspect of this is our ability to understand text written in strange ways (in a forced dialect, as in Huckleberry Finn, or internet l33t (i.e., "leet") speak), impaired speech, and even normal conversation flow that often includes self-interruptions or changes in thought and generation of new and unique words [13]. According to Sproat, the machine "would just break down on that kind of stuff" because it is not currently capable of understanding anything outside of the normal range of variation [10]. "Humans are really adaptable," describes Hollinshead Seitz, "we'll figure out what are the rules, we'll play by the rules because it will make something work well for us…computers have a really really really hard time with that." Bates et al. suggest that the limited discourse capabilities of current NLP systems can be overcome by human adaptability, which we do effectively [13]. Our innate human ability to adapt has impacted how NLP algorithms are realized, but has also limited our ability to model human language in machines. Adapting to unusual language, claims Sproat, is "one big challenge that really I don't think anyone is thinking about too much" [10].

In addition to the limitations of suprasegmental information like intonation and gestures and the problem of human adaptability, there is a very basic limitation that is a result of the deep learning algorithms that are currently popular within the field. While often the results that deep learning algorithms produce are fairly accurate and powerful, they are frequently uninterpretable. Cavar claims that these algorithms are basically black boxes, that "you don't have transparency over the model—you don't know why it's doing certain things, you don't actually even understand what it learned" [18]. For example, a visual deep learning algorithm could be trained over a dataset containing images of dogs and cats, and can be trained to easily distinguish these animals. Can it then take the set of statistical rules it has developed and apply them to a dataset with completely different types of objects, like coffee mugs? "The answer to that seems to be clearly yes," says McAllester. "These deep networks have learned more than just the categories that they were trained on—they've learned some general representation…but we don't know what it is" [9]. Understanding what the system has learned, therefore, about dogs and cats that is applicable to identifying coffee mugs is not possible. The lack of transparency, our inability to interpret the results of deep learning algorithms, may not inherently be a limitation. The algorithms work, after all, and produce the desired results. The larger problem here is trust. "Any bit of information that's given to us," claims Hollingshead Seitz, "we want to know the source, we want to know whether we can trust it." Without knowing why the algorithm has come to its conclusion, we cannot fully trust its conclusion. This may seem superficial when the system is deciding whether an object is a cat or a coffee mug, or which antecedent a pronoun refers to, but becomes significantly more problematic if it is instead deciding how to respond to a life or death situation in a self-driving car. "There's this sort of level of trust between humans and we want to be able to have that with any kind of artificial intelligence as well," says Hollingshead Seitz. This trust is absent with deep learning algorithms due to their lack of transparency and interpretability. Instead, says Hollingshead Seitz, "you just have to see how often it's right and you don't get to know why it thinks that." Algorithmic interpretability is not only a problem within NLP, however, and seems poised to be at the forefront of ethics debates regarding the use of algorithms, heuristics, and the application of artificially intelligent (or even increasingly autonomous) systems. As NLP and artificial intelligence algorithms begin to play an even larger role in our daily lives, this question of interpretability and trust becomes even more important.

One of the biggest problems for current NLP algorithms, especially with the recent focus on deep learning, is data. To learn and develop new rules, deep learning algorithms may require "at last a hundred thousand instances" of a language phenomenon, according to McAllester [9]. Sproat calls these deep learning methods "data hungry," which can be problematic for a number of reasons [10]. It can be difficult to get humans to annotate enough data, according to McAllester, and without annotation the machine may not understand enough about the language to develop any useful rules [9]. Gershgorn describes the laborious task of preparing data for the generation of a lip-reading NLP algorithm, where the direction the actor was facing, amount of lighting, and even the sentence structure all had to be standardize. "When other researchers," he writes, "pointed out that using such specialized training videos weren't applicable to real-world results, author Nando de Freitas defended the results of his paper, noting that other video sets the team tried were too noisy. The other videos they tried were each too different from the last for the AI to draw meaningful conclusions—meaning a perfect data set just doesn't exist yet" [21]. Moreover, even before annotation, finding enough data at all may be difficult—"if you're trying to model something very specific or very individual, like you're trying to model one person's language and how that language changes over time, or you're trying to model the language of three to four-year-old children and maybe you have 10 of them or even a hundred of them, that's not enough for deep learning" claims Hollingshead Seitz. Sproat suggests that while there may not be any aspect of language that deep learning cannot tackle, it is largely a matter of whether there is enough data for deep learning to approach the problem [10].

Perhaps some of the most surmountable low-level limitations for NLP algorithms are related to their speed and their size. In order for any natural language interface to be usable, it has to interpret human language in about as much time as it would take a human. One Linguistic Technologist from Sensory, Inc. described the algorithm she was currently working on as "one of the fastest algorithms for syntactic parsing," saying that "it runs in $O(N)^3$ time…and that's considered like pretty good" [22]. This means that the time performance of her NLP algorithm is directly proportional to the cube of the size of the input data set, so with every new piece of input data the time it takes to run the algorithm will increase three times. "So one big limitation," she claims, "is the runtime of all of these algorithms." Not only does it take a long time to run these algorithms, but it takes a long time to train them as well. The Linguistic Technologist claims that it "takes forever to train a model…the more data you have, just the longer it takes." Moreover, as human language models become more advanced and more intricate, they also take up more room. Sensory, Inc. would "train these acoustic models and then we would have to have the whole model fit on the device," claims one Linguistic Technologist. While their runtime recognition code was relatively small and took up little space, these trained models containing proprietary parameters and weights were pretty big. Ensuring that entire NLP systems can fit on pocket-sized devices, then, is another problem currently being tackled by the field. Constant improvements to algorithms as well as to computer hardware make the size and time restrictions on NLP algorithms limitations that, while certainly impactful, are likely able to be overcome.

Together, researchers in the fields of computational linguistics, computer science, and even cognitive science are working to address the other, less-surmountable limitations of NLP algorithms. Currently work is being carried out at the Toyota Technology Institute on machine analysis and understanding of intonation and stress, according to Cavar [18]. While some NLP algorithms still run slowly, improvements to hardware and software alike and access to large scale processing servers through companies like Amazon and Google are allowing for speed reductions to within acceptable limits for human users. But while progress is being made in accounting for such low-level limitations, there are other restrictions that plague NLP algorithms at a much higher level.

4 High-Level Limitations

One of these high-level limitations is a lack of a full understanding of many aspects of linguistics. This limitation played a role in accounting for suprasegmental features, such as gesture and intonation, but plagues NLP research at an even more fundamental level as well. As Cavar claims, "the real limitations are actually that we just don't understand, for example, semantics, meaning, intentionality…nobody knows what consciousness is and why people do things, you know, and so programming and imitating that is just highly difficult" [18]. Basic aspects of language, such as semantics and what language means, have been examined for years, and attempts to model such aspects within machines are necessary for NLP systems. However, these attempts are significantly hindered by our lack of understanding of the very thing we are attempting to model. McAllester even questions our approach to understanding semantics, saying that "semantics is about the relationship between language and reality, so the question is what is reality?" [9]. Sensory, Inc.'s Linguistic Technologist claims that "maybe the day that we define what it means to be conscious is the day that we are able to create machines that are conscious," but until that day the best that we can hope for are machines that still surprise us when they work correctly.

However, at the other end of the spectrum, the application of too much knowledge has also proven problematic within NLP. Systems that performed well using statistical language modelling often lacked any applied language rules to guide their understanding. Linguists would often try to apply language rules to the systems, because intuitively this is how language works, but the application of these rules actually caused the systems to worsen. As McAllester summarizes, "it's this strange thing that's happened in AI for decades… that you think you see all this structure and you try to use it and it only hurts the performance of the system" [9]. The reduction in performance with the application of language-based rules is perhaps best described, however, in Frederick Jelinek's declaration "anytime a linguist leaves the group the recognition rate goes up" [23]. The more we understand about language, and the more of this understanding we try to give NLP systems, the worse these systems tend to perform.

There are still, however, areas of linguistics that we do not currently know enough about, and this lack of knowledge is significantly hindering the progress of NLP algorithms. Unfortunately, these areas include weighty questions like "what is consciousness?" and "what does it mean to have understanding?" that have been the focus of philosophical debate for millennia. While current research is underway to tackle just these questions, the answers will not be found easily. Additionally, the application of too much knowledge has proven troublesome to NLP algorithms, and limiting statistical processing with too many rules has actually worked to reduce the success of machine translation software. McAllester has not yet given up hope, however, saying that he's "completely unwilling to give up on the idea that there are things and there are relationships between them and that's really what language is talking about" [9]. Brew identifies this lack of understanding of how we as humans make language as one of the "fundamental limitations" [11]. Many researchers think that early attempts to bound machine understanding with these rules were clumsy compared to more nuanced methods available today, and that returning to this structured idea in conjunction with statistical learning and in the form of neural networks may prove even more successful than either method independently.

However, there is, one further limitation of NLP algorithms lying beneath even these ones. Multiple experts in the field were asked whether machines would ever fully understand human language, and was met on more than one occasion with a laugh. It turns out that there is vast disagreement within the field of computational linguistics on whether this is possible, and much of it hinges upon the definition of “understanding.” As Sproat asks, “you had asked originally a higher-level question on natural language understanding and where that’s going to be in a number of years and are we ever going to get to full understanding, so I guess my question for you would be what do you mean by understanding?” [10]. This question has been at the heart of NLP research since the beginning, when Alan Turing himself declared that his work on determining whether machines can think “should begin with definitions of the meaning of the terms ‘machine’ and ‘think’” [24]. While current research in cognitive science and even philosophy are attempting to answer this question, without full comprehension of what it means to understand language and not just be programmed to reply to language it is unlikely we will ever be able to reproduce this phenomenon in machines. Brew further identifies this lack of understanding of what we are trying to accomplish, saying that “if you can tell me precisely what human capability is I would be really really surprised” [11]. Even Hollingshead Seitz says that our ability to enable machine understanding with NLP algorithms “depends on what you mean by understanding.”

One aspect of this problem identified by many of the experts was embedding. Cavar says embodiment in the real world may be a prerequisite to understanding, and that some experts claim that “you can understand the world only if you’re part of it and the computer and algorithm is never part of it” [18]. Many companies are trying to bypass the issue of embodiment by providing the machine with enough data about the real world that it does not matter whether the machine is actually a part of it. This strategy is problematic due to the sheer amount of data required, however; Cavar says that while companies like Google claim to just need more memory and to collect more information about the world, the problem is that at its heart it is a “never-ending story and the amount of data it just demands, you can’t cope with that” [18]. Brew and Sproat also identify embodiment as a limitation preventing NLP algorithms from ultimately progressing all the way to machine understanding of language [10, 11].

Despite these fundamental limitations, McAllester predicts that NLP algorithm advancements will progress significantly in the coming years, bypassing the limitations described in this paper and achieving full machine understanding of human language. He gives this a 30–50% chance of happening in the next five years [9]. Brew is similarly an optimist, claiming that we are already starting to see machines that humans prefer interacting with over other humans. While he readily admits that full understanding may not happen, he thinks we are “already quite surprisingly far along” and is optimistic about the future [11]. Others are less ready to make predictions—Sproat claims that he is hesitant to make any predictions because “things have changed so much and because well you know…predictions don’t have a very good status in the world anymore” [10]. Despite this, he suggests that recent improvements have helped to whittle down the remaining problems and barriers to machine understanding, but identifies a core of understanding that we are still not getting at. Because of this, he claims that he is skeptical that we will ever reach a full level of machine understanding. Still other experts claim that this understanding will never happen, due to lack of embodiment in the real world or the amount of data that full understanding of language requires, or our lack of understanding of what language actually is.

5 Multimodal Interface

One method for compensating for many of the low- and high-level limitations of speech-based NLP is to expand beyond only speech. Natural language communication occurs through numerous extra-verbal methods, including the previously discussed deixis. Incorporating gestural information to the data analysis may provide for a more intuitive natural language interface and compensate for some of the limitations described above.

One area where a multimodal interface speaks directly to current limitations of speech-based NLP systems is incorporating suprasegmental information. Incorporating gestural information allows a system to better interpret deictic statements such as “put that there” [25]. Current research efforts by the authors have been focused on combining gestural and verbal information in an intuitive manner to allow users to instruct and operate unmanned aerial vehicles (UAVs). Implementing such a multimodal interface is difficult because it requires not only building a system that can correctly interpret multiple different kinds of input signals but also because it has to do so in a way that is intuitive to the user. While natural communication frequently utilizes gestures and speech in conjunction, forcing their combination in a non-intuitive manner can actually reduce success and increase complexity. In recent research, users who had the opportunity to test an initial multimodal interface indicated that they were neutral about using such an interface in the future, indicating that the difference in effectiveness of the gesture and speech aspects of the interface often made operation more trouble than it was worth [25].

Incorporating gestural data into a multimodal interface can also help a machine cope with the problem of human adaptability. While current systems are often flummoxed by over-enunciation, unusual dialects, and the complexities of normal speech patterns, incorporating a second set of data can provide a method of error correction for the system. If the speech data are ambiguous, simultaneous gestural data can provide a clue as to the correct interpretation. For example, if a user wants to tell the machine to move over to the left ten feet, the combination of dialect, enunciation patterns, and noisy environment might result in a phrase that could either be interpreted as “move over left ten feet” or “move forward left ten feet.” However, if this phrase was accompanied by a simultaneous gesture pointing not forward left but rather just left, the machine could disambiguate between the choices.

Regarding the specified high-level limitations, incorporating gestural data into a natural language interface helps to compensate for the embodiment issue. By providing additional data of how communication occurs between humans, a multimodal interface provides additional and necessary information to the machine interface. This in turns helps the system to better interpret natural language commands and interactions, resulting in a more intuitive interface. However, incorporating this data does increase the amount of data processing necessary for the interface, further increasing the data processing issue already plaguing NLP researchers. Nevertheless, with an accurately specified and limited dictionary and grammar of acceptable and expected speech and gesture components, the resulting multimodal natural language interface can help provide a better picture of human communication without overburdening the system with data.

Incorporating gestural information to create a multimodal natural language interface cannot compensate for all of the limitations in speech-based systems. Issues of understanding language, imitating consciousness, and interpreting deep learning advancements are still at the heart of many research questions. However, combining speech and gesture information can help to mitigate other limitations, resulting in a more robust and intuitive natural language interface.

6 Conclusion

This paper has outlined numerous limitations faced by NLP algorithms, including suprasegmental information, the lack of necessary data, the interpretability problem, the issue of embodiment, and our lack of understanding what exactly we are asking machines to replicate. This paper also suggested ways in which a multimodal natural language interface, combining gestural information with speech information, could work to mitigate problems encountered due to the limitations of a speech-only interface. Many of these limitations are the focus of current research investigations within the field, but some are more difficult to overcome. While it seems likely that many of these limitations can be overcome, some may prevent NLP algorithms from ever fully allowing machines to interpret and understand human language. McAllester is a firm believer that "everything we can do can be done by computers" because he believes that humans are, in fact, machines. He also believes that NLP algorithms will progress to the point that machines will have full and true understanding of human language. But this has its own set of consequences: as McAllester says, this means that "presumably we're not the end of the story, presumably greater intelligence is possible" [9].

References

1. Saffer, D.: Designing Gestural Interfaces: Touchscreens and Interactive Devices. O'Reilly Media, Inc., Sebastopol (2008)
2. Becker, K.C.: Developing a Speech-Based Interface for Field Data Collection. Diss., Texas A&M University (2016)
3. McNeill, D.: Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, Chicago (1992)
4. Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W.: Natural language processing: an introduction. J. Am. Med. Inform. Assoc. **18**(5), 544–551 (2011)
5. Lewis-Kraus, G.: The Great A.I. Awakening. The New York Times Magazine. http://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html. Accessed 14 Dec 2016

6. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing, vol. 999. MIT Press, Cambridge (1999)
7. Deng, L., Dong, Y.: Deep learning: methods and applications. Found. Trends Signal Process. **7**(3–4), 197–387 (2014)
8. Nielsen, M.A.: Neural networks and deep learning, 2016. http://neuralnetworksanddeeplearning.com/. Accessed 21 Dec 2016
9. McAllester, D.: Interviewed by author. November 30, 2016
10. Sproat, R.: Interviewed by author. December 7, 2016
11. Brew, C.: Interviewed by author. December 1, 2016
12. Hollingshead Seitz, K.: Interviewed by author. December 11, 2016
13. Bates, M., Bobrow, R.J., Weischedel, R.M.: Critical challenges for natural language processing. In: Challenges in Natural Language Processing, pp. 3–34 (1993)
14. Bänziger, T., Scherer, K.R.: The role of intonation in emotional expressions. Speech Commun. **46**(3), 252–267 (2005)
15. Nakassis, C., Snedeker, J.: Beyond sarcasm: intonation and context as relational cues in children’s recognition of irony. In: Proceedings of the Twenty-Sixth Boston University Conference on Language Development. Cascadilla Press, Somerville, MA, pp. 429–440 (2002)
16. Liberman, M., Prince, A.: On stress and linguistic rhythm. Linguist. Inq. **8**(2), 249–336 (1977)
17. Bolt, R.A.: ‘Put-that-there’: voice and gesture at the graphics interface. In: Maybury, M.T., Wahlster, W. (eds.)Readings in Intelligent User Interfaces, pp. 19–28. Morgan Kaufmann Publishers Inc., San Francisco (1998)
18. Cavar, D.: Interviewed by author. November 8, 2016
19. Chandarana, M., et al.: A natural interaction interface for UAVs using intuitive gesture recognition. In: Savage-Knepshield, P., Chen, J. (eds.) Advances in Human Factors in Robots and Unmanned Systems, pp. 387–398. Springer International Publishing, Berlin (2017)
20. Bulyko, I., Ostendorf, M., Siu, M., Ng, T., Stolcke, A., Çetin, Ö.: Web resources for language modeling in conversational speech recognition. ACM Trans. Speech Lang. Process. **5**(1), 1 (2007)
21. Gershgorn, D.: Oxford University’s lip-reading AI is more accurate than humans, but still has a way to go. Quartz. http://qz.com/829041/oxford-lip-reading-artificial-intelligence/. Accessed 07 Nov 2016
22. Sensory, Inc. Linguist Technologist. Interviewed by author. December 11, 2016
23. Jurafsky, D., Martin, J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson, Upper Saddle River, NJ (2009)
24. Turing, A.M.: Computing machinery and intelligence. Mind **59**(236), 433–460 (1950)
25. Chandarana, M., et al.: Fly like this: Natural language interfaces for uav mission planning. In: Proceedings of the 10th International Conference on Advances in Computer-Human Interaction. ThinkMind (2017)

Crew Performance and Situation Awareness in Three UAS GCS Layouts

Derek McColl[1(✉)], Jean-Francois Gagnon[2], Simon Banbury[3], Robert Arrabito[1], Nada Pavlovic[1], Fred Williams[4], Mario Charron[4], and Ming Hou[1]

[1] Defence Research and Development Canada – Toronto Research Centre, Ottawa, ON, Canada
{Derek.McColl,Robert.Arrabito,Nada.Pavlovic,
Ming.Hou}@drdc-rddc.gc.ca

[2] Thales Canada, Research and Technology, Québec, QC, Canada
Jean-Francois.Gagnon@ca.thalesgroup.com

[3] C3 Human Factors Consulting Inc., Montreal, QC, Canada
Simon.Banbury@c3hf.com

[4] Department of National Defence, Ottawa, QC, Canada
{Fred.Williams2,Mario.Charron}@forces.gc.ca

Abstract. The aim of Canada's Joint Unmanned Surveillance and Target Acquisition System program is to acquire an Unmanned Aircraft System (UAS) for Royal Canadian Air Force's domestic and international operations. This UAS will be capable of complementing existing reconnaissance, surveillance, target acquisition, and engagement capabilities. Defence Research and Development Canada has developed a UAS ground control station simulator to investigate crew layouts, airworthiness certification, UAS crew training technologies and strategies, as well as novel human-machine interfaces. This paper investigates crew performance and situation awareness in three layouts of the ground control station: distributed, classroom, and boardroom. Preliminary results indicate that the boardroom layout outperforms the other layouts in terms of crew communication, teamwork and situation awareness.

Keywords: Unmanned Aircraft System · Ground control station · GCS layout · Crew performance · Situation awareness

1 Introduction

Unmanned Aircraft Systems (UAS) include the Unmanned Aircraft Vehicle (UAV), the ground control station (GCS), and the human crew. UAS are a low-risk force multiplier for military operations, allowing militaries to perform a variety of mission types (Intelligence, Surveillance, Reconnaissance (ISR), force protection, exercise support, etc.) without putting an aircrew's life at risk [1]. Canada's Joint Unmanned Surveillance and Target Acquisition System (JUSTAS) program's objective is to acquire UAS for the Royal Canadian Air Force to conduct both international and domestic operations [2]. The JUSTAS UAS will be able to perform ISR, target acquisition, and strike missions in all weather conditions [2]. The aim is for the UAS to increase Canada's domestic artic

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_19

and maritime awareness as well as supplement current ISR, target acquisition, and strike capabilities [2].

Typically, in current UAS operations the pilot and sensor operator are collocated while the intelligence personnel analyze the UAS sensor data from their respective headquarters. This spatial separation of UAS flight crew and the intelligence cells limits the communication among the teams, as the distributed teams tend to see themselves as independent from each other. As a result, the integration of various intelligence feeds (full motion video, electronic support measures, etc.) occurs after the mission and as such does not contribute to decision making in real time. Research has shown that it is difficult for distributed teams to create and keep cohesion [3, 4]. Additionally, collaborative technologies have not yet been able to alleviate these issues [5, 6].

The JUSTAS Concept of Operations [7], has defined the composition of a Canadian UAS crew to include the pilot and sensor operator as well as Canadian Armed Forces (CAF) members from the intelligence analyst community, specifically image analysts and electronic warfare analysts. By collocating the teams, it is expected that the crew will better exploit the information gained from the UAS sensors during the mission. This should then lead to an overall increase in UAS mission effectiveness.

However, even with collocated teams, it is still critical that the crew be organized into a room-scale workspace layout that facilitates successful completion of the mission objectives. A GCS layout is defined by the orientations and locations of the UAS crewmembers' workstations in a room. The objective of a layout design is to facilitate operator communication and interaction with each other while taking into account operator sensory limitations (e.g. a large physical separation makes it difficult for two operators to communicate face-to-face) [8]. Quality of communications and interactions between operators has a significant effect on teamwork and mission success [8]. Defence Research and Development Canada (DRDC) has been tasked to study various GCS layouts for the JUSTAS UAS crew in relation with mission effectiveness and team performance.

DRDC has developed a UAS GCS simulator to support the JUSTAS program. Specifically, DRDC is using this simulator to study JUSTAS UAS GCS crew layouts, airworthiness certification, UAS crew training technologies and strategies, as well as novel human-machine interfaces [9–12]. This paper presents the preliminary operator performance and situation awareness results of a Human Factors Engineering (HFE) trial performed by DRDC to investigate UAS crew performance in three different GCS layouts using a UAS GCS simulator.

2 GCS Simulator and Roles

DRDC's UAS GCS simulator, called Testbed for Integrated GCS Experimentation and Rehearsal (TIGER), includes six crew workstations: Air Vehicle Operator (AVO), Payload Operator (PO), Image Analyst (IMA-A), Image Reporter (IMA-R), Electronic Warfare Analyst (EW-A), and Electronic Warfare-Reporter (EW-R). TIGER was designed to be modular and re-configurable to allow for the repositioning of crew workstations to investigate crew performance in various GCS layouts.

The TIGER crew workstations can be divided into three cells. The first cell consists of the AVO and PO. The AVO is the aircraft pilot and crew commander. The crew commander ensures the UAS crew successfully completes mission objectives and has final responsibility for all weapon engagements. The PO's main function is to control the UAV sensor position and orientation. The second cell consists of the IMA-A and IMA-R. The IMA-A screens the UAV sensor video for potentially important events. The IMA-R reviews the events identified by the IMA-A, determines which information is important, and generates intelligence reports for the AVO and UAS tasking authority. The third cell consists of the EW-A and EW-R that together support the UAS mission with electronic support measures with a partition of responsibilities similar to the image analysts.

Each crew workstation has three 27" inch monitors positioned low on their desks, allowing line of sight over the monitors between co-located crewmembers. In addition to job-specific software for each crewmember, all of the workstations have a map display, communications software, an option to view the UAV sensor video feed, keyboard, mouse, and radio headset. The AVO and PO workstations also have joysticks and throttles for flight and sensor control.

TIGER includes five additional workstations: two experimenter stations for data collection and three white cell workstations for simulation control. The experimenter stations include tools for recording observations and collecting simulation data for analysis. The white cell stations consist of the Computer Generator Forces (CGF) operator, a Role Player (RP), and the Instructor Operating Station (IOS). The CGF operator controls the states of the entities that inhabit the simulation environment as observed through the UAV sensors. For example, if the TIGER crew is providing over-watch to a convoy, the CGF operator controls the appearance, route, and speed of the convoy. During experiments, an actor uses the RP station to simulate an entity outside the UAS

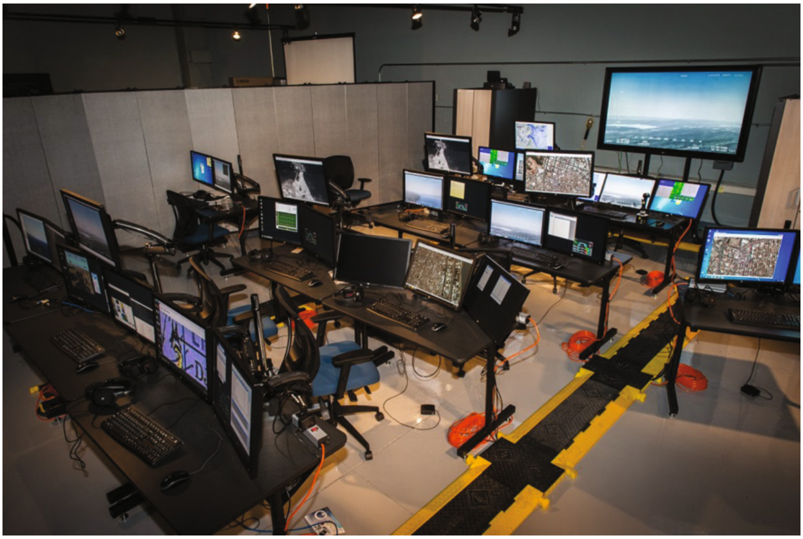


Fig. 1. TIGER at DRDC-Toronto Research Centre.

crew with whom the crew needs to communicate. The instructor uses the IOS station to observe crew performance during the mission as well as to coordinate actions between the crew, RP and CGF operator. TIGER is shown in Fig. 1.

3 GCS Layout Trial Methods

3.1 GCS Layouts

DRDC conducted a HFE trial, using TIGER, based on a within-subjects experimental design to investigate three different UAS GCS layouts: (1) distributed, (2) classroom, and (3) boardroom. Each GCS layout changes the physical location of each crew workstation, but does not make any changes to the function of the individual workstations.

1. *Distributed.* The distributed layout has each crew cell physically partitioned from each other, namely the AVO and PO in one location, IMA-A and IMA-R in another, and EW-A and EW-R in another, Fig. 2. The distributed layout simulates having the IMA and EW personnel located in their respective headquarters and the AVO and PO in theatre or in a third separate location.

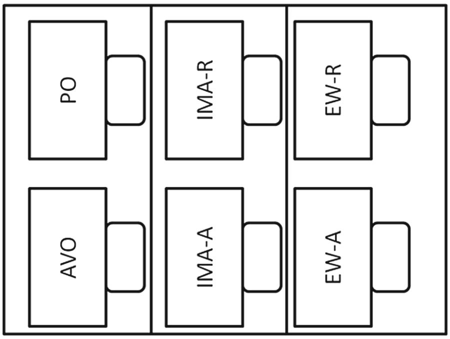


Fig. 2. Distributed GCS layout: each cell is physically separated from the others by partitions.

2. *Classroom.* The classroom layout has all six crew members in the same location, with the AVO and PO in the first row facing forward, the IMA-A and IMA-R in the second row facing the AVO and PO, and the EW-A and EW-R in the third row facing the IMA-A and IMA-R, Fig. 3. The classroom layout was used in previous TIGER HFE trials [10].

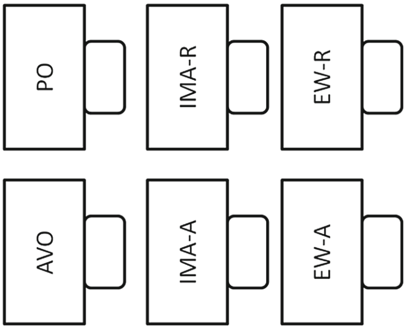


Fig. 3. Classroom GCS layout: the crew is collocated all facing the same direction.

3. *Boardroom.* The boardroom layout has the AVO and PO facing towards the other crewmembers who are all in a line facing the AVO and PO, Fig. 4. This layout was identified to meet crew interaction requirements based on a model that predicts required visual and verbal communications among crewmembers. This model was generated from an UAS crew cognitive task analysis [13] and Subject Matter Expert (SME) interviews.

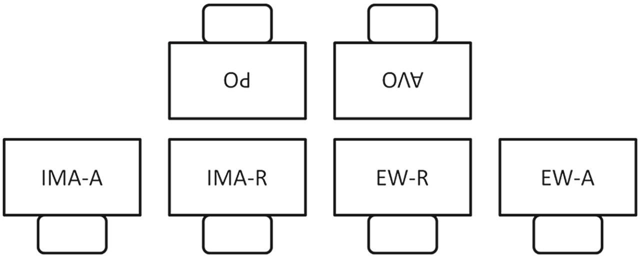


Fig. 4. Boardroom GCS layout: the crew is collocated with the AVO and PO facing the other crewmembers.

The goal of the proposed study is to evaluate the effectiveness of each of the three GCS layout options for supporting individual and collective duties of a UAS crew during a combat mission.

3.2 Participants

The participants consisted of one crew of six CAF personnel experienced in their individual crew roles in real CAF operations. Their experience includes operating in the Canadian Heron UAS Detachment and on CP-140 Maritime Patrol Aircraft. None of the crewmembers had previously worked together. Experience ranged from 3 months (IMA-A) to 4 years (AVO) in their roles on UAS or related platforms, for details see Table 1. Each of the participants had at least 10 years of service in the CAF.

Table 1. Crew experience.

Crew position	Platforms	Duration	Years of service
AVO	Multiple	4 years	23
PO	Multiple	7 months	28
IMA-A	CP-140	3 months	10
IMA-R	Heron, CP-140	16 months	19
EW-A	Heron	9 months	10
EW-R	Multiple	10 months	18

3.3 Crew Training

Crew training occurred in three phases. The first phase had the crew participate in a 2-day course covering Rules of Engagement (ROE), the Law of Armed Conflict (LOAC) and human performance in military aviation. The second phase consisted of a one-half day individual workstation training to familiarize each crewmember with the hardware, software, functions, capabilities and limitations of their workstation on TIGER. The third phase had the crew perform a training scenario twice in order to practice completing mission objectives as a team and practice skills learned in the previous two training phases. Completing the training scenario twice allowed the crew to identify and correct for individual and team performance issues that manifested during the first run. Crew training was performed in a GCS layout different than the three presented in Figs. 2, 3 and 4. The training GCS layout had the crew collocated, with each cell facing away from the other cells.

3.4 Simulated Mission Scenarios

The GCS layout evaluation consisted of the crew completing one evaluation mission scenario for each TIGER layout. The order of GCS layouts was chosen randomly. Each evaluation scenario lasted approximately two hours with additional time needed for the IMA-R and EW-R to generate reports after the mission. One evaluation scenario was performed per day to allow time for TIGER to be reconfigured into a new layout for the next evaluation. The evaluation scenarios were designed to be distinct by following different narratives yet equivalent in terms of the functions and tasks to be carried out by the crew. Having different evaluation missions ensured the crew would not perform the same scenario twice across experimental conditions. All of the scenarios are representative of JUSTAS missions.

The three simulated scenarios are named “Objective Phoebe”, “Objective Oceanus”, and “Objective Cronus.” All the scenarios take place in a coastal city. Objective Phoebe had the crew monitor a known insurgent making package drop-offs at various buildings then attend a meeting with other insurgents. The crew was then tasked with engaging an armed insurgent observed leaving the meeting. Objective Oceanus had the crew locate the position of a downed friendly rigid hull inflatable boat and report the position for a rescue. Afterwards, the crew observes insurgents kidnapping a friendly in the city. The crew informed their tasking authority of the kidnapping location to support a rescue.

The crew then engaged armed insurgents that escape the kidnapping rescue. Objective Cronus had the crew maintain over-watch for a friendly ground-based task force attempting to extract a high value target. During this task, the crew engaged in improvised explosive device and rocket propelled grenade threats to the task force.

During each of these missions, the crew needed to complete three main tasks: (i) pattern of life surveillance, (ii) target surveillance, and (iii) target engagement. The pattern of life task consisted of collecting intelligence over areas of interest. Target surveillance consisted of moving the UAV and sensor in an effort to keep vehicles or people of interest in the field of view of the UAV sensor in the urban environment. Engaging a target consisted of using ROE and LOAC in a decision to legally strike a target with UAV weapons.

3.5 Evaluation Metrics

Several objective and subjective observations and measures were gathered during training and the evaluation missions, including: training effectiveness, crew team task performance, Situation Awareness (SA), mental workload, and participant feedback. SA and workload questionnaires were completed both during and after evaluation missions. This paper will focus on reporting crew team performance and post-mission SA measures.

A UAS SME used a behavioral marker checklist with nine items to measure crew performance. The checklist includes markers for both individual and team performance. The SME rated the crew on a 5-point Likert scale for each marker (5 being "very good, 1 being "very poor"). Five team based behavioral markers, adapted from [14], were used to judge crew effectiveness at (i) monitoring, (ii) communication, (iii) conflict resolution, (iv) cross-checking, and (v) coordination and prioritization. A sixth team-based marker was used to judge IMA and EW teamwork and how well these crew members shared information between their cells.

The Situation Awareness Rating Technique (SART) provides a validated and practical subjective rating tool for the measurement of SA, based on personal construct dimensions associated with SA [15]. Instead of numeric Likert-scales, a graphical display of the rating scales was utilized, where the length of line from the left hand side of the scale to the participant's mark represents a respective rating score for one item. The possible range is between 0 ("low") and 10 ("high"). Each crewmember was asked to rate their situation awareness using SART immediately after each evaluation mission.

4 Results

4.1 Behavioral Marker Checklist Results

The individual and team behavioral marker results are shown in Table 2. For the distributed GCS layout, the crew was rated "good" (4) for all the markers except for "UAS positioning to optimize sensor performance" and "IMA and EW Teamwork" for which the crew was rated "fair" (3). No conflicts occurred during the distributed layout evaluation.

Table 2. Crew behavioral marker data for three GCS layouts.

	Behavioral marker	Distributed	Classroom	Boardroom
AVO	UAS positioning to optimize sensor performance	3	4	5
	UAS positioning to optimize weapon performance	4	4	5
PO	Payload management to provide useful imagery	4	4	5
	Payload management to provide useful targeting information	4	4	5
IMA	Use of image management tools to provide useful imagery	4	3	4
	Use of image management tools to provide useful targeting information	4	3	4
	Visual sensor imagery reporting	4	3	4
EW	Use of EW tools to provide useful contextual information	4	3	4
	ELINT reporting	4	4	4
Teamwork	Monitoring	4	3	5
	Communication	4	3	5
	Conflict resolution	N/A	N/A	4
	Cross-checking	4	4	5
	Coordination and prioritization	4	3	5
	IMA and EW teamwork	3	2	5

In the classroom GCS layout, the crew was rated “good” for the individual markers of the AVO and PO. The three IMA individual markers were rated “fair”. The EW marker “Use of EW tools to provide useful contextual information” was rated “fair” and the “ELINT reporting” was rated “good”. The team behavioral markers were rated as “fair” except for cross-checking, which was rated “good” and “IMA and EW Teamwork” which was rated “poor” (2). No conflicts occurred during the classroom layout evaluation.

In the boardroom GCS layout, the AVO and PO were rated “very good” (5) and the IMA and EW cells were rated “good” on the individual markers. The crew was rated “very good” on all the team markers except for conflict resolution, which was rated “good”.

Overall, the crew performed well for each of the three GCS layouts, with all except one behavioral marker rating being “fair” or higher. The boardroom layout resulted in the highest ratings. The boardroom layout was the only layout in which the crew obtained “very good” ratings, both on individual and team markers. It should also be noted that the crew obtained lower team marker ratings for the classroom than the distributed layout.

4.2 Situation Awareness Results

The SA measures data was aggregated and averaged by cell and GCS layout, Fig. 5. Overall, observed self-rated SA was higher on average, for all cells, in the boardroom condition when compared to the first two conditions.

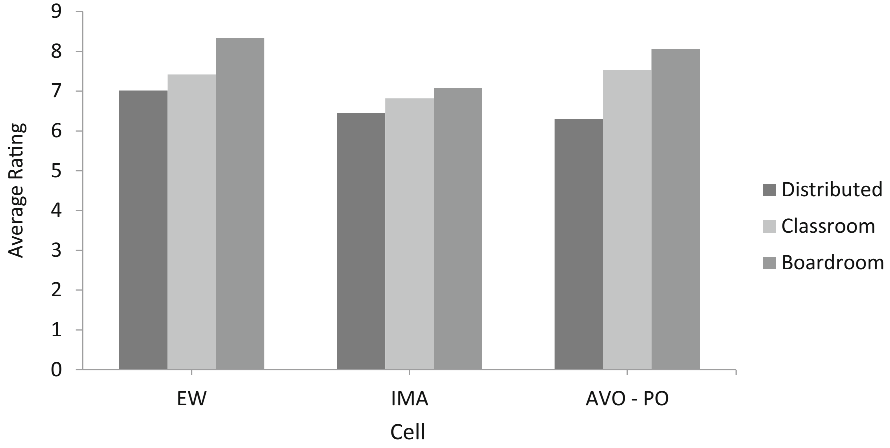


Fig. 5. SA ratings aggregated by cell and GCS layout.

5 Conclusions

DRDC has conducted an HFE trial to investigate how UAS crews perform at individual and team levels when conducting missions in different GCS layouts. The TIGER UAS GCS simulator was reconfigured into three layouts for evaluation: (i) distributed layout, (ii) classroom layout, and (iii) boardroom layout. The results of this preliminary study indicate that the boardroom layout was associated with improved teamwork behavior (monitoring, communication, coordination and prioritization, etc.) performance and self-rated SA than the distributed and classroom TIGER layouts. The main limitation of this study is its reliance on a single crew, mainly due to the availability of suitable CAF personnel with the requisite knowledge and experience in UAS operations. All efforts were taken to reduce the effects of this limitation, including the development of the three simulated mission scenarios that had the crew perform the same basic tasks in different orders and contexts, prohibiting the crew from anticipating their own tasks ahead of time. Additionally, the training with multiple missions emphasized crew familiarity TIGER in order to limit the possibility of a training effect during the evaluations. Future work will include evaluating the three GCS layouts with multiple crews and a variety of missions.

Acknowledgements. We would like to thank everyone who helped with the experiment, especially Capt. Sonny Rose, Leo Boutette, Adele Mininni, John Townsend, Juliana Fung, Jim Clark, Kevin O’Mahony, and Sylvain Pronovost.

References

1. Cook, K.: The silent force multiplier: the history and roles of UAVs in warfare. In: IEEE 2007 Aerospace Conference, pp. 1–7. IEEE Press, New York (2007)
2. Garrett-Rempel, D.: Will JUSTAS prevail? Procuring a UAS capability for Canada. R. Can. Air Force J. **4**(1), 19–31 (2015)
3. O'Leary, M.B., Mortensen, M.: Go (con)figure: subgroups, imbalance, and iso-lates in geographically dispersed teams. Organ. Sci. **21**(1), 115–131 (2010)
4. Singer, M., Grant, S.C., Commarford, P., Kring, J., Zavod, M.: Team performance in distributed virtual environments. Technical Report 1118, US Army Research Institute for the Behavioral and Social Sciences, p. 70 (2001)
5. Cordova, A., Keller, K.M., Menthe, L., Rhodes, C.: Virtual collaboration for a distributed enterprise. Technical Report. RAND Corporation (2013)
6. Driskell, J.E., Radtke, P.H., Salas, E.: Virtual teams: effects of technological mediation on team performance. Group Dyn. Theory Res. Pract. **7**(4), 297–323 (2003)
7. 1 Canadian Air Division: JUSTAS concept of operations (CONOPS). Canadian Armed Forces (2013)
8. Wang, W., Baker, K., Moreau, R.: Communication requirements and workspace layout effectiveness for the Joint Operations Centre (JOC) in CAGE IIIB. DRDC Scientific Report DRDC-RDDC-2015-R258 (2015)
9. Hou, M.: Testbed for integrated GCS experimentation and rehearsal (TIGER) development summary I. DRDC Technical Report, DRDC-RDDC-2015-L282 (2015)
10. McColl, D., Banbury, S., Hou, M.: Testbed for integrated ground control station experimentation and rehearsal: crew performance and authority pathway concept development. In: Lackey, S., Shumaker, S. (eds.) Virtual, Augmented and Mixed Reality. LNCS, vol. 9740, pp. 433–445. Springer, Heidelberg (2016)
11. Covas-Smith, M., Grant, S.C., Hou, M., Joralmon, D.Q., Banbury, S.: Development of a testbed for integrated ground control station experimentation and rehearsal (TIGER): training remotely piloted aircraft operations and data exploitation. In: Unmanned Systems, pp. 1–11. Association for Unmanned Vehicle Systems International (AUVSI) (2015)
12. McColl, D., et al.: Authority pathway: intelligent adaptive automation for a UAS ground control station. In: Human-Computer Interaction International Conference 2017, pp. 1–14 (In press)
13. Banbury, S., Baker, K., Proulx, R., and Tremblay, S.: TA1: unmanned air system operator information flow and cognitive task analyses. DRDC Contract Report DRDC-RDDC-2014-C307 (PA) (2014)
14. Baker, K., Banbury, S.: Force level tactical command and control in a littoral environment. Report I: Literature Review and Mission Analysis. DRDC Contract Report CR2008-999 (2008)
15. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: Situational Awareness in Aerospace Operations (AGARD-CP-478), pp. 3/1–3/17. NATO-AGARD, Neuilly Sur Seine (1990)

Investigation of Gesture Based UAV Control

Brian Sanders[1(✉)], Dennis Vincenzi[2], and Yuzhong Shen[3]

[1] Department of Engineering and Technology, Embry-Riddle Aeronautical University,
Worldwide, Daytona Beach, FL, USA
sanderb7@erau.edu

[2] Department of Aeronautics, Undergraduate Studies, Embry-Riddle Aeronautical University,
Worldwide, Daytona Beach, FL, USA
vincenzd@erau.edu

[3] Department of Modeling, Simulation, and Visualization Engineering,
Old Dominion University, Norfolk, VA, USA
YShen@odu.edu

Abstract. The purpose of this study is to investigate the functionality and effectiveness of hand gestures to control flight operations and onboard equipment of unmanned aerial vehicles (UAVs). This study utilizes the Leap Motion Controller (LMC) to sense hand motions. The LMC is a gesture based, human-computer interface (HCI) device. For this application, a primitive gesture library was developed by considering fundamental UAV operations. The LMC application programming interface was then used to implement an interpreter to identify the motions and translate them into UAV performance and control commands. Careful attention was paid to identifying and developing natural, intuitive, and common hand gestures that can be sensed in the LMC field of view. It is anticipated that this work will provide the framework from which to develop a gesture based library for actual UAV control and other LMC applications.

Keywords: Hand gestures · Leap Motion Controller · Hovercraft · UAV

1 Introduction

Current unmanned systems industry, specifically the unmanned aerial system (UAS) market, has been experiencing significant growth in recent years due to maturation and advances in related technology, increased application opportunities, availability of key components and materials, and solidification of Federal Aviation Administration (FAA) rules and regulations governing legal use and requirements for use of unmanned vehicles within the National Airspace System (NAS) [1, 2, 3]. A significant amount of literature has been written about the need to design better displays for UAS but little effort has been expended to develop controls that are innovative, take advantage of new technology, and are natural and intuitive in design. Instead, sophisticated UAS have relied on legacy displays and controls, and have paid little attention to new technologies that have recently become available for use. Commercially available small unmanned aerial systems (sUAS) have traditionally been designed and controlled using legacy hand-held

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_20

controllers that provide the standard miniature joystick control interface to control the basic maneuvering functions of the unit (Fig. 1).

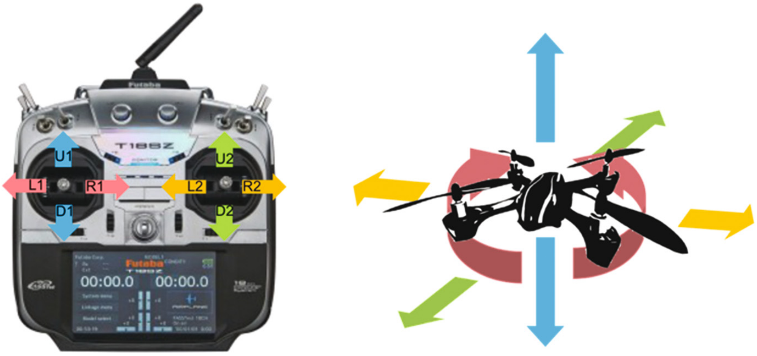


Fig. 1. Typical legacy controller and movement designations for sUAS control.

Traditional control interfaces are typically 1 dimensional (1D) or 2 dimensional (2D) devices that allow the user to interact with a system in a limited manner [4]. Keyboards are 1D input devices that allow for text input and activation of preprogrammed functions via a sequence of key/text inputs. Mice have expanded input capabilities into a 2D framework but input is still limited to menu item selection or “hotspots” on a graphical user interface (GUI). Both of these control devices, while functional and useful, are limited in nature and not very intuitive in terms of control movement, input, and function, and they are often slow and time consuming as control through these devices often requires a series or sequence of inputs to achieve the desired end state.

Legacy control devices such as the one pictured above (Fig. 1) are better, but still an attempt to translate 2D input into movement through a 3 dimensional (3D) space or environment. Integration with touch-sensitive devices such as phones and tablets are beginning to emerge on the market to replace or augment discrete physical controls and information displays [5]. However, these devices are, in many cases, simply electronic or digital versions of the same 2D legacy control devices. These devices typically combine electronic visual displays with touch input, and sometimes electronic input (GPS, accelerometers, and automation for example).

The gaming industry recognized the potential to create more robust visualizations and has concentrated on developing environments that move away from the standard 2D environments into richer and more robust 3D environments using 3D stereoscopic displays [4]. Virtual Reality (VR) displays or VR-like displays are now affordable and commonplace, and are regularly used as the display of choice when immersion into a 3D environment is preferred. High resolution displays on phones or inexpensive Helmet Mounted Displays (HMDs) such as the Oculus Rift have begun to replace and augment visual systems to develop environments that serve as displays for vehicle parameters as well as provide an egocentric view from the UAS camera. The capability exists for technology to provide more information with realistic visual perspectives similar to looking through a Heads Up Display (HUD) on a manned aircraft. Utilization of these types of technologies, if designed correctly, can result in a more realistic visual display that provides the information needed for successful operation with minimal training requirements.

1.1 Combining New Technology for New Displays and Controls

The objective of developing new technology and new approaches to Human Machine Interfaces (HMIs) is to increase system efficiency and reduce cognitive workload on the individual operator. The term "unmanned system" is a misnomer; since there is a human operator present in the system, the system will always be "manned" in some way. The only difference in the case of unmanned systems is that the operator is not collocated with the vehicle.

Being separated from the vehicle places the human in a unique position and provides a different operational perspective in that many of the environmental cues normally present in manned flight scenarios are no longer present and available to the human operator. For Visual Line of Sight operations (VLOS), the operator must rely on visual observation; for Beyond Visual Line of Sight (BVLOS) operations, the user must rely on instruments and displays. Developing displays using new technology described above to enhance the visual display for the human operator makes sense from a human performance perspective.

Improvements in communication capabilities coupled with development of new virtual-friendly environments have created the potential for a HMI that both takes advantage of technological innovations and encourages the development of new types of interfaces [5]. In order to be successful and readily accepted by UAS users in general, the new technology and HMIs must be easier to use and more intuitive in nature. They must produce a combined system performance that is better than existing HMIs in that workload is reduced, situation awareness is increased, and system safety and reliability is enhanced.

1.2 Gesture Based Approach to Command and Control

Drones are widely used around the world for a variety of purposes including aerial videography, photography, and surveillance [6]. Successful accomplishment of these tasks requires the execution of a series of basic maneuvering functions of a sUAS that, when combined, result in movement (take off, acceleration, change in direction, change in altitude, transit from one point to another, hover, and landing for example).

Gestures and visual signals are common in military and aviation domains. A series of standard gestures (hand and arm signals) has been in used for many years to transmit information from one person to another. Gestures are movements of human body parts - usually the arms, head, and hands - that provide contextual meaning [7]. Development of a gesture based approach for sUAS operation may be a viable alternative for implementation into command and control interfaces using technology that is designed to recognize gestures.

A gesture based approach can free the operator from having to hold and operate a multi-joystick, multi-button based controller by correlating the UAS operations to a set of fluid, intuitive, natural, and accepted set of hand gestures. This in combination with emerging display configurations could be used to create systems in which the vehicle operator can observe the vehicle position, monitor its operations, stay abreast of vehicle performance parameters, and have the ability to control the sUAS. All of this could be accomplished without the cumbersome necessity of holding a cryptic control device. A simple gesture control interface can make the task of piloting much easier and more intuitive in nature [6].

To date, studies examining the functionality and effectiveness of virtual interfaces using gestures as the primary method of interaction with a system have been scant. Design of new interfaces that take advantage of emerging technologies are required to complement and enhance the capability of the human component and the system in terms of performance. Some gesture related research centers around development of computer algorithms that would allow robotic systems to recognize gesture commands in the field as part of military teams. Other research focuses on virtual reality environments integrated with optical sensors to recognize and measure movement, velocity, and patterns of movement, of fingers and hands, and then translates those gestures into commands. These commands are typically interpreted with the intent of changing the state of something in the virtual environment (turning something on or off, for example).

Limitations must also be considered regardless of whether the receiving unit is a human or computer. The U.S. Military has developed a standard set of visual signals for use in combat operations. Effective operations depend on clear and accurate communication among ground units and supporting aviation units [8, 9]. Limitations of visual signals include range and reliability which are dependent upon visibility, and this may affect the degree to which these visual signals are understood or misunderstood. The same is true in computer environments. The degree of recognition is dependent upon many factors including noise present in the environment, accuracy and consistency of the gesture, and resolution of the receptor.

Hamilton et al. [7] conducted research that focused on developing the ability for robotic systems to understand military squad commands. The long term goal was to develop the capability to integrate robots with ground forces as seamless teammates in combat operations. Lampton et al. [10] conducted research using a gesture recognition system integrated with a virtual environment in an attempt to measure the accuracy and effectiveness of a VR based gesture recognition system. The researchers selected 14 basic and accepted hand gestures commonly used in the field by U.S. Army personnel. In general, the results were mixed in terms of recognition and accuracy. Many of the gestures were problematic in terms of tracking, recognition, or both [10].

Gesture based technology is integrating into our everyday life via applications such as cell phones and touch screen computers and affordable technology. The Leap Motion Controller (LMC) is a relatively recent technology that can capture and track hand motion with a sensor just slightly bigger than a USB flash drive. Studies have been reported ([10, 11, 12, 13]) that describe the accuracy of the LMC less than a mm for static conditions and on the order of a mm for dynamic conditions. Scicali and Bischof [14] developed several games to gauge user performance in different 3-D environments. They obtained excellent general information about several usable gestures. McCartney et al. [4] used data collected data from over 100 participants to train a 3D recognition model based. They reported an accuracy rate of 92.4% with the goal of trying to gain support for the creation of a gesture-based language. Some innovative applications have recently recently been reported. Staretu and Moldovan [16] used the LMC to control a robotic gripper and Sarkar et al. [6] used it to control some basic motions of a UAV. The application investigated in this study is the gesture control of UAVs. The study will develop the beginnings of a gesture library by conducting a task decomposition for control of a representative, recreational UAV and matching the task to the capability of the LMC. The objective is to identify and design gestures that are natural and intuitive for incorporation into a gesture based HMI for UAV control.

2 Methodology

There are two components of this investigation. The first focuses on defining a gesture library that matches UAV flight control requirements, such as climb and descend, to common and natural gestures to meet those needs. The general philosophy is to blend natural gestures while optimizing the LMC capabilities as illustrated in Fig. 2. For this investigation a typical recreational hovercraft was selected and a task breakdown analysis was conducted to identify tasks associated with vehicle control and onboard equipment control. This vehicle type was selected due to its simplicity but yet multifunction capability such as camera and automated navigation control functions. Identified control tasks were then matched to potential gestures that are incorporated within the LMC's capability.

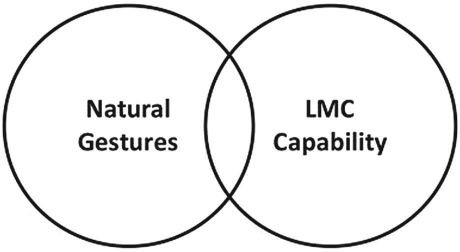


Fig. 2. Metrics for gesture selection

The second component is to capture gesture data from the LMC to assess the human performance vs the technology capability. It is of interest to assess the reliability and fidelity of the LMC for the current application. This is achieved by capturing and analyzing data related to hand motion and gestures. There are 4 gestures built into the LMC Application Programming Interface (API) and are shown in Fig. 3. They are circle, swipe, key tap, and screen tap. The circle gesture is recognized when a finger moves in a circular motion. The swipe gesture is a linear movement of a finger and is functional in any direction. The Key Tap gesture is a quick, downward tapping movement by a finger. The screen tap gesture Screen Tap is a quick, forward tapping movement by a finger. Which finger is assoicated with these gestures is definable. Table 1 shows relevant attributes associated with the gesture as well as parameters for defining when a gestures is recognized. More details on these can be found at the LMC website [17].

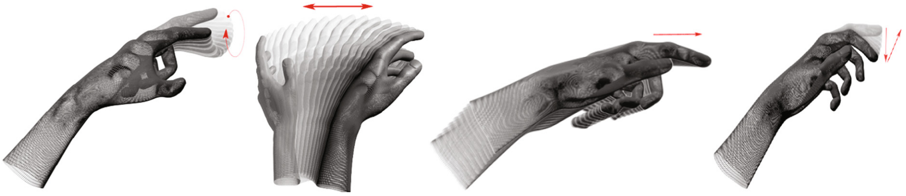


Fig. 3. (a) Circle gesture, (b) swipe gesture, (c) key tap, (d) screen tap [17]

Table 1. LMC built-in gestures and associated attributes

Gesture	Attributes
Circle	• Radius of the circle (mm), Progress from start (radians) • Gesture Recognition Parameters: minimum arc length and radius
Swipe	• 3D direction vector, Velocity (mm/s), Starting point of gesture • Gesture Recognition Parameters: min velocity and length
Key Tap	• 3D direction vector • Gesture Recognition Parameters: min downward velocity, history, min distance required for the movement
Screen Tap	• 3D direction vector, Position where the screen tap is registered • Gesture Recognition Parameters: min forward velocity, history, min distance required for the movement

Attributes associated with the hand motion can also be assigned to vehicle control tasks. For example, Fig. 4 shows the normal and direction vectors associated with a hand. Using these vectors, it is possible to track hand rotation (pitch, roll yaw). It is also possible to track the location of the hand in three dimensional space. One final capability of interest is the ability to select either right or left hand for which to designate control functions. In the next section these capabilities along with the built-in gestures are matched to potential vehicle control tasks.

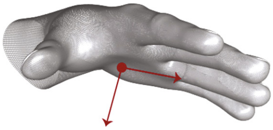


Fig. 4. Palm shown normal and direction vectors [17]

3 Results and Discussion

3.1 UAV Control Task Breakdown and Gesture Matching Analysis

The first step in the task breakdown and gesture matching analysis is to identify the functions associated with flying and operating a representative recreational hovercraft. These are then matched to LMC gesture capability described above. This task breakdown is shown in the first column of Table 2. It is partitioned into three categories: flight control, component control, and camera control. The description of the flight control is related to what the operator wants to happen rather than how the vehicle does it. For example, the desired action is for the vehicle to climb or turn. This motion can then be translated to the pitch, roll and yaw of the vehicle. Those performance attributes can be determined by the internal control logic of the air vehicle and are transparent to the operator.

Table 2. Hovercraft control actions and gestures

Vehicle action	Gesture	Built in LMC gesture
Flight control		
Climb/descend	Hand pitch up/down	
Move left-right	Hand direction left or right	
Move backward-forward	Palm motion back-forth	
Move diagonally	Palm motion back-forth	
Speed	Swipe motion	✓
Stop	Hands leave FOV	
Control initiation	Key tap	✓
Component control		
Power on/off	Screen or key tap	✓
Various modes (i.e. automated positioning)	Screen or key tap	✓
Camera control		
Rotate (horizontal and vertical)	Circle gestures	
Zoom-in, zoom-out	Scaling motion	✓

These desired actions were then matched potential gesture. Gesture selection was based on consideration of natural association and common association. For example, pitching the hand up and down or left and right is a natural hand motion to control those vehicle flight maneuvers. On the other hand, common gestures are defined as those accomplished on most touch screen interfaces such as increasing and decreasing the distance between the thumb and forefinger to represent zoom-in and zoom out actions. Most of the descriptions should be self explanatory with the exception of the control initiation. This is an action that tells the controller that the next series of gestures is intended for vehicle control and anything before that should be ignored.

Figure 5 shows the leap motion coordinate system while Figs. 6, 7 and 8 show the data captured by the LMC. Figure 6 shows the path created by the index finger conducting a circle gesture. Figure 6a is when the figure is pointing in the horizontal direction while Fig. 6b the index finger is pointing vertically. The latter may be a more natural gesture for rotating a camera. Figure 7 shows path (7a) and velocity (7b) data for a swipe motion. The displacement data is smooth and consistent while the velocity data is repeatable but has a noticeable variation. This is mostly likely a function of the human performance limitations to accurately control speed rather than the LMC. Figure 8 shows the x vs z displacement when trying to move the hand forward and backward and then left to right as well as rolling, pitching, and yawing motion. As reported earlier by Weichert et al. [13] the LMC itself is highly accurate (i.e., sub-millimeter accuracy), but additional filtering of the data may be required to account for variations in human performance when it comes to control of the vehicle. Otherwise it may result in undesirable performance effects.

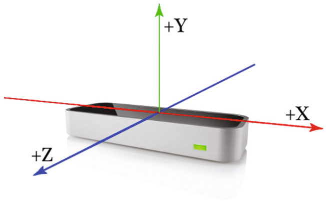
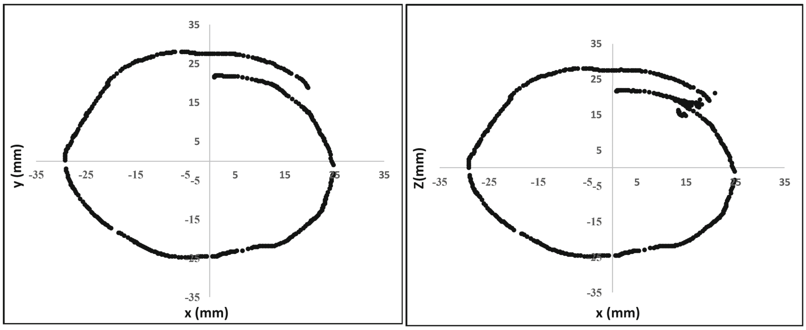


Fig. 5. Leap motion coordinate system [17]



(a) (b)

Fig. 6. Circle gesture data (a) index finger horizontal, (b) index finger vertical

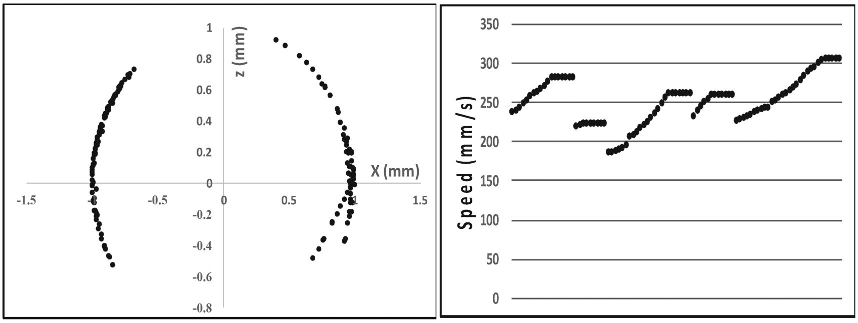
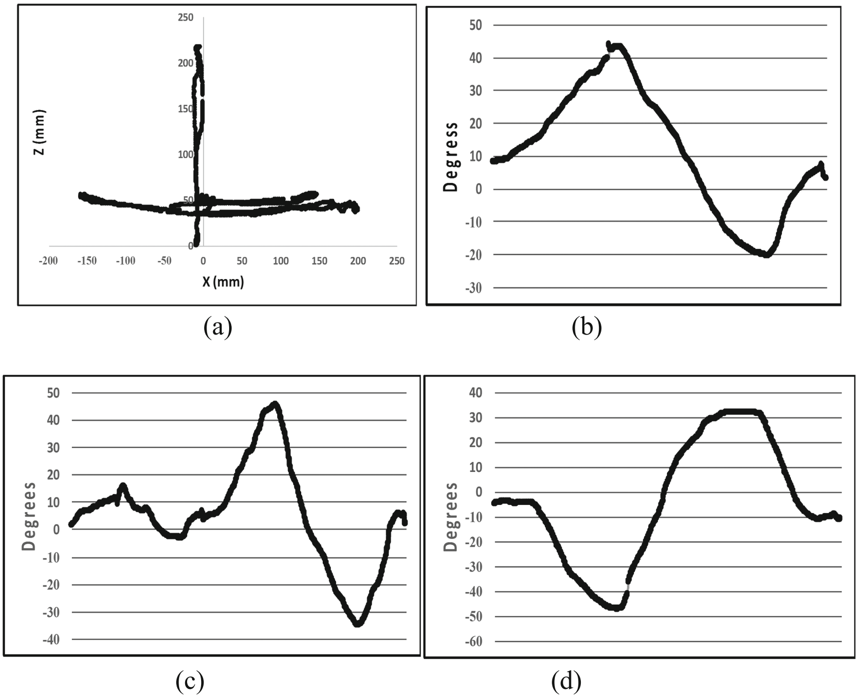


Fig. 7. Swipe data – acquired by swiping index finger over the LMC.



(a) (b)

(c) (d)

Fig. 8. Hand motion data– (a) x vs z movement, (b) pitch, (c) roll and (d) yaw.

4 Summary and Way Forward

This study investigated the suitability of hand gestures to control recreational hovercraft style drones. This was accomplished by conducting a decomposition of typical tasks and identifying natural and common gestures to control the flight operations and onboard equipment such as cameras. Gestures were matched to the capability of an optical sensor (a Leap Motion Controller) that can capture and interpret them for translation to UAV control instructions. Several suitable gestures were identified that are within the capability of the LMC controller and meet the desired vehicle control requirements. The LMC was highly accurate but some additional filtering may be required to smooth out variability in human performance. Future work will include conducting flight simulations to evaluate human performance metrics in more detail.

References

1. Terwilliger, Brent A., Ison, David C., Vincenzi, Dennis A., Liu, Dahai: Advancement and application of unmanned aerial system human-machine-interface (hmi) technology. In: Yamamoto, S. (ed.) HCI 2014. LNCS, vol. 8522, pp. 273–283. Springer, Cham (2014). doi:10.1007/978-3-319-07863-2_27
2. U.S. Department of Transportation, John A. Volpe National Transportation Systems Center: Unmanned aircraft system (UAS) service demand 2015-20135: literature review and projections of future usage. (Report no. DOT-VNTSC-DoD-13-01). (2015). http://ntl.bts.gov/lib/48000/48200/48226/UAS_Service__Demand.pdf (2013). Accessed 2015
3. U.S. Department of Defense. Unmanned systems integrated roadmap FY2013-2038. (Report no. 14-S-0553). http://www.defense.gov/pubs/DOD-USRM-2013.pdf (2013)
4. McCartney, R., Yuan, J., Bischof, H.: Gesture recognition with the leap motion controller. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV). 2015, vol 3 (2015)
5. Balog, C., Terwilliger, B., Vincenzi, D., Ison, D.: Examining human factors challenges of sustainable small unmanned aircraft system (sUAS) operations. In: Savage-Knepshield, P., Chen, J. (eds.) Advances in Human Factors in Robots and Unmanned Systems of the Series Advances in Intelligent Systems and Computing, vol. 499, pp. 61–73. Springer, New York (2016)
6. Sarkar, A., Patel, K.A., Ram, R.K.G., Capoor, G.K.: Gesture control of drone using a motion controller. Paper presented at the 1–5. (2016). doi:10.1109/ICCSII.2016.7462401
7. Hamilton, M.S., Mead, P., Kozub, M., Field A.: Gesture recognition model for robotic systems of military squad commands. Paper no. 16089. Presented at the 2016 Interservice/Industry, Training, Simulation and Education Conference, Orlando, FL (2016)
8. ARI Research Note 2003–2006: Gesture recognition system for hand and arm signals. In: Lampton, D.R., Knerr, B., Clark, B.R., Marting, G.A., Washburn, D.A., Rosas-Anderson, C.J. (eds.) United States Army Research Institute for the Behavioral and Social Sciences (2003)
9. Army Field Manual FM 21-60, September 1987
10. Lampton, D.R., Knerr, B.W., Clark, B.R., Martin, G.A., Washburn, D.A., Army Research Inst for the Behavioral and Social Sciences Alexandria VA.: Gesture Recognition System for Hand and Arm Signals (2002)
11. Bachmann, D., Weichert, F., Rinkenauer, G.: Evaluation of the leap motion controller as a new contact-free pointing device. Sensors **15**(1), 214–233 (2015). doi:10.3390/s150100214
12. Guna, J., Jakus, G., Pogacnik, M., Tomazic, S., Sodnik, J.: An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors **14**(2), 3702–3720 (2014). doi:10.3390/s140203702
13. Weichert, F., Bachmann, D., Rudak, B., Fisseler, D.: Analysis of the accuracy and robustness of the leap motion controller. Sensors **13**(5), 6380–6393 (2013). doi:10.3390/s130506380
14. Scicali, A., Bischof, H.: Usability study of leap motion controller. In: Proceedings of the International Conference on Modeling, Simulation and Visualization Methods (MSV), vol. 39 (2015)

15. Lu, W., Tong, Z., Chu, J.: Dynamic hand gesture recognition with leap motion controller. IEEE Signal Process. Lett. **23**(9), 1188–1192 (2016). doi:10.1109/LSP.2016.2590470
16. Staretu, I., Moldovan, C.: Leap motion device used to control a real anthropomorphic gripper. Int. J. Adv. Rob. Syst. **13**(3), 113 (2016). doi:10.5772/63973
17. Leap Motion Developer retrieved March 2, 2017 from. https://developer.leapmotion.com/. Accessed 2 March 2017

Tactile Applications for Enhanced Performance in Naturalistic Settings

How Close Is Close Enough? A Multimodal Analysis of Temporal Matching Between Visual and Tactile Signaling

Catherine Neubauer(✉)

US Army Research Laboratory, Adelphi, MD, USA
catherine.neubauer@gmail.com

Abstract. Research has shown beneficial performance gains from concurrent multimodal presentation of visual and tactile signaling. Studies have also suggested the importance of closely matching or emulating the spatial characteristics of tactile signaling to its visual counterpart, resulting in intuitive tactile signals that are easily learned and that provide immediate benefits in the absence or concurrent presentation of visual signaling. The purpose for this study was to inform display design regarding how closely the tactile signaling should match the visual signaling temporally, before the observer detects the difference. Participants observed a visual signal presentation of six different circular patterns, that spatially matched a concurrent tactile presentation, with the visual presentation being temporally faster, slower, or the same speed as the tactile presentation. Results showed that participants were better at identifying a difference between the visual and tactile stimuli when the visual stimuli were faster. The incremental nature of the faster and slower visual presentations results in helpful guidelines for multimodal display design on how perceptible the temporal difference is between tactile and visual modalities.

Keywords: Multimodal presentation · Tactile signaling · Wearable technology

1 Introduction

Communication is an integral part of life. We communicate with others through many channels with the use of our eyes, voice, body language, and even touch. These systems are seldom used in isolation. Humans often prefer multi-modal forms of communication, which ultimately allow us to communicate more effectively. For example, the colorful story teller uses a combination of verbal and nonverbal cues to entertain the audience while often making swooping gestures with the arms and hands, not needed for meaning, but adding something to the story that would sorely be missed if the speaker stood absent from life. The slightest mismatch of multi-modal presentation can cause confusion and misunderstanding, even resulting in perceptual errors [1]. While visual and auditory combinations are the most prevalent form of multi-modal presentation, recent advances in technology are allowing the more frequent use of tactile communication.

The potential benefits for tactile displays have been widely noted. Both accuracy and response times can be improved with the use of this technology. For instance, an

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_21

experiment conducted by [2], found that tactile displays substantially improve accuracy, on target acquisition. This task was simultaneously performed alongside visual stimuli, relative to a condition in which a singular mode of visual or tactile presentation was used. As helpful as this tactile technology has been, there are still many questions and areas that need to be explored within the realm of tactile communication and signaling. These questions include, but are not limited to, discrepancies between modalities, timing discrepancies, and their detection. Research has shown that when the visual and tactile displays were displayed in a similar directional pattern, the perception of their speed was faster, compared to when they were displayed in opposite directions, which resulted in reports of a slower perceptual speed of stimuli [3].

Similarly, [4] created a tactile display that was used to create the illusion of movement, which was presented on the arm. They found that response times were much faster when the visual and tactile stimuli were matched in the same direction. Additionally, these authors also found that reaction times were slower when the visual and tactile stimuli were presented in differing directions. Findings such as these suggest that it is essential that the relationship between vision and touch is best received when the respective stimuli are presented in congruent modes. Additionally, tactile stimulation can result in a definite preference amongst its users. Here, this preference results in learning where participants are capable of identifying tactile patterns as the duration of the tactile stimulus moves from 80 to 320 ms [5]. In an effort to improve marksmanship, [2] found that tactile feedback can greatly improve target acquisition, alerting the individual when spatial deviations occurred. This research found that tactile feedback greatly improved accuracy and decreased response time amongst individuals. Additionally, understanding the physiological constraints of a user is vital when developing this tactile display. More specifically, identifying the ideal settings for participants to receive the haptic display is essential in order to produce the best performance outcomes, even under high physiological stress [6].

The implications that tactile displays present, when individuals are engaged in high workload situations, is great. For example, [7] found that tactile displays effect performance when individuals engage in complex activities. In complex environments, the relationship between cross-modal spatial cues and visual stimuli can greatly affect target response, with tactile cues shortening response time. For a more complete review of these and other tactile studies see also [8, 9]. The purpose of the current study was to identify when individuals can detect incongruence in timing between two presented modalities (e.g. visual and tactile). It was expected that participants would be able to detect incongruence between the visual and tactile displays within 60 ms of the constant speed of the tactile belt, which is set at an inter-stimulus interval of 140 ms.

2 Experimental Method

2.1 Experimental Participants

A total of forty participants (25 females and 15 males) ranging in age from 18 to 25, volunteered to participate in this study. Participants volunteered in this study through the University of Central Florida's (UCF) online extra credit system, SONA-SYSTEMS, which was offered in their undergraduate psychology class.

2.2 Experimental Materials and Apparatus

The tactile system used in the present study was developed using a plunger-type vibrotactile actuator (hereafter referred to as a "tactor"). The model C2 tactors, manufactured by Engineering Acoustics, Inc. (see Fig. 1) are essentially acoustic transducers that displace 200–300 Hz sinusoidal vibrations onto the skin. This frequency range, in combination with their 17 g mass, is sufficient to activate the skin's Pacinian corpuscles. The C2's contactor is 7 mm, with a 1 mm gap separating it from the tactor aluminum housing. The C2 is a tuned device, meaning it operates well only within a very restricted frequency range, in this case the optimal frequency of 250 Hz.

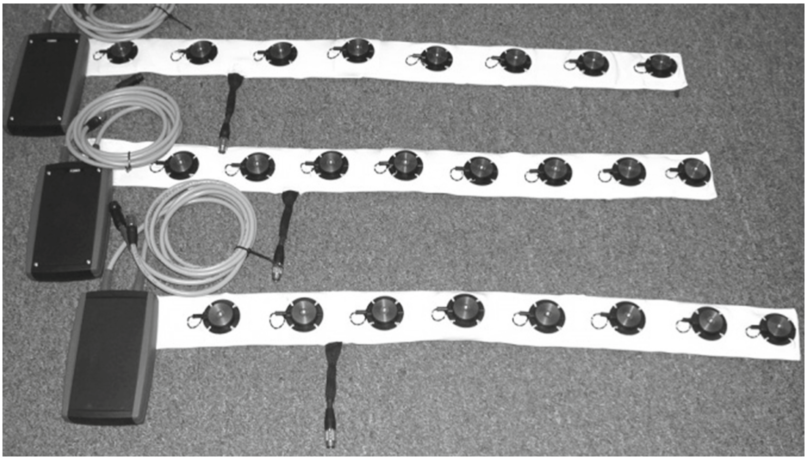


Fig. 1. Three tactile displays belt assemblies are shown above along with their controller box.

A tactile presentation, which emulated the visual presentations, was created, and corresponded to six circular patterns that were presented to participants. The six visual stimuli patterns were, "Full Circle Right/Clockwise, Full Circle Left/Counter Clockwise, Half Circle to the Front Right/Clockwise, Half Circle to the Front Left/Counter Clockwise, Half Circle to the Back Right/Clockwise, and Half Circle to the Back Left/Counter Clockwise". Very short video clips were presented to the participants, which ranged from a total speed of 200 to 1920 ms. Each video clip was presented in one of seven different speeds, or onset intervals, which corresponded to one of seven different onset and offset timings, (40, 80, 100, 140, 180, 200, and 240 ms). The on and offset timings of the small red dots created the illusion (Phi Phenomena) of circular movement. The 40 ms onset interval created a 320 ms total video for the directions Full Circles Right and Left, etc. The tactor onset timing was 140 ms for each tactor or 1120 ms full circle (the tau phenomena created the illusion of movement around the abdomen).

A dell multisync computer with a 19″ liquid crystal display was used to present the LabVIEW 8.5 National Instruments application, which presented a Windows media player viewing of the visual stimuli, executing the tactile display via a Bluetooth connection, in concert with the visual stimuli. Participants were also asked to wear sound attenuating headphones, which blocked any possible sounds effects the tactile display produced.

2.3 Experimental Design and Procedure

Participants viewed 42 different videos of a red dot traveling in a prescribed circular pattern, which was analogous to the tactile display layout. The red dot was present in one of eight locations sequentially, which made up the circular path and was present at that location for a certain duration and then present at the next prescribed location for that same duration until the pattern was established (see Fig. 2). The resulting onset and offset resulted in the appearance of circular movement (Phi Phenomena). Participants viewed the red dot traveling at seven different speeds, which ranged from 40 to 240 ms (ms) of stimulus onset interval. Additionally, the videos depicted the six aforementioned different circular patterns: (Full Circle Clockwise and Counter Clockwise, Half Circle Forward Clockwise & Counter). The middle speed, 140 ms, exactly matched the tactile display presentation, which was set at a constant speed of 140 ms. The 42 different combinations (randomized) of visual and tactile presentation were presented to the participants. The different combinations of the 42 different videos were representations

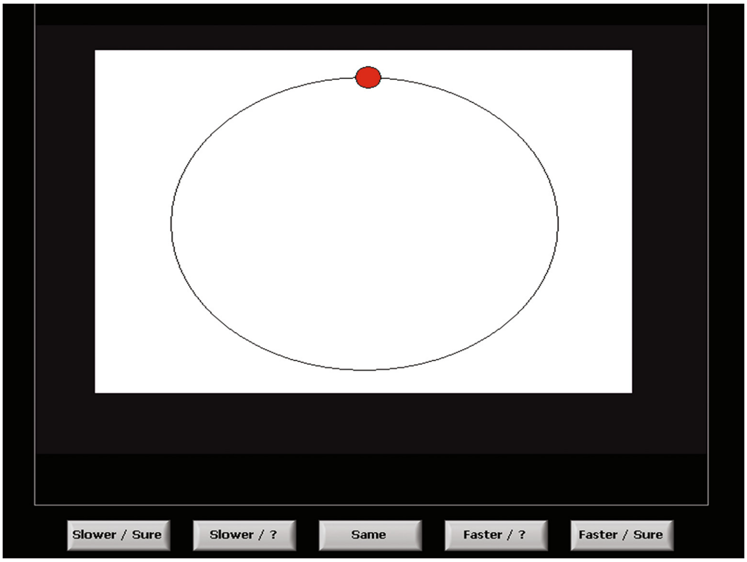


Fig. 2. A computer screen shot showing what the participant viewed as the red dot traversed the ellipse in one of the six aforementioned patterns. The participant clicked on the appropriate selection based on their perception of the visual stimuli compared to the simultaneously received tactile stimuli. The six choices reading from left to right are: Slower/Sure, Slower/?, Same, Faster/?, Faster/Sure.

of faster, slower, or perfectly matched visual stimuli with the tactile display. The participants then performed a forced choice response as to whether the visual presentation was slower, faster, or the same speed as the tactile display. The forced choices also allowed participants to make a confidence rating of their selection.

Each participant performed 168 trials, which represented 4 trials of each of the 42 different combinations. The order of the trials was randomized for each participant by a randomization algorithm inside the LABVIEW application. The entire experiment took less than 45 min.

3 Results

An analysis was conducted using SPSS 11.5, with an alpha level set at $p \leq .05$. The results were analyzed with special emphasis on accuracy and response time with respect to the directional stimuli presented. During the trials, participants were given the option of selecting five options: slower/sure, slower/unsure, same, faster/unsure, and faster/sure. Participants were encouraged to use the far ends of the five point scale to show an indication of their confidence in their selection.

The accuracy rates for the seven different speeds for the directional patterns Full Circle Right and Left, resulted in the greatest accuracy for the speed, 240 ms to the right with an accuracy of 96% as compared to the other six speeds (40, 80, 100, 140, 180, 200 ms) which resulted in the following accuracy rates, 94, 92, 78, 84, 44, and 94 percents respectively. Additionally, the greatest accuracy for the same pattern in the opposite direction, to the left, resulted in the greatest rate of accuracy for the speed, 240 ms, with an accuracy rate of 92%, as compared to 90, 86, 68, 57, 10, and 84 percents, for the respective speeds 40, 80, 100, 140, 180, and 200 ms. There was no statistical difference in the presentation of the visual signal or tactile signal clockwise or counterclockwise (Fig. 3).

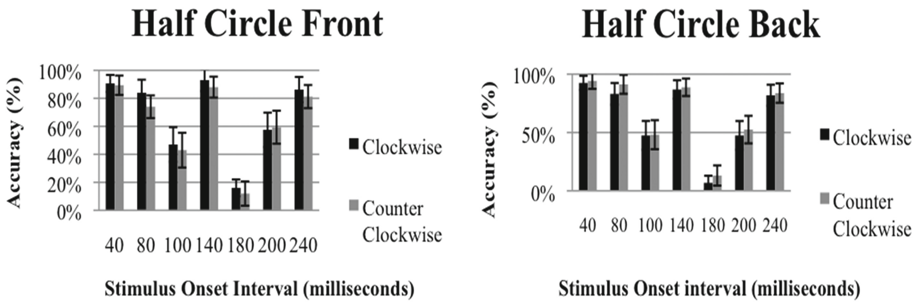


Fig. 3. Accuracy by visual onset interval. The tactile display onset interval was 140 ms meaning the 140 stimulus onset interval was the same as the tactile stimulation.

The accuracy rates for the seven different speeds for the directional patterns Half Circle Forward Right and Left resulted in greatest accuracy for the speed, 140 ms with an accuracy of 93% as compared to the other six speeds (40, 80, 100, 180, 200, 240 ms) which resulted in the following accuracy rates, 91, 84, 47, 16, 58, and 86 percents respectively. Additionally, the greatest accuracy for the same pattern in the opposite direction, to the left, resulted in the greatest rate of accuracy rate for the speed, 140 ms, with an accuracy rate of 89%, as compared to 74, 43, 88, 12, 59, and 81 percents, for the respective speeds 80, 100, 140, 180, 200, and 240 ms, again no difference in accuracy between direction.

Additionally, the fastest response time for the same pattern in the opposite direction, to the right, resulted in the fastest response time also for the speed 40 ms, with a response time of 3306 ms, as compared to 4370, 4847, 3656, 4601, 5210, and 4529 ms for the respective speeds 80, 100, 140, 180, 200, and 240 ms. Participants showed lower accuracy rates as the visual stimuli became slower than the tactile stimuli. Participants were able to better distinguish between when a visual signal was 40 ms faster, compared to when the visual signal was 40 ms slower.

4 Discussion

Along with the usability of tactile displays, the future implications are also an exciting area to explore, which have shown that tactile displays do help performance. Although the tactile display system is a very helpful and convenient way to increase performance, there is also another small detail that must be kept in mind: the tactile display must match the visual display as closely as possible.

As we have seen with the patterns Half Circle Front and Back, participants showed a marked decrement in performance when incongruence between the visual and tactile display was presented in an excess of 40 ms from the baseline temporal display of 140 ms, 100 and 180 ms respectively. Consequently, participants did best when the visual and temporal displays were congruent, but also at the tail ends of the visual display (e.g. 40 and 240 ms). Incongruence was rapidly detected usually within 60 ms of consequent incongruence; however, participants were better at detecting incongruence when the visual display was faster than the tactile display.

Additionally, response times were slowest at 240 ms, which would support the notion that participants were not as accurate in detecting incongruence when the visual display was presented slower than the tactile display. Display design should be aware that while it is important to maintain congruent timing, stimuli may be 40 ms off and individuals would rarely notice incongruent effects. This research shows that visual stimuli matter much more than temporal and/or any other kind of stimuli that is presented to the participant. It is important to remember which sense is affected the most; which ensures that performance is affected in the way design would want.

Because tactile communication can affect performance, it is crucial that more research be conducted to understand exactly what condition can and does affect performance. Additionally, it is important that future research employ participants in higher workload situations to discover maximum performance capabilities, as the workload and performance level of this experiment was minimal. The benefits of tactile communication are vast. The discrepancies that can occur between different sensory modalities can be diminished with the use of the display system. There are of course limits to the reality with which we can feel movement on the skin, but the benefits in terms of performance is certainly encouraging in the realm of communication.

References

1. McGurk, H., MacDonald, J.: Hearing lips and seeing voices. Nature **264**, 746–748 (1976)
2. Oron-Gildad, T., Downs, J.L., Gilson, R.D., Hancock, P.A.: Vibrotactile guidance cues for target acquisition. IEEE Trans. Syst. Man Cybern. **37**, 1–10 (2007)
3. Bensmaia, S.J., Killebrew, J.H., Craig, J.C.: Influence of visual motion on tactile motion perception. J. Neurophysiol. **96**, 1625–1637 (2006)
4. Gray, R., Tan, H.Z., Young, J.J.: Do multimodal signals need to come from the same place? Crossmodal attentional links between proximal and distal surfaces. In: Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces, pp. 437–441. IEEE Computer Society, Los Alamitos (2002)
5. Jones, L.A., Sarter, N.B.: Tactile displays: guidance for their design and application. Hum. Factors **50**, 90–111 (2008)
6. Merlo, J.L., Stafford S.C., Gilson, R., Hancock, P.A.: The effects of physiological stress on tactile communication. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco, CA (2006)
7. Ferris, T.K., Sarter, N.B.: Cross-modal links among vision, audition, touch in complex environments. Hum. Factors **50**, 17–26 (2008)
8. Gilson, R.D., Redden, E.S., Elliott, L.R.: Remote tactile displays for future soldiers (Tech. Rep. ARL-SR-0152.) Aberdeen Proving Ground, MD: Army Research Laboratory (2007)
9. Geldard, F.A.: The language of the human skin. In: Proceedings of the 14th International Congress of Applied Psychology, vol. 5, pp. 26–39 (1962)

Identifying Errors in Tactile Displays and Best Practice Usage Guidelines

Bruce J.P. Mortimer[1] and Linda R. Elliott[2](✉)

[1] Engineering Acoustics Inc., 406 Live Oaks Blvd, Casselberry, FL 32707, USA
bmort@EAIinfo.com
[2] HRED, ARL, US Army, Fort Benning, GA 31905, USA
linda.r.elliott.civ@mail.mil

Abstract. Wearable tactile cueing provides a significant opportunity for performance and safety improvement during human-in-the-loop tasks. However, wearable technology and the human factors associated with tactile cueing presents specific challenges and many pathways for potential errors. This paper reviews tactile cueing displays, usage characteristics and identifies potential error pathways. We use a model for tactile salience to describe adaptive cueing and case-specific examples of potential errors. We propose that tactile designers work towards intelligent systems that recognize user responses, adapt the tactile signal characteristics and close the loop between the display and the task.

Keywords: Tactile displays · Error pathways · Systems engineering · Cueing

1 Introduction

The sense of touch offers a relatively untapped and intuitive channel for communication and orientation. Tactile displays comprise of a number of small, wearable vibrotactile actuators, or tactors, that are distributed over the body. Tactile arrays have been integrated into clothing (e.g., vests, body suits, gloves and head attire) or implemented as wearable components that are often worn over clothing (for example torso-worn belts). Tactile arrays have the potential to provide effective cueing even under situations where the conventional communication channels such as visual, audio and even vestibular become disorientated [1, 2]. Meta-analytic investigations have shown that tactile cueing can enhance performance when added to visual cues, particularly under conditions of high workload [3]. Research has demonstrated that properly implemented tactile cueing yield significantly faster and more accurate performance than comparable spatial auditory cues [4].

Torso-mounted tactile displays have proven effective for navigation and communication in military field evaluations and were very highly regarded by experienced Soldier-evaluators [5–7]. These displays, when integrated with GPS, enabled Soldiers to navigate at night, hands-free (allowing the Soldier to hold his/her weapon) and eyes-free (allowing focused attention to surroundings as opposed to a display). In those studies, torso-mounted displays were proven effective for Soldier covert communications.

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_22

Any human interface display technology can be susceptible to errors. Tactile cueing displays can be intuitive, but human factors and systems designers have usually had less experience with this modality. We argue that they are susceptible to system design errors (implementation and design defects), and more often, insufficient consideration of environmental context, task demands, and user characteristics. A system may be built to specifications, and yet still be unsuited for purpose, when specifications are not matched to context of use. Therefore, designers need to be cognizant of these potential error pathways and in addition, develop operator training that minimizes and/or compensates for any errors. We discuss potential error pathways and suggest approaches for the identification and mitigation of errors.

2 Outlining a Systems-View of Tactile Cueing

2.1 Human Response to Vibrotactile Stimuli

Mechanoreceptors in the skin are responsible for detecting contact, deformation and motion across the skin surface. The type, location and density of mechanoreceptors varies with skin type and body location. Human response to vibrotactile stimuli can be defined in terms of our vibration detection threshold, the stimulus, and our spatial and temporal resolution. Spatial resolution can be measured in terms of two-point discrimination and single point localization, and as may be expected, there are large variations in perceptual performance across different body sites. There are many interactions between variables, especially time and space during perception [8]. Interactions can be in terms of perceptual changes, sensory illusions or masking. Another factor that must be recognized is sensory adaptation, where the response to a stimulus may vary with time. Adaption is mechanoreceptor and stimulus specific.

Most cueing constructs have utilized stimuli that fall into the vibration window between 200–300 Hz where the body is most sensitive to vibrations [9]. Although lower frequency stimuli can be distinct, the sensitivity for detection on the torso at 100 Hz is about 20 dB (Re 1 μm) lower than at 250 Hz [10]. Thus, to be perceptible, lower frequency 100 Hz vibration amplitudes must be ten times greater than vibrations at 250 Hz.

It has been estimated that the information capacity of the fingertip is 100 bits/s and this represents a practical limit for an experienced, undistracted subject [11]. When factors such noise, distractions, workload and postural movement is included, our tactile information capacity decreases [12]. There is also a difference between feeling a stimulus, and being able to assign meaning to that sensation. Therefore, the extension of tactile orientation cues to tactile command or language symbology is not straightforward.

The issue of tactile commands requires careful attention to the various factors that influence the production and perception of stimuli. It is also well known that the human channel capacity for information cannot be ignored when designing a human in the loop system [13]. Channel capacity varies with a number of stimulus and situation variables, including individual differences among users, like gender, age, and experience. Human perceivers are affected by their environment, task, workload, multisensory interactions, their experience, training, and expectations when presented with stimuli. These factors can all affect the accuracy and reliability of user interactions with the tactile display [14]. We have modeled the interaction between the environment, task, user and the tactile technology using tactile salience.

2.2 Tactile Displays

Tactile displays can be classified into; feedback, sensory substitution, sensory augmentation, communication and alerts systems [9]. The tactile display cueing components are usually comprised of combinations of relatively short tone-burst vibratory patterns. These patterns will have a natural or learned user association with a situational variable. Thus, the tactile display will comprise of distributed (wearable) actuator hardware as well as a means for the delivery of sensory stimuli (e.g., patterns/tactons/tactions).

One design approach to optimize tactor and taction effectiveness has been to configure tactors to have a "contactor" that oscillate perpendicularly to the skin, surrounded by a housing and radial gap [12]. The moving "contactor" is lightly preloaded against the body. When an electrical signal is applied, the "contactor" oscillates perpendicular to the skin, while the surrounding skin area is "shielded" with a passive housing. Using these criteria, all EAI's tactors are designed to minimize the effects of loading by the skin or garments [15].

An effective tactile display configuration for an intuitive communication of direction information is a torso worn belt [16]. It should be noted that the waist circumference varies between 22″ (5th percentile) to 28″ (95th percentile) for females and, 30″ to 40″ males [17]. Therefore, care must be taken in designing belts with different sizes and/or materials that can accommodate different waist sizes. The tactile cueing belt approach can be extended by using two rows of different types of advanced tactors to provide a wider design variability in the potential tactile stimulus. The dual belt approach (shown in Fig. 1) has been used to communicate both navigation information and incoming alerts [7]. In these experiments, the EAI EMR tactor was used primarily for navigation signals and the C-3 tactors were used to provide incoming alerts from a simulated robot asset. The C-3 produces a highly salient, "sharp" sensation as it operates at 250 Hz while the EMR provided a lower frequency, comfortable but less salient signal [14].

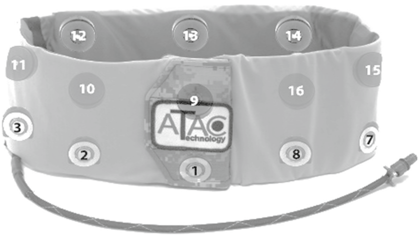


Fig. 1. Dual Belt (Engineering Acoustics, Inc.) tactile display comprising of a row (1–8) of eight EMR motor based tactors and a row (9–16) of eight C-3 tactors. The Dual Belt is constructed using stretchable material, with each tactor in a pod. The belt closure in the front using an overlap to accommodate the belly tactor (1 and 9) and electronic connection through an egress connector.

2.3 Model for Human Performance – Tactile Salience

We have previously developed a model based on tactile salience to describe the perception pathway [18]. 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. Tactile salience as mediated by three core factors; characteristics pertaining to the user, the technology, and the environment and their interactions. A simplified model for tactile performance is shown in Fig. 2.

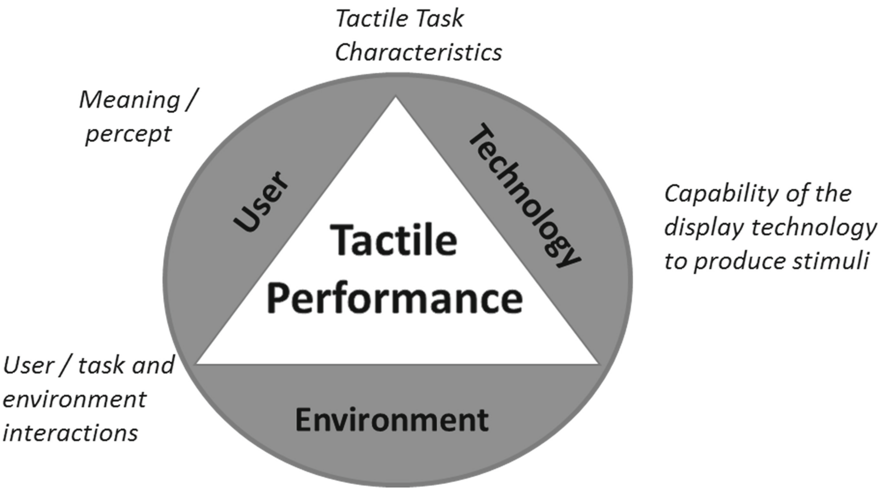

Tactile Task Characteristics
Meaning / percept
User
Technology
Tactile Performance
Capability of the display technology to produce stimuli
User / task and environment interactions
Environment

Fig. 2. Simplified model for viewing the interactions between the user, the tactile display (user-interface) technology and the environment.

Controlled comparisons showed some tactile patterns to be more salient than others [14]. 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 [19]. Simply changing the frequency of activation from slower to faster can change perceptions of urgency [9]. Patterns may be communicated across multiple body locations, such that an additional cue in a particular location can increase salience and indicate urgency.

Our research has investigated how Soldiers interact with tactile displays and we have developed some experience of user interface design and usability engineering (UE). Generally, these effects are measured in terms of a specific task performance measure. However, errors can, and do occur in tactile displays. It is beneficial to design systems and studies with an understanding of the potential error pathways.

3 Potential Errors

The tactile salience model can be a useful initial framework for investigating and identifying potential error pathways. Tactile devices and hardware are often developed and tested in laboratory or controlled environments while the intended usage may be significantly different (for example, indoors vs. outdoors). Therefore, our framework for discussing tactile display errors must include the user, environment and technology. While we separate by these categories, it will become evident that many errors arise from a systems perspective that considers the interrelationships among these three core sources. For example, technology characteristics pertaining to a particular system may be effective for one set of users, but not another. In the same way, the same technology may be effective in one environment or set of task demands but not another. Thus, we first list sources of error due to the technology, then we discuss sources due to mismatch of technology with environment, task demand, and user characteristics.

3.1 Sources of Error Related to the Technology

By technology, we refer to the characteristics of the tactile array and supporting components. If the tactor does not vibrate, or the vibratory stimulus is not detectable (i.e. not salient) to the user, an error of omission occurs. There are several component systems that may be involved in tactor display activation. Sensors are often responsible for activating tactile cues (i.e., tactions). Examples include use of tactions that indicate altitude from ground to helicopter pilots, dependent on altitude sensors, GPS cues that drive tactions cuing navigation direction, or software that identifies high priority information for taction alerts. Sensor systems often determine a figure of merit, or estimate of their reliability (and this may also change during a mission). Clearly, tactile display information must be based on accurate sensor data.

Tactor actuators are generally mechanically robust and relatively reliable. However, there are many pathways for technology failure including electrical and wireless connectivity, breakages and failures. Such problems may also prevent or distort tactile pattern generation or tactile activation. Errors may also arise due to power supply. If batteries are used, they must be adequate for the anticipated duration of use. Best practice guidelines include including mechanisms for robust communication transmission, error checking and system status monitoring for all tactile display system components.

Diagnosis of problems due to technology is a well within the realm of standard engineering design and use. Many other problems can arise when the system leaves the laboratory setting and is used in applications by users. These problems can be anticipated through consideration of the environment context, including task demands, and user characteristics. We discuss environment and task demands separately; however, it should be noted they can interrelate greatly. The following section addresses sources of error arising from a mismatch of technology that work, but is not suited to the environment, the task demands, and/or the user.

3.2 Sources of Error Related to User, Task and Environment

Mismatch with Environment. Tactors are versatile and have been used in many different environments including underwater [21], high vibration (noisy) rotorcraft [22] and even in space [23]. However, tactile technology characteristics may differ among these environments; a system that is effective for an indoor low-noise environment will likely not be effective in conditions of high noise. Similarly, systems must be designed to be compatible with underwater or in high altitude aircraft. Environmental factors that must be considered include regular exposure to water, vibration, wind, sand, humidity, and heat. Tactor system sensors also have potential limitations that are due to the environment (for example, operation in GPS-denied settings such as caves and tunnels). Each setting and use case may require hardware, stimuli and suitable mounting configurations to be effective. Mitigation involves appropriate design and iterative testing in conditions that replicate the environmental context, and is regularly performed within the realm of engineering design.

Mismatch with Task Demands. Task demands can have a large impact on the effectiveness of tactile display systems. Technology that is effective for stationary tasks may not be adequate when the user is highly active (e.g. crouching, crawling, running). Tactor characteristics associated with strength of signal and resistance to loading effects [15] greatly affect the ability of a user to perceive signals when activity levels are high [26]. Our understanding of tactile perception has also been typically based on a laboratory or a controlled environment. Simply changing the posture (sitting to standing or lying) can potentially change the perception.

Tasks that require high levels of attention, memory (e.g., short term memory), and/or cognition (e.g., information processing and decision making) contribute to overall cognitive workload [27]. As the user's cognitive workload increases, cognitive bottlenecks or demands from other sensory channels can rapidly decrease the effectiveness of complex tactile displays that involve recognition and discrimination of multiple tactile array signals. In these situations, tactile cuing should strive to alleviate the workload, through attention management, information prioritization (e.g., high priority alerts), and alleviation of higher workload tasks (e.g., tactile cues as augmenting or replacing visual cues) [25]. An example of this is the use of tactile direction cues for ground waypoint navigation, such that the user need not consult or interpret visual map-based displays; instead, they get a torso belt single direction cue that indicates "go this way". Tactile direction cues have been shown to be very rapidly understood, with little or no training, and thus are a good example of an intuitive design.

There is a balance between intuitive tactor display constructs, message length and task demands. Tactile stimuli patterns and stimuli must also be designed in terms of the intended usage, task and informational flow. Single stimuli can be missed while continuous stimuli may be ignored or masked. Changes may not be noticed. Therefore, stimuli must be robust and recognizable and what constitutes effectiveness may change based on the task, user-workload and environment.

Mismatch with User. The user is part of a human in-the-loop-system. There are many user preferences regarding what constitutes an intuitive, understandable tactile display output. There are also many individual differences [24] and these can be related to age, training (expectations) and prior experience. Care should be taken to evaluate displays on sufficient numbers of representative user populations (rather than samples).

Form and fit of the tactile system to the user's body is often a critical factor; wearable designs should preferably use design techniques that position the tactors with a pre-load against the user and recognize the potentially wide variation in the user's anatomical features. The tactor itself can potentially detect the body of the user by sensing the mechanical impedance load or physiological measures [20]. As described in Sect. 2, the threshold for detection of vibrotactile stimuli depends on the stimuli, the body location and many interacting factors that generally act to increase the threshold. We recommend that a best practice tactor display stimulus should also have a dynamic range that is at least 30 dB (Re 1 μm) above the threshold of vibration sensitivity to account for practical increases in the required threshold.

Users can potentially assemble the wearable tactor display incorrectly. Even if the array is sized correctly, wearable arrays can be mis-localized (for example, the front tactor not corresponding to the belly), or installed incorrectly (for example, upside down or even inside-out). Training is vital for novice users and in controlled studies should be scripted and user proficiency should be confirmed.

Users can also perceive stimuli in unintended manners. This is particularly important when designing tactile messages. If the tactor pattern occurs during activities when the user is focused on other tasks, the cueing can be unwanted or parts of the tactile pattern could be neglected (missing data). Multiple tactors that are simultaneously ON can be masked, while patterns that are too long too short may be missed. User's may also be subject to change blindness [27] where changes in the display are unnoticed. Delays in cueing information can lead to instability. Best practice guidelines would therefore include an identifying (attention focusing) feature at the start of a tactile message pattern to prevent missing the start of a pattern (can't recognize the middle or end) and utilizing tactile salience and multimodal sensory modalities to increase message priority.

4 Discussion

Table 1 provides an overview of the potential error pathways we have outlined for tactile display systems, and some example errors.

System and technology errors must be identified early and the user notified. This is standard practice for the technology components in the system. It is somewhat more difficult to recognize user and display use errors. The user may use the display technology incorrectly or the task and environment may change and potentially render the display less effective. User errors may be mitigated with effective user training and competency, but if they are unrecognized they can result in unsafe and ineffective tactile display use. The addition of smart technology that can adapt displays to situational and individual factors should be driven by theory and research-based guidelines [28].

Table 1. Potential error pathways in tactile displays.

Category	Description of the potential error	Impact
Technology	Hardware faults, sensor & communication errors.	Display error
User	Array usage errors, incorrect tactile pattern recognition, insufficient tactile salience & incorrect display data. Mode recognition & effectiveness of training	Display use error
Environmental	Specifications for tactile system sensors and displays operating environment exceeded	Display error
Task	Tactile display not effective during high-workload. Tactile display not adaptive to user requirements, information not recognized	User errors

There are several mechanisms that can be introduced to increase trust in the system. Primary to user trust is development of a system that is reliable and effective in the context of use. While tactile systems are generally passive, in terms of user interaction, care must be taken to ensure ease of use—in terms of preparing the system before use (e.g., creating or choosing signals and meanings), as well as during use. Signals should be easily felt, distinguished, and interpreted. If the tactile system is used to communicate verbal concepts (e.g., multiple alerting signals), training must ensure immediate and easy recognition. When tactile communications are added to visual and audio display, the resulting multimodal display has been associated with higher levels of trust, and allow users to not only identify potential errors but also use redundant information to synthesize and combine data more effectively than individual data streams.

Ultimately, systems should be designed to adapt to dynamic context – to recognize that as the environment, task demands, and/or user-workload change – the display must be adaptive and change the salience of the information. Adaptive systems can also be included intelligent systems where the response or reactions of the users is used to determine whether the user has processed the display information. Tactile display systems to date have not usually been bidirectional in nature, to allow the user to acknowledge or query a message construct. This is an option for intelligent systems where gestures and other user-interfaces can be used to provide a naturalistic user response.

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. Raj, A.K., Suri, N., Braithwaite, M.G., Rupert, A.H.: The tactile situation awareness system in rotary wing aircraft: flight test results. In: Proceedings of the RTA/HFM Symposium on Current Aeromedical Issues in Rotary Wing Operations, pp. 16.1–16.7. RTO NATO, Neuilly-sur-Seine, France (1998)
2. van Veen, H.A.H.C., van Erp, J.B.F.: Providing Directional Information with Tactile Torso Displays, EuroHaptics, Dublin (2003)

3. Elliott, L.R., Coovert, M.D., Redden, E.S.: Overview of meta-analyses investigating vibrotactile versus visual display options. 13th International Conference on Human-Computer Interaction, San Diego, CA (2009)
4. Gori, M., Vercillo, T., Sandini, G., Burr, D.: Tactile feedback improves auditory spatial localization. Front. Psychol. **5**, 1121 (2014)
5. 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)
6. Prewett, M.S., Yang, L., Stilson, F.R., Gray, A.A., Coovert, M.D., Burke, J., Redden, E., Elliot, L.R.: The benefits of multimodal information: a meta-analysis comparing visual and visual-tactile feedback. In: Proceedings of the 8th International Conference on Multimodal Interfaces, pp. 333–338. ACM (2006)
7. Elliott, L., Mortimer, B., Cholewiak, R., Mort, G., Hartnett, G., Pettitt, R.: Salience of tactile cues: an examination of tactor actuator and tactile cue characteristics. Technical Report ARL-TR-7392. Army Research Laboratory Human Research and Engineering Directorate (2015)
8. 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)
9. Jones, L., Sarter, N.: Tactile displays: guidance for their design and application. Hum. Factors **50**, 90–111 (2008)
10. Bolanowski, S., Gescheider, G.A., Verrillo, R.T.: Hairy skin: psychophysical channels and their physiological substrates. Somatosens. Mot. Res. **11**(3), 279–290 (1994)
11. Tan, H.Z., Durlach, N.I., Reed, C.M., Rabinowitz, W.M.: Information transmission with a multifinger tactual display. Percept. Psychophys. **61**, 993–1008. https://doi.org/PEPSBJ (1999)
12. Redden, E., Carstens, C., Turner, D., Elliott, L.: Localization of tactile signals as a function of tactor operating characteristics. Technical Report ARL-TR-3971. Army Research Laboratory Human Research and Engineering Directorate (2006)
13. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**(2), 81 (1956)
14. Elliott, L., Mortimer, B., Cholewiak, R., Mort, G., Zets, G., Pomranky-Hartnett, G., Pettitt, R. Wooldridge, R.: Salience of tactile cues: an examination of tactor actuator and tactile cue characteristics. ARL-TR-7392. US Army Research Laboratory, Aberdeen Proving Ground, MD (2015)
15. Mortimer, B., Zets, G., Cholewiak, R.: Vibrotactile transduction and transducers. J. Acoust. Soc. Am. **121**, 2970 (2007)
16. Gilson, R., Redden, E., Elliott, L.: Remote tactile displays for future Soldiers. Technical Report ARL-TR-3971. Army Research Laboratory Human Research and Engineering Directorate (2007)
17. Gordon, C.: Anthropometric survey of U.S. Army personnel: Methods and summary statistics. U.S. Army Natick Soldier Research, Development, and Engineering Center. Technical report Natick/TR-15/007, Natick, Massachusetts (2012)
18. Elliott, L.R., Mortimer, B.J.P., Cholewiak, R.W., Mort, G.R., Zets, G.A., Pittman, R.: Development of dual tactor capability for a Soldier multisensory navigation and communication system. 15th International Conference on Human Computer Interaction (HCI), Las Vegas (2013)

19. Brown, L., Brewster, S., Purchase, H.: Tactile crescendos and sforzandos: Applying musical techniques to tactile icon design. CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 610–615 (2006)
20. Gilson, R., Brill, J., Mortimer, B., Zets, G.: Electromagnetic field tactile display interface and biosensor. US Patent 7,696,860 B2
21. Dobbins, T., Samways, S.: The use of tactile navigation cues in high-speed craft operations. In: Proceedings of the RINA conference on high speed craft: technology and operation, pp. 13–20. The Royal Institution of Naval Architects, London (2002)
22. Rupert, A.H.: An instrumentation solution for reducing spatial disorientation mishaps. IEEE Eng. Med. Biol. **19**, 71–80 (2000)
23. van Erp, J.B.F.: Tactile Displays for Navigation and Orientation: Perception and Behavior. Mostert and Van Onderen, Leiden (2007)
24. Hancock, P., Elliott, L., Cholewiak, R., van Erp, J., Mortimer, B., Rupert, A., Schmeisser, Redden, E.: Tactile cueing to augment multisensory human-machine interaction. Ergon. Des. **23**(2), 4–9 (2015)
25. Elliott, L.: Bruce Mortimer, Context Sensitive Tactile Displays for Bidirectional HRI Communications, AHFE Conference, Orlando, July 29 2016
26. Apkarian, A.V., Stea, R.A., Bolanowski, S.J.: Heat-induced pain diminishes vibrotactile perception: a touch gate. Somatosens. Motor Res. **11**(3), 259–267 (1994)
27. Wickens, C.: Multiple resources and performance prediction. Theor. Issues Ergon. Sci. **3**(2), 159–177 (2002)
28. Mortimer, B., Elliott, L.: Information transfer within human robot teams: multimodal attention management in human robot interaction. IEEE Conference on Cognitive and Computational Aspects of Situation Management, Savannah, GA (March 2017)

Can You Feel Me Now? Wearable Concept for Soldier Communications

Gina Hartnett[1](✉), Linda Elliott[2], Lisa Baraniecki[3], Anna Skinner[3],
Kenyan Riddle[4], and Rodger Pettitt[2]

[1] Army Reserach Laboratory, Human Research and Engineering Directorate,
Fort Rucker, AL, USA
Regina.A.Hartnett.Civ@Mail.Mil
[2] Army Reserach Laboratory, Human Research and Engineering Directorate,
Fort Benning, GA, USA
{Linda.R.Elliott.Civ,Rodger.A.Pettitt.Civ}@Mail.Mil
[3] AnthroTronix, Inc., Silver Spring, MD, USA
{Lisa.Baraniecki,Anna.D.Skinner}@atinc.com
[4] Aptima, Inc., Performance Assessment and Augmentation Division,
Orlando, FL, USA
KRiddle@aptima.com

Abstract. Soldiers and other tactical teams often need critical information regarding direction and distance. The challenge then becomes, how best to convey this information to someone who must act as quickly as possible, particularly when that person is out of line of sight. In this study, we compared three methods to communicate direction and distance: (a) tactile for both, (b) mixed - tactile for direction and audio for distance, and (c) audio for both. Twenty Soldiers participated in each communication condition. Performance measures included response time and accuracy of direction and distance. Results suggested that the mixed condition reduced time needed to accurately determine direction and distance to a threat compared to the all-tactile or all-audio communication conditions.

Keywords: Wearable concepts · Soldier · Dismount · Warfighter · Communications · Instrumented glove · Tactile · Tactile belt · Pointing · Gestures · Human factors · Human-systems integration · Systems engineering

1 Introduction

1.1 Background

A future vision of Soldier-robot working teams assumes higher levels of robot intelligence and semi- independence, using metaphors of tactical execution like a robotic "wingman" or military working dog [1, 2]. Continuous and deliberate control would no longer be necessary; instead, Soldier-robot interactions would be closer to a tactical two-way dialogue. Thus, it is critical to improve human-robotic interaction (HRI) capabilities enabling Soldiers and robots to communicate with each other quickly and easily.

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_23

Advanced systems have been developed that can assess direction and distance to a particular location. The goal is to be able to convey location information to a person, as naturally as one would point to a location, when two people are standing together (e.g., the robot went around "that" building). The challenge is how to convey that information when a pointing gesture cannot be seen. One example is the use of a tactile interface to communicate direction [3], such as in waypoint navigation [4], where Soldiers navigated more quickly and easily when tactile direction cues were provided. In the same manner, tactile displays were used to convey the same information as Army hand and arm signals [5].

A recent capability, demonstrated during this effort, is the ability of an instrumented glove to convey direction and distance information, by means of a pointing gesture, to another person who is out of line of sight. The instrumented glove provides direction and distance information through a pointing gesture (direction), combined with the use of finger gestures to count distance information. These new capabilities allow rapid messaging of direction and distance information to a receiving Soldier. In this study, we investigated the effectiveness of tactile and audio communication channels to indicate direction and distance without use of visual cues (e.g., no visual map display or visible pointing gestures).

1.2 Purpose

The current effort sought to investigate advanced concepts in multisensory interfaces to help operators understand spatial information more quickly and easily. Direction and distance information was conveyed to Soldiers using three methods: (a) direction and distance through audio communications, (b) direction by tactile belt and distance through audio, and (c) direction and distance through the tactile belt. In this Soldier-based evaluation, our goals were to collect performance-based measures to assess these capabilities based on accuracy and speed of comprehension.

1.3 Gestures for Target Location

A preliminary goal for this evaluation was to demonstrate an instrumented glove that can convey direction and distance information through pointing gestures combined with finger-based indication of distance. Pointing gestures for HRI have been developed over several years, either to convey direction information or to clarify ambiguous speech-based commands [6]. While the pointing gesture is natural and intuitive, recognition of "where" and "what" can be challenging depending on task context. Advancements in instrumented glove technology are enabling determination of azimuth from a point gesture; when combined with a global positioning system (GPS)-based wearable device, both direction and distance can be determined through sensors within the glove [4].

Automated electronic capture of hand and arm signals via instrumented glove technologies enables commands to be initiated and instantaneously sent to all team members simultaneously without requiring line-of-sight. These electronic signals can be presented to both human and robot team members. The sensors necessary for gesture recognition are small, lightweight, and can be unobtrusively integrated into warfighters' current field gloves. Hand movements can thus be used for standard hand signal commands and can be presented to human team members via a variety of modalities.

A prototype system developed by AnthroTronix was demonstrated to researchers in the US Army Research Laboratory's Human Research and Engineering Directorate located at Fort Benning, GA. The system was able to convey information regarding cardinal direction and distance through use of a pointing gesture for direction combined with a counting gesture (number of fingers held straight) for direction. The system was generally reliable though somewhat subject to variances due to GPS calibration. For this effort, our evaluation was not focused on the technology capability of the instrumented glove, but rather on identification of the best means to convey that information to another person out of line of sight. Results of this study thus generalize beyond gestures to include any messaging of spatial information to the operator.

1.4 Tactile Interface

While the instrumented glove provides the means for gestural signals out of line of sight, the reception is accomplished through a torso-mounted belt with vibrating tactors. The haptic modality has proved to be a reliable and covert conduit for the conveyance of critical information during infantry tactical operations. For example, van Erp [3] showed that a localized vibration on a waist belt could easily and accurately be interpreted as a direction in the horizontal plane since it is intuitive to infer direction from the torso, which is relatively stable. Recently, torso-mounted haptic displays have proven very effective for navigation in field evaluations [4]. These interfaces, if integrated with GPS, enable dismount warfighters to navigate in low-visibility conditions, hands-free (allowing the Soldier to hold his/her weapon), mind-free (not having to pace count) and eyes-free (allowing focused attention to surroundings rather than a visual display) [4, 7, 8]. Torso-mounted interfaces have also proved effective for warfighter communications. However, haptic systems must be integrated with visual and control systems to support optimal display of certain types of complex information and to enable map-based situation awareness and easy input of waypoints. Multi-modal information presentation supports redundancy and enables warfighters to attend to the individual modality or combined information channels of choice in any given situation. Additionally, intelligent wearable computing devices allow warfighters to communicate with each other, obtain information, and control remote devices without impeding their ability to perform tasks in a field environment [9]. In this study, we included the tactile modality as an option that was expected to be more easily and quickly recognized for spatial information, particularly for direction cueing.

2 Equipment

2.1 COMMAND System

The Communication-based Operational Multi-Modal Automated Navigation Device (COMMAND) integrates an instrumented glove (Fig. 1) for automated gesture-based communication and control (not the focus of this study), a haptic display belt, and a GPS-enabled ruggedized handheld computer. The COMMAND technology is designed to support gesture recognition, navigation support, wireless communication, robotic control, and multi-modal information presentation. The haptic interface can receive signals from the instrumented glove (i.e., pointing gestures), enabling Soldier-Soldier communications. The glove contained ten 9-axis sensors (3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer). Data from the glove sampled at a rate of 100 Hz. The glove was tethered to a tablet, which was used to transmit the wireless command signal.

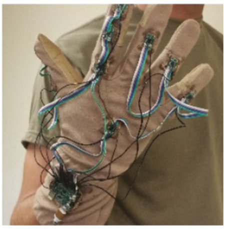


Fig. 1. AnthroTronix instrumented glove

2.2 NAVCOM Tactile System

EAI's ATA Dual Belt (shown in Fig. 2) represents a state-of-the-art, wearable vibrotactile array, suitable for a wide variety of military, biomedical, research, and commercial applications. The EMR tactor (shown in Fig. 3) is a miniature vibrotactile transducer optimized to create a strong localized sensation on the body. This tactor uses an eccentric motor in a proprietary and patented configuration to provide low-frequency, high-displacement contactor vibration. The C-3 tactor (shown in Fig. 4) is a miniature vibrotactile transducer optimized to create a strong localized sensation on the body. A body-referenced arrangement of tactors activated individually, sequentially, or in groups, can provide intuitive "tactile" instruction to a user.

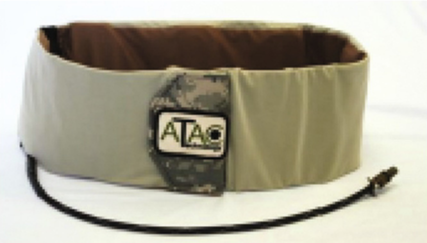


Fig. 2. EAI dual tactile belt

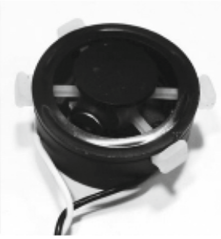


Fig. 3. EMR tactor

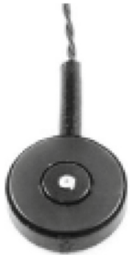


Fig. 4. C-3 tactor

2.3 Smartphone Integration

A GPS-enabled smartphone was used for data processing and signal communication (Fig. 5) performed on a Samsung Galaxy S4 device. The smartphone included a touchscreen and visual display, an Android operating system, custom gesture recognition software, and tactor controller software, as well as embedded GPS and wireless communication capabilities. The operators did not interact directly with the smartphone.

Fig. 5. Samsung galaxy GPS-enabled smartphone

3 Method

3.1 Participants

Twenty Soldiers recruited from the Officer Candidate School at Fort Benning, GA, participated in this study. All participants had a BS degree or higher—two had PhDs. Age ranged from 22 to 32 (average = 26.04). Twelve were female. Three participants were left-handed. Uniform size ranged from XS to L.

3.2 Procedures

Soldier-participants were briefed on the purpose of the target localization experiment. They were told they would be trained on information received through audio and/or tactile displays. After training, they participated in the three communication conditions, with each condition providing information on 10 targets. Performance data were collected through trained observers. After all performance sessions were complete, the participants filled out questionnaires pertaining to each condition.

3.3 Training

Soldiers were first trained on the signals for the commands they would receive. For example, the number of vibrations would indicate the distance, in 10 s of meters away.

Additionally, the direction of the threat would be indicated by the location of the vibration on the belt. Likewise, in the audio condition, the distance would be conveyed by voice over the "radio" (smartphone speaker). Once in place, the Soldier faced north. Soldiers were told that another Soldier spotted a threat and would be communicating that threat either by "radio" or by tactile belt. Soldiers were trained until proficient in each condition.

3.4 Target Localization Task Demands

After training, each Soldier responded to incoming information to interpret and measure accuracy of the cardinal directions (N, NW, NE, W, SW, S, SE, and E) (see Fig. 6). The letter R was always placed facing north. The Soldier responded to incoming information by facing the letter corresponding to the incoming direction, stating the letter representing that direction, and stating the distance information (Fig. 7).

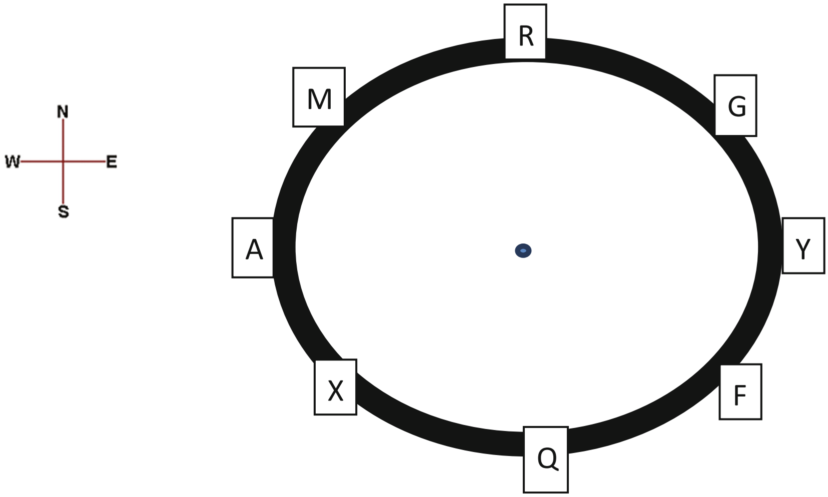


Fig. 6. Representation of cardinal directions. Each soldier stood in the center of the circle, facing north (starting position prior to each trial).

For the all-audio condition, the distance and direction were received by audio over the "radio". For example, the Soldier would hear 30 m west. Given that prompt, the Soldier would turn and face the direction indicated and verbalize the distance heard. The Soldier's correct response should be "A 30 m". Note that in the audio condition, it

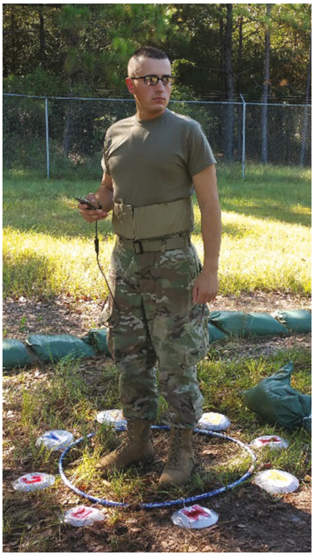


Fig. 7. Soldier preparing for cue facing north (R)

was the added task of the Soldier to know that he was facing north and to determine which direction was west. The Soldier would then face that direction and give the corresponding letter found at that location on the direction ring. The distance would be a repeat of what was heard and remembered from the “radio” (Fig. 8).

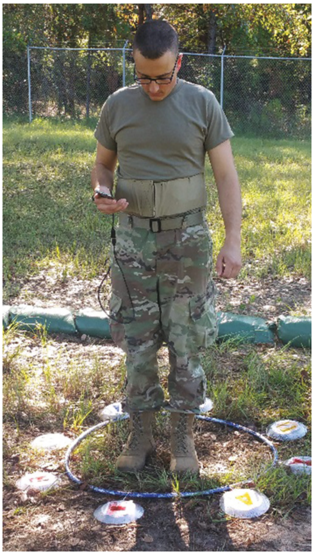


Fig. 8. Soldier moving in the direction of threat (M)

In the all-tactile condition, the Soldiers received a vibration on the belt that corresponded to the location of the threat. Additionally, the number of vibration taps felt would indicate the distance (in 10 s of meters) away from the threat. Upon feeling each cue, the Soldiers would turn their bodies to face the direction they felt on their torso. This direction corresponded to a letter on the direction ring. For example, if the Soldier received two vibrations at his 3 o'clock, the correct response should be "Y 20 m".

Finally, in the mixed condition (direction via tactile belt and distance via audio), the Soldier would receive a vibration on the belt that corresponded to the location of the threat and an auditory cue that indicated the distance away from the threat. For example, if the Soldier received a vibration at his 6 o'clock and heard via the "radio", 40 m, the correct response should be "Q 40 m". After each response, the Soldier would turn to face north again and prepare for the next signal. This was true in all conditions.

3.5 Target Localization Performance Measures

Communication performance regarding direction/range included speed of Soldier response to incoming information (i.e., number of seconds, logged by the data observer). Accuracy of direction response was scaled by counting the number of direction responses between the correct answer and the given answer. If the response was completely accurate, accuracy (error) = 0; if the response was one off, accuracy was = 1, and so on. Any equipment failures were noted.

4 Results

The three main variables reflected performance in the target localization task. Direction indicates the degree to which the direction indicated by the Soldier was correct. Distance indicates the degree to which the distance indicated by the Soldier was correct. Time reflects the time taken for Soldier response.

4.1 Direction

In the tactile condition, direction was indicated by a tactile direction cue. In the Audio condition, direction was indicated by an audio speech cue indicating a cardinal direction (N, S, W, E, NW, etc.). In the mixed condition, the direction was indicated by a tactile direction cue.

Errors were measured by the number of direction categories in which the response was off in location from the correct answer. A zero was given if the Soldier indicated the direction correctly: 1 if he or she was one off, 2 if two off, etc. Table 1 provides the mean percentage of error by condition as portrayed in Fig. 9. While the error percentage was somewhat lower in the mixed condition, the difference was not statistically significant (F 2, 38 = 1.035, p = 0.365, $\eta\rho^2 = 0.05$).

Table 1. Mean percent error in direction commands

Channel	Mean	Std. Dev.	N
Tactile-tactile	0.115	0.11	20
Tactile-mixed	0.065	0.10	20
Audio-Audio	0.125	0.21	20

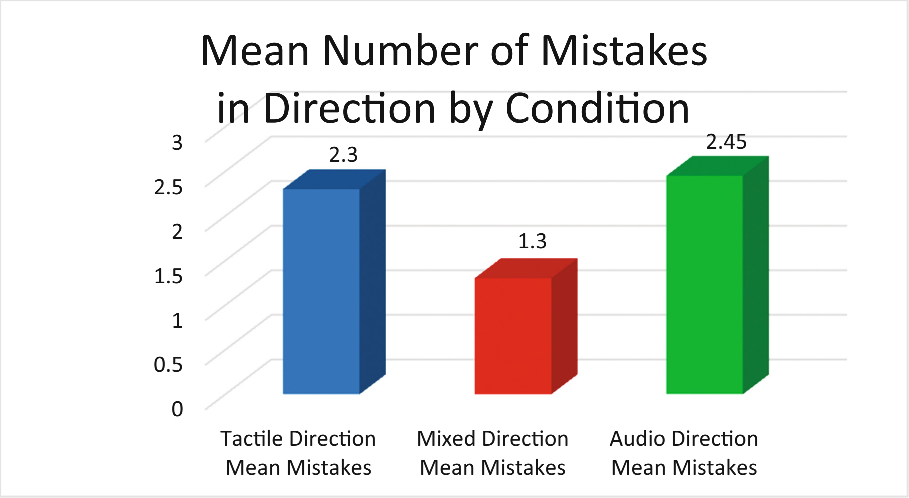


Fig. 9. Mean number of mistakes in direction by condition

4.2 Distance

In the tactile condition, distance was indicated by a succession of tactile cues, which must be counted. In the Audio condition, distance was indicated by an audio cue (i.e., voice recording) stating the distance (i.e., 20 m), such that the Soldier need only repeat the information correctly. In the mixed condition, the distance was indicated by the voice recording.

Errors were measured by the number of direction categories that were off. Categories were in multiples of 10 m. A zero was given if the Soldier indicated the direction correctly: 1 if he or she was one off, 2 if two off, etc. Table 2 provides the mean percentage of error by condition. The percentage was significantly lower in the mixed condition (F 2, 18 = 6.998, p = 0.006, $\eta\rho^2 = 0.44$) (Fig. 10).

Table 2. Mean percent error in distance commands

Channel	Mean	Std. Dev.	N
Tactile-Tactile	.085	0.113	20
Audio-Mixed	.000	0.000	20
Audio-Audio	.090	0.231	20

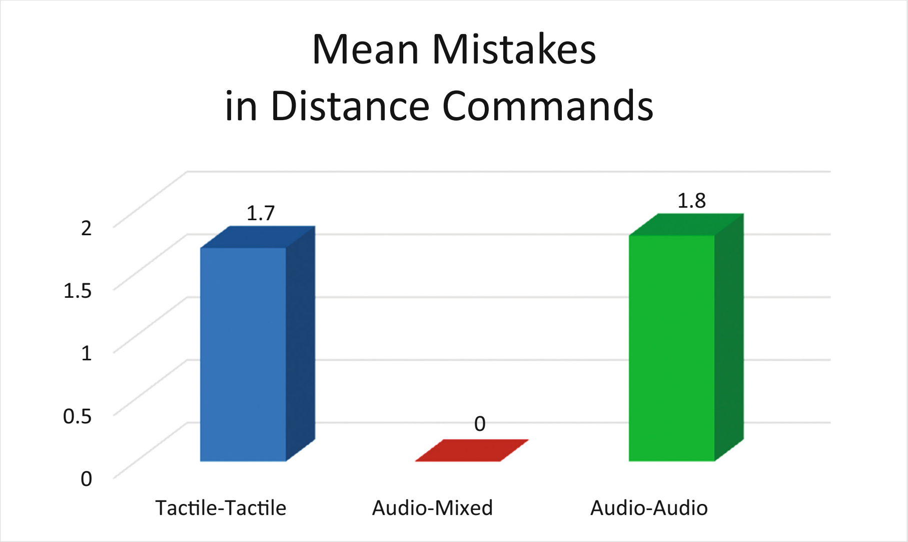


Fig. 10. Mean mistakes in distance by condition

4.3 Time

A stopwatch was used to record the number of seconds for the Soldier to respond for each target, in each condition. Table 3 shows mean time for each condition, as portrayed in Fig. 11. Mean time for the audio condition was much higher. Times were significantly different (F 2, 38 = 183.40, p = 0.00, $\eta\rho^2 = 0.90$).

Table 3. Mean response time

Channel	Mean	Std. Dev.	N
Tactile-Tactile	3.6300	0.46915	20
Audio-Mixed	2.9600	0.32509	20
Audio-Audio	5.1950	0.60739	20

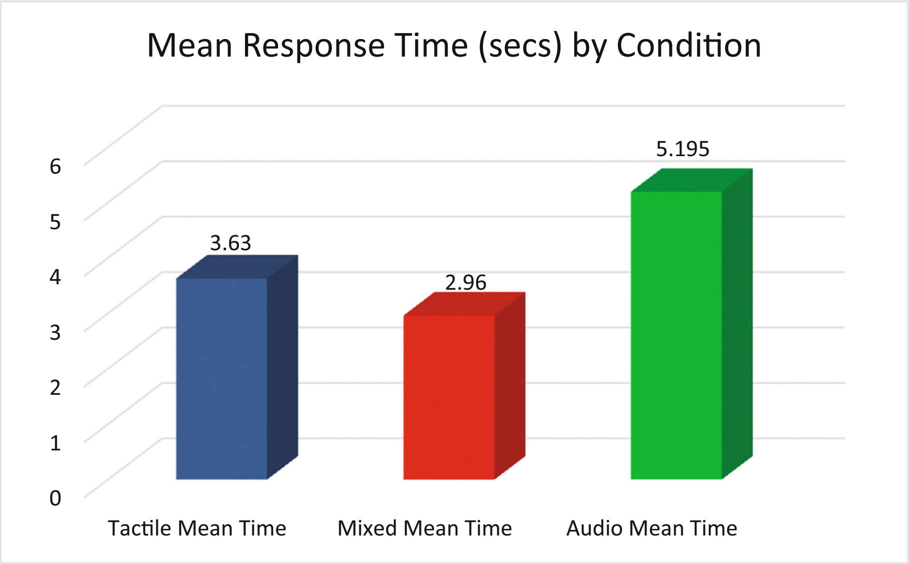


Fig. 11. Mean response time by condition

5 Conclusion

The purpose of this study was to identify the best options for communicating spatial information to an operator. In this study, we compared three methods of communication of direction and distance: (a) tactile-both direction and distance through a tactile display, (b) mixed-tactile for direction and audio for distance, and (c) audio-both direction and distance through audio communications. Twenty Soldiers participated in the three communication conditions, with each condition providing information on 10 targets. Performance measures included response time and accuracy of direction and distance.

Results indicated that use of the instrumented glove and tactile belt reduced time needed to accurately determine direction and distance to a threat in the mixed condition (tactile for direction and audio for distance) compared to the all-tactile or audio communication conditions. Recent systems that integrate torso-mounted tactile displays have allowed Soldiers to more quickly understand, communicate, and respond to battlefield dynamics. Given these technology developments, further research and guidelines are needed to optimize the use of these multimodal options.

Acknowledgements. This report was accomplished through the support of the US Army Research Laboratory Human Research and Engineering Directorate programs regarding multi-year investigations of human-robot interaction.

References

1. Phillips E., Rivera, J., Jentsch, F.: Developing a tactical language for future robotic teammates. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 54(1), pp. 1283–1287 (2013)
2. Redden, E., Elliott, L., Barnes, M.: Robots: the new team members. In: Coovert, M., Foster, L. (eds.) The Psychology of Workplace Technology. Published by Society for Industrial and Organization Psychology: Frontier Series (2013)
3. van Erp, J.B.F.: Presenting directions with a vibro-tactile torso display. Ergonomics **48**, 302–313 (2005)
4. Pomranky-Hartnett, G., Elliott, L., Mortimer, B., Pettitt, R.: Soldier based assessment of dual-row. ARL-TR-7397 (2015)
5. Pettitt R., Redden, E., Carstens, C.: Comparison of Army Hand and Army Signals to a Covert Tactile Communication System in a Dynamic Environment. ARL-TR-3838 (2006)
6. Perzanowski, D., Schultz, A., Adams, W., Marsh, E.: Using natural language and gesture interface for unmanned vehicles. NRL; Code 5510. Report No.: ADA435161. Navy Center for Applied Research in Artificial Intelligence, Washington, DC (2000)
7. Elliott, L., Schmeisser, Redden, E.: Development of tactile and haptic systems for U.S. Infantry navigation and communication. In: Proceedings of the 14th International Conference of Human Computer Interaction, Human Interference and the Management of Information. Interacting with Information Lecture Notes in Computer Science, 2001, vol. 6771, pp. 399–407. Springer, Orlando, FL (2011). doi:10.1007/978-3-642-21793-7_45
8. Elliott, L., Redden, E.: Reducing workload: a multisensory approach. In: Savage-Knepshield, P. (ed.) Designing Solder Systems: Current Issues in Human Factors. Ashgate, Farnham (2013)
9. Vice, J., Lockerd, A., Lathan, C.: Multi-modal interfaces for future applications of augmented cognition. Foundations of Augmented Cognition, pp. 21–27 (2005)

Development of a Vibrotactile Metronome to Assist in Conducting Contemporary Classical Music

Patrick Ignoto(✉), Ian Hattwick, and Marcelo M. Wanderley

Input Devices and Music Interaction Laboratory (IDMIL),
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT),
McGill University, Montreal, Canada
{patrick.ignoto,ian.hattwick}@mail.mcgill.ca,
marcelo.wanderley@mcgill.ca

Abstract. Contemporary classical music conductors are often required to perform music that requires a click track for accurate tempo following. Auditory click tracks can be obtrusive as they block an ear and interfere with the conductor's perception of the music being performed. We propose a device and framework that uses the tactile sense to provide a similar functionality.We outline the unique requirements such a system would require gathered from interviews with an expert conductor. A series of prototypes were developed and presented to the conductor in order to design a device to meet these requirements. A test conducted to verify the current system's timing performance showed a mean error of 0.54% compared to the audio click track.

Keywords: Interface design · Music technology · Vibrotactile display

1 Introduction

Contemporary orchestral conductors frequently perform pieces that require an ensemble to synchronize with a pre-recorded electronic audio track. Audio click tracks, played through an earpiece placed in one of the conductor's ears, are often used to assist in this task as they are inexpensive and easy to use. However, these devices are intrusive, blocking an ear and interfering with the conductor's perception of the music being performed, and they are generally designed for use in music with a prominent and consistent rhythmic pulse.

Using a vibrotactile display [1], we can render vibrotactile sensations based on signals sent via a computer. This allows a conductor to feel the tempo rather than hear it. We present our work in creating a system which allows for the flexible creation of vibrotactile pulses in order to assist an expert conductor in performing music that normally requires an audio click track, therefore not directly interfering with the conductor's hearing of the ensemble performance and being appropriate for use with contemporary classical music.

This paper discusses the unique requirements and implementation of such a device. The vibrotactile click track was designed in close collaboration with Professor Guillaume Bourgogne, director and conductor of the McGill Contemporary Music Ensemble

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_24

(CME). User interviews with Prof. Bourgogne were used to determine the requirements for the specific application and to refine the system's features and usability. Laboratory measurements were then used to evaluate the performance of the device. The current version of the metronome meets the requirements gathered through the user interviews, including the ability to translate an existing audio file containing a click track into a vibrotactile click track, the ability to fully customize the vibrotactile pulses sent to the performer, and being able to represent a wide variety of rhythmic and tempo variations in any given contemporary musical piece.

2 Vibrotactile Notification in Music

Haptics is the field of research concerning the sense of touch, typically used for both kinesthetic and tactile sensations [2]. In computer music, the synthesis of tactile stimuli is often used by researchers to notify or give feedback to musicians [3]. Previous research has given some credence to developing a vibrotactile click track system. This section outlines some previous studies that verified the use of vibrations as a metronome in music and examines systems that used vibrations for notifications to live concert performers. As well, it introduces the Vibropixels, the vibrotactile display system that we used in the development of our vibrotactile click track.

2.1 Vibrotactile Metronomes

Giordano and Wanderley [4] performed a pilot study that compared guitar performance with both an auditory and tactile metronome. Using a prototype system built with off the shelf parts, they asked four guitar players to play a G major scale at 60 and 120 BPM using both an auditory and tactile metronome. The tactile metronome output discrete haptic pulses much like the auditory metronome outputs clicks for each subdivision of time. The data they gathered from this pilot study preliminarily showed that a tactile metronome can reliably cue a performer to the tempo with accuracy comparable to an auditory metronome.

Much of the body of research involving beat detection with vibrotactile stimulation is with regards to tapping to the beat of rhythms rather than keeping time with a metronome. Ammirante et al. [5] studied tapping the beat of the rhythms with vibrotactile stimulation. Using voice coils embedded in a chair, they asked participants to tap to the beat of a simple and complex rhythm played using auditory, tactile, and bimodal (auditory-tactile) cues at large and small magnitudes. The inter-tap interval was measured to determine how well they tapped along to the cues. The data found that people could follow along with the tactile cues as well as the auditory cues for simple rhythms if the tactile cues were sufficiently prominent (large magnitude condition).

Commercial solutions for vibrotactile metronomes, such as the Soundbrenner Pulse [6], are also available. The Soundbrenner Pulse is a watch-like wearable device that is controlled via a smartphone app, or via MIDI beat clock generated by a Digital Audio Workstation (DAW). The smartphone app allows customizing the pulse to only 9 available presets and outputs discrete haptic pulses only. Changing the tempo or time signature with the smartphone app requires manual user input. Changing the tempo via MIDI beat clock is possible if the DAW project is setup correctly beforehand, but changing the time signature requires manual user input. Therefore, this device is geared towards beginners and studio musicians who need to be tightly synchronized to a single tempo and time signature. Since many works in the contemporary music repertoire can change tempo or time signature multiple times within the same piece, the commercial solution could not be used for the development of this project.

2.2 Live Vibrotactile Notification

Some products have also been developed which send vibrotactile notifications to live performers. The *NeVIS* system by Hayes and Michalakos [7], is a networked vibrotactile communication system that can be used by improvisational performers for notification and interaction. Using a custom haptic device with three coin-cell actuators, the system sends cues to a group of improvisational musicians. These cues include denoting sections in a piece, tempo information, and communication between performers.

Schumacher et al. [8] incorporated vibrotactile notifications into the CIRMMT Live Electronics Framework (CLEF), a Max-based modular framework for use with live performers in mixed music pieces. Using a hardware prototype with two coin-cell actuators, a module for tactile feedback was added to the CLEF software by the developers. This vibrotactile notification module allowed for two types of modalities, an individual mode where each of the actuators were triggered individually to allow for discrete pulses or continuous motions between the actuators, and a balanced mode which is used to display relative values by changing the intensity ratios between the two actuators. They evaluated the use of tactile notifications by running tests with a performer. One test evaluated a haptic click track modality, where discrete haptic pulses were used to convey tempo information. Another test notified performers of the location of a sound source in a sonic spatialization system, which was an irregular and non-deterministic notification. In both cases, the performer said that the tactile notifications were very effective and unobtrusive. A later study by Frid et al. [9] quantified more perceptual information about tactile modalities using the CLEF system. From the data generated by the experiments, the authors provide suggestions for designing tactons (tactile icons) [10] in musical settings, such as duty cycle scale and haptic inter-onset times.

2.3 Vibropixels

The Vibropixels, developed at the Input Devices for Music Interaction Laboratory (IDMIL) at McGill University, is a reconfigurable, scalable vibrotactile display system [11]. A transmitter connected to a PC wirelessly sends control messages, generated by an interactive software (Max/MSP), to one or many different Vibropixels in the network, because the Vibropixels are individually or group addressable. The parameters of the control message represent an envelope for the vibrotactile pulse and the receiving Vibropixel outputs a pre-programmed shape based on this envelope. Two motors, a coin cell and a pager motor, are available and individually controlled allowing for different textures of haptic pulses to be generated. The Vibropixels modular design allows them to be reconfigured for use in a wide range of applications. They can be used as a wearable device and are easily programmable and customizable. For these reasons, the Vibropixels were chosen as the platform for developing the tactile metronome (Fig. 1).

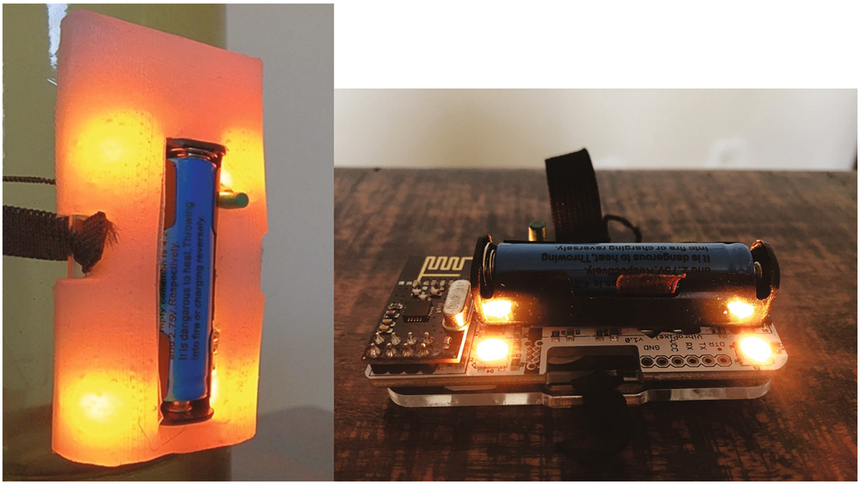


Fig. 1. Vibropixels, a wearable vibrotactile display system used in developing the vibrotactile click track system. The left image shows a Vibropixel with its silicone cover. The right image shows a Vibropixel without the silicone cover.

3 Design Requirements

A series of user interviews were performed with Professor Bourgogne to gather the requirements necessary for the initial metronome design. The interviews aimed to highlight what the device should do, how the vibration patterns should feel, as well as general usability requirements. The following section summarizes all the major requirements such a device needed based on interview answers.

The System Should Be Able to Convert Already Existing Click Tracks Made for Works in the Repertoire. Audio click tracks are only used for pieces where an orchestra must synchronize to a pre-recorded electronic audio track. However, there are many existing pieces where this is the case, and so audio click track files already exist for these pieces. Since the click tracks contain the precise tempo information for a musical piece, the vibrotactile click track should be modelled from this, in order to reduce the work required to program the precise timing information. Therefore, the system must be able to take an audio click track file, extract the information from it, and convert it to the vibrotactile system.

The System Should Handle the Aesthetic Qualities of Contemporary Music. Contemporary works in the repertoire typically do not have a single tempo or time signature throughout the piece. Previous works on haptic metronomes are geared towards session musicians or beginners who need to synchronize to one tempo and time signature. Manual user input is required to change these settings and might cause missed beats if the user is not fast enough. Live performances of contemporary works need a system where the tempo and time signature can change automatically, without missing any beats. As an example, we were given the audio click track to the musical piece Charge by Raphaël Cendo. The piece changes tempo and time signature several times within its 15-min runtime.

Downbeats Should Feel Different than Upbeats. The downbeat marks the first beat in a measure of music. This is useful to a conductor for determining time signature information. These are typically denoted in audio click tracks with a distinct click and should similarly be denoted in a vibrotactile click track.

Pulses Should Feel Continuous and Not Feel Like Individual, Discrete Impulses as in Previous Works. Previous works used discrete haptic impulses in order to convey tempo information. This means they quickly rise to a maximum vibration, held there for a specified amount of time and then went down to no vibration, much like an audio click track, where percussive sounds are used to designate time subdivisions. However, Prof. Bourgogne preferred something that flows more naturally than discrete impulses do. In the interviews, he uses the example of a pendulum motion to describe how he believes the pulses should feel, where the pendulum alternates between two positions, but never stops moving.

Pulses Should Have a Ramp Up to a Maximum Peak. A ramp up to a maximum peak allows for anticipation of the next beat. This gives the conductor a better sense of when the next beat occurs, especially when there is a change in tempo.

The Peak of a Pulse Should Be Where the Click is Found in the Audio Click Track. The maximum peak of the vibration should be located where the click of the audio track is heard. This allows for synchronization of the vibrotactile click track with the audio click track. It is also a good way to evaluate how well the vibrotactile click track outputs the timing information generated using the audio click track. This should be as distinguishable as possible to have a good sense of the pulse in the click track.

Length of a Pulse Should Be Dependent on the Tempo. In other words, faster tempos should have shorter pulses than longer tempos. This, much like one of the previous requirements, allows for anticipation of the next beat. At a faster tempo, the ramp up will be much faster than at slower tempos because there is less time between pulses so the overall pulse will be shorter. This, according to Prof. Bourgogne, would allow a conductor to "feel the variations of tempo with natural motions".

4 Implementation

An iterative design approach was used in the implementation of the device [12]. Prototypes were quickly designed and tests were run to see if they met the requirements.

Prototypes were then presented to Prof. Bourgogne and his comments were used to refine the design and the requirements. This section discusses some early design choices that affected the final prototype and discusses the final prototype implementation in detail.

Early prototypes initially did real-time audio analysis of the click track in the software Max/MSP. The audio click track was loaded into the patch and using the *bonk~* object by Puckette [13], a real-time analysis of the percussive envelopes in the audio was performed. This allowed for identification of the downbeats and upbeats since they had distinct percussive envelopes. When either envelope was identified, a pre-set vibrotactile envelope was triggered and the associated control message was transmitted to the Vibropixel. This allowed for discrete haptic pulses to be sent out based on the real-time analysis of the audio click track. Further refinements accessed the firmware of the device and added two new types of haptic pulse. When one of these types of haptic pulses were triggered with a unique control message from the transmitter, the Vibropixel would display this type of pulse until triggered off. The haptic pulses were based on comments gathered from the user interviews. One was based on an exponential rise and decay model and the other was based on a triangular model. Both types of haptic pulses prevented the motor from stopping completely, which means it vibrated at a low "base amplitude" specified in the control message so that the pulses felt continuous, which is one of the requirements.

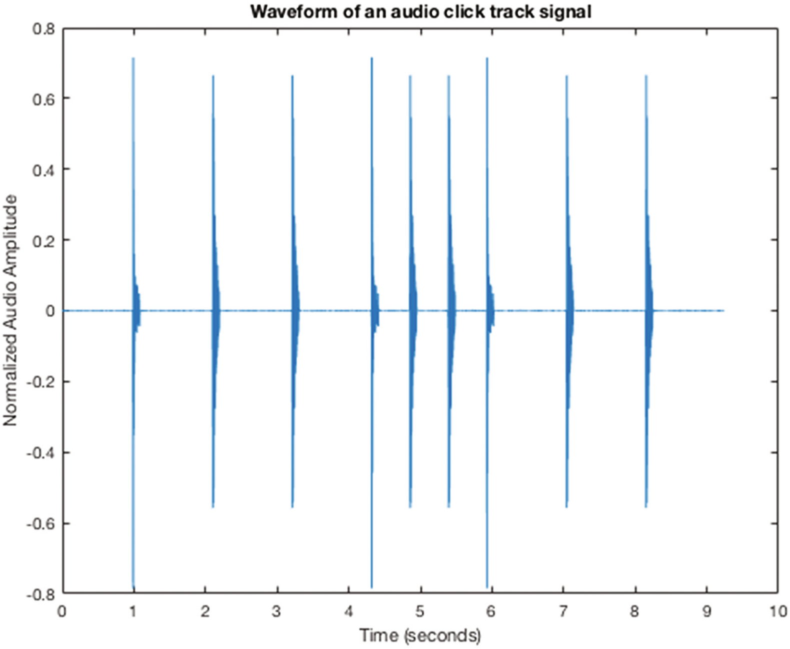


Fig. 2. Waveform sample of an audio click track for the contemporary piece *Charge* by Raphaël Cendo. Note the discrete impulses for each click, the different pulses for the downbeats, and the brief change in tempo.

In order for the haptic pulse lengths to vary with the tempo, the time between clicks, also called the Inter-Onset Interval (IOI), was measured and changed the lengths of the pulses according to the measured IOI. This worked for consistent tempos but when a test was run where the tempo changed frequently, such as what is seen in Fig. 2, it was clear that timing issues were prevalent. This was because the idea that the IOI between clicks was equivalent to the pulse's length was flawed. This is true when the clicks are at equal intervals, but when there is a change in tempo, the IOI before and after a click are different and should have an effect on the pulse length. A more precise way to map the IOI to the length of the haptic pulse is by saying the length of a haptic pulse associated with a click is equal to half the previous IOI (for the rise time) plus half the next IOI (for the decay time). This method of determining the IOI would yield much more precise results.

The final prototype implemented the more precise mapping of IOI to haptic pulse length. However, calculating this in a real-time context was no longer possible without considerable buffering. Instead, a pre-processing script was written in MATLAB that analyzed the audio file containing the click track, extracted the timing information from it, and calculated the necessary pulse lengths and timing requirements. The MATLAB script imports the audio file containing the click track and uses the *findpeaks* function to find the local maxima for each click. From the locations returned by this function, the IOI from the previous click and the IOI for the next click are determined, using the method mentioned above to map the IOI to the length of the haptic pulse. The rise time (half the IOI from the previous click), decay time (half the IOI from the next click), and envelope length (rise time + decay time) for each corresponding haptic pulse is determined from these values; these are key variables in the Vibropixel control message that are used to shape the pulses. To ensure synchrony with the audio click track, the first haptic pulse is triggered midway between the start of the audio click track and the first audio click, so that the peak of the haptic pulse lines up with the first audio click. All subsequent pulses are delayed by the pulse length of the previous, again to ensure peaks line up with the audio clicks. This effectively means that the haptic pulses are triggered in between two audio clicks (see Fig. 3).

This timing information and associated Vibropixel control messages were written to a plain text file that was readable by the *coll* Max object. The *coll* object in the Max environment reads in this text file and ensures that the control messages are sent out to the Vibropixel at the time specified. To determine how well the timing information is displayed by this system, laboratory measurements were performed and are presented in the following section.

The Vibropixels have two motors, a coin cell motor and a pager motor. The prototypes so far were using the coin cell motor only, since it provides a smoother feel than the pager motor. Since the design was only using the coin cell motor, it was decided that the pager motor could be used for distinguishing downbeats. The real-time analysis from earlier prototypes was reintegrated and used to trigger the pager motor when a downbeat is detected, outputting a short 25 ms pulse at the maximum amplitude. This provided a texture to downbeats and ensured that they are easily distinguishable from the other pulses.

The user interview that followed demoed the two haptic pulses for the prototype described above. Professor Bourgogne found the exponential model to be too weak and the lengths were not distinguishable. The triangular model was at the strength necessary but it was more difficult to pinpoint the pulse exactly since it felt rounded at the peaks and not as accurate as the exponential rise and decay model. However, he suggested that if we used the triangular model with a very short pulse from the pager motor (like what was providing the texture on downbeats) on every beat, it may work better. This was quickly set up in the software during the interview and it was demoed again. This changed the prototype to something he said was close to what he imagined this would feel like. It had a curve associated with what was discussed in user interviews, as well as an ictus that allowed a more accurate sense of the tempo.

The MATLAB pre-processing script was modified after the interview to output the pager motor pulses at the locations of the clicks in the audio click track, instead of using the real-time processing, for better timing synchrony. With this prototype, all of the requirements gathered from the user interviews were met except for downbeat distinguishing. It was decided by Professor Bourgogne that this was not as critical as the other requirements. With this, work can begin on integrating the vibrotactile system in a concert setting.

5 Timing Evaluation

Two tests were performed on the prototype implementing the pre-processing script in MATLAB, before the interview with Professor Bourgogne to evaluate the timing of the vibrotactile click track. The first test compared the sound output of the audio click track to the vibration output of the vibrotactile click track. By clamping an analog accelerometer to a Vibropixel, we were able to visualize the vibrations on the device using an oscilloscope. The Max patch output the audio click track and the vibrotactile click track at the same time. Both signals were monitored on two channels of an oscilloscope. Figure 3 shows the measurements gathered from the oscilloscope. This was performed with the triangular model of haptic pulses, without the pager motor on every beat. Note that the amplitude envelope of the vibration, determined with the *envelope* function in MATLAB, is shown rather than the actual accelerometer data for clarity.

The oscilloscope measurements demonstrate that the peak envelope of the vibrotactile pulse is well synchronized to where the click is found in the audio click track output, save for the first pulse, which is later than others, likely due to the motor having to start up from a state of complete rest. The pulse lengths adapt to a tempo change and are smaller for faster tempos. It also demonstrates the rounded peaks that Professor Bourgogne mentioned in the interviews.

The second test measured how much the time between pulses displayed by the Vibropixel deviated from the exact value generated by the MATLAB pre-processing script. This was to give a good sense of how accurate the display length is, and how well the Vibropixel displays timing information. The time delay between triggering two pulses calculated by the MATLAB pre-processing script is the length of a vibrotactile pulse associated with a click in the audio click track (half the previous IOI plus half the next IOI from the audio click track). The error between this measured value and the ideal value is representative of inaccuracies in timing control message output in Max and the wireless transmission latency. The entirety of the vibrotactile click track for Cendo's *Charge* was played on the Vibropixel and the pulse lengths were recorded for each of the 1240 clicks in the piece. This was performed four times to produce four data sets. The data sets were compared to the ideal lengths calculated from the audio click track to determine the error. A box plot with the error of each data set was drawn.

The initial box plot was generated with the measurement error in milliseconds. This box plot demonstrated that the envelope lengths were mostly shorter than what was required. This is likely from inaccuracies in timing in the Max patch and on the Vibropixel itself. Much of the data is within a relatively small range. As well, there was a consistency across each test which shows that the wireless transmission latency is consistent between tests and not random. However, there were a few very large outliers. Looking at the raw data, the values were all at consistent points in the test, during passages with very long pulse lengths due to a rest. Compared to the pulse lengths involved, the error was small, so in order to visualize the data better, the percent error was computed and similar box plots were generated.

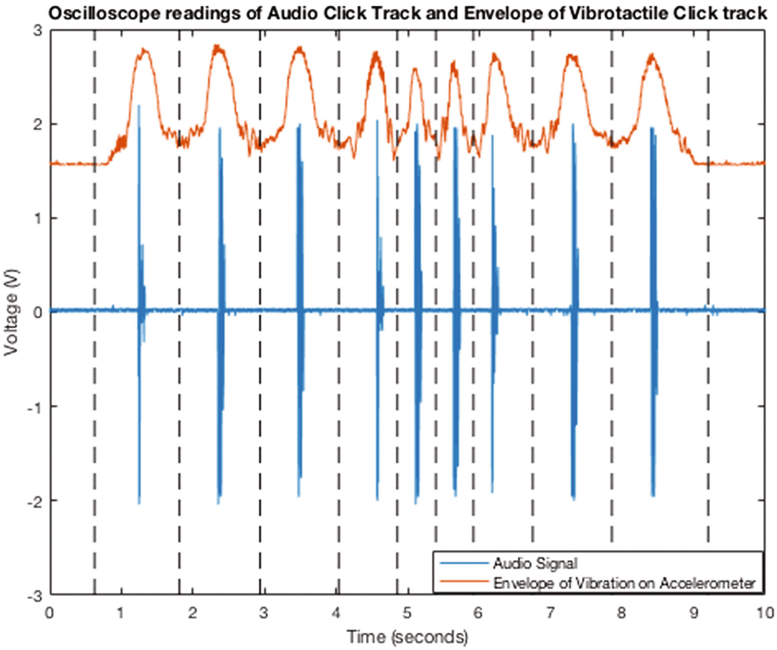


Fig. 3. Oscilloscope measurements showing audio click track with vibrotactile click track. The vibrotactile click track measurements were taken from an accelerometer clamped to the Vibropixel and the amplitude envelope was calculated using the MATLAB *envelope* function. The dotted lines represent where each haptic pulse is triggered. Note that the peaks of the haptic pulse line up with the audio clicks and the rising before the click and falling after the click.

The box plot for percent error in pulse length (seen in Fig. 4) shows that the percentage the pulse length deviates from the ideal is fairly consistent and low. The highest outlier for each data set is that of the first pulse. This is likely because there was some additional delay in starting the completely still actuator. However, the error is still quite low for that pulse, at an average of 4 ms and it is part of an initial count-in (not part of the music piece itself). Overall, the vibrotactile click track was shorter by 5 s of the overall 15 min and 34 s runtime of the piece. Future work on this project will involve compensating for such timing deviations.

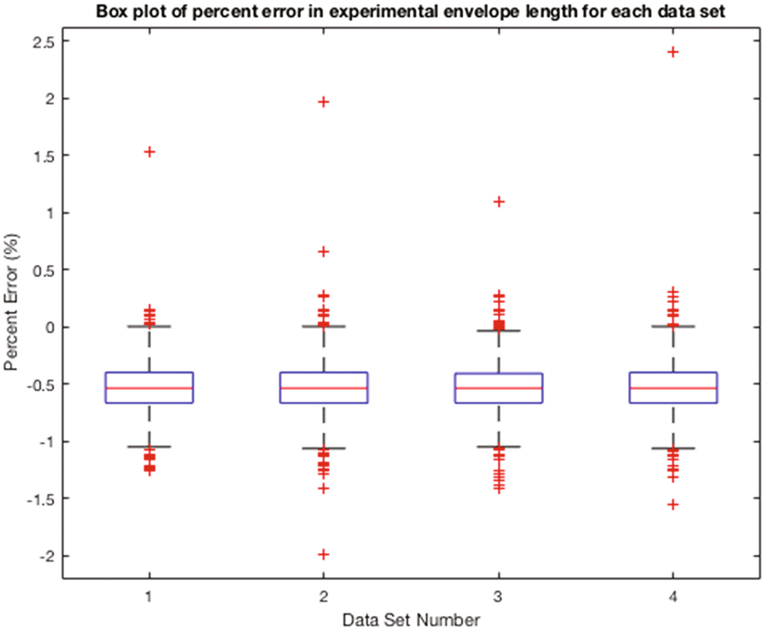


Fig. 4. Box plot showing percent error in pulse length for 4 tests that played the audio click track for the entirety of Cendo's *Charge*. A negative percent error signifies the pulse was longer than the actual value. Note that the variability between tests is minimal and percent error is rather low on average.

6 Conclusion

We presented a flexible device and framework that can be used as a vibrotactile click track for contemporary classical music performance. This platform is unique from other vibrotactile metronome solutions because the system is designed to convert the existing repertoire of contemporary music click tracks to this new vibrotactile system. As well, it uses more natural feeling pulses to convey tempo information to the conductor.

We presented the many requirements a contemporary music conductor has for such a system, gathered through interviews with Professor Guillaume Bourgogne. A device and framework was iteratively designed and implemented using an existing hardware platform based on these requirements. The initial tests and user interviews have been promising in adequately cueing a performer to the tempo using this unique continuous pulse.

Future work will examine integrating the device in a concert setting, and refining the timing information displayed by the Vibropixel. As well, verifying if the system can be used by contemporary music conductors, other than Professor Bourgogne, to aid in their performances is something we wish to explore in the future.

References

1. Choi, S., Kuchenbecker, K.J.: Vibrotactile display: perception, technology, and applications. Proc. IEEE **101**(9), 2093–2104 (2013)
2. El Saddik, A.: The potential of haptics technologies. IEEE Instrum. Meas. Mag. **10**, 10–17 (2007)
3. Birnbaum, D., Wanderley, M.M.: A systematic approach to musical vibrotactile feedback. In: Proceedings of the International Computer Music Conference (ICMC), Copenhagen, Denmark, pp. 397–404 (2007)
4. Giordano, M., Wanderley, M.M.: Follow the tactile metronome: vibrotactile stimulation for tempo synchronization in music performance. In: Proceedings of the SMC Conference, Maynooth, Ireland (2015)
5. Ammirante, P., Patel, A.D., Russo, F.A.: Synchronizing to auditory and tactile metronomes: a test of the auditory-motor enhancement hypothesis. Psychon. Bull. Rev. **23**, 1882–1890 (2016)
6. Soundbrenner: World's First Wearable for Musicians. http://www.soundbrenner.com
7. Hayes, L., Michalakos, C.: Imposing a networked vibrotactile communication system for improvisational suggestion. Organ. Sound **17**, 36–44 (2012)
8. Schumacher, M., Giordano, M., Wanderley, M.M., Ferguson, S.: Vibrotactile notification for live electronics performance: a prototype system. In: Proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR), pp. 516–525 (2013)
9. Frid, E., Giordano, M., Schumacher, M.M., Wanderley, M.M.: Physical and perceptual characterization of a tactile display for a live-electronics notification system. In: Proceedings of the Joint International Computer Music Conference (ICMC) and Sound and Music Computing Conference (SMC), Athens, Greece, pp. 954–961 (2014)
10. Brewster, S., Brown, L.M.: Tactons: structured tactile messages for non-visual information display. In: Proceedings of the Fifth Conference on Australasian User Interface, vol. 28, pp. 15–23. Australian Computer Society, Inc., Darlinghurst, Australia (2004)
11. Hattwick, I., Franco, I., Wanderely, M.M.: The Vibropixels: a scalable wireless tactile display system. In: Proceedings of the Human-Computer Interaction International Conference, Vancouver, Canada (2017)
12. Nielsen, J.: Iterative user-interface design. IEEE Comput. **26**, 32–41 (1993)
13. Puckette, M.S., Apel, T., Zicarelli, D.D.: Real-time audio analysis tools for Pd and MSP. In: Proceedings of the 1998 International Computer Music Conference, pp. 109–112. International Computer Music Association, San Francisco, CA (1998)

Understanding the Key Drivers of Human-Machine Trust: Lessons Learned and Future Directions

A Theoretical Conceptualization for Overtrust

Alan R. Wagner and Mollik Nayyar(✉)

Department of Aerospace Engineering, The Pennsylvania State Institute,
University Park, PA, USA
{alan.r.wagner,mxn244}@psu.edu

Abstract. This paper presents and develops a theoretical basis for understanding overtrust. Overtrust describes a phenomenon in which an individual's calibration of trust does not match the true risk associated with a situation and/or trustee. This paper describes trust in terms of risk and an associated risk equation. The risk equation is then used to decompose overtrust to investigate the factors that influence overtrust. We conclude with a discussion of potential experiments focused on evaluating these theoretical ideas and future work.

Keywords: Human-robot trust · Social robots · Overtrust · Risk · Human factors · Human systems integration

1 Introduction

People tend to overtrust autonomous systems. In 1995, while traveling from Bermuda to Boston, the Royal Majesty cruise ship ran aground because the ship had been left on autopilot for 34 h and the autopilot malfunctioned [1]. Studies have found that overtrust in automated systems for detecting cancer result in certain types of cancers being overlooked [2]. Our own research has shown that during an emergency evacuation, some individuals will follow and stand by a robot that has stopped moving, moves in circles, or repeatedly crashes into a wall in spite of the risk to their well-being [3]. As robots leave the lab and enter battlefields, schools, and hospitals, these examples show that people may become over-reliant on and overtrusting of these systems.

For people working with automation, overtrust is an important consideration; for people working autonomous social robots, overtrust may have far-reaching consequences. Driverless vehicles and autopilot technologies are now available to the public and have already resulted in deaths [4]. Autonomous aerial vehicles are poised to become taxis services and autonomous weapons have been deployed to actively participate in wars [5]. These technologies and others suggest that are on the cusp of new age of autonomous machine usage. It is thus critical that we investigate the issues surrounding human overtrust of autonomous systems.

This paper investigates overtrust of autonomous robots, which, unlike automation, tend to socially engage people and promote anthropomorphism. Unlike factory automation, anthropomorphic social robots may increase a person's tendency to overtrust, and are thus an important avenue for further study. We present a conceptualization of overtrust. Overtrust is defined as a person's tendency to trust a person or autonomous

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_25

system too much [6]. Relatedly automation bias is a person's tendency to view a machine's suggestions favorably, even when contradicted by other information [7]. The purpose of this paper is to begin to develop the theoretical underpinnings related to overtrust. Successfully doing so may inform the development of experimental studies or develop a crisper distinction between overtrust and human-robot trust.

The remainder of the paper begins by discussing the related work in this area. Next, as a conceptual foundation, we present our theoretical formulation of trust. Overtrust is then described as a special case during which a person's overestimates trust. We then suggest ideas for further empirical study and conclusions.

2 Related Work

A reasonable amount of research has focused on the effect of overtrust on automation. Overtrust can result when individuals are asked to attend to several tasks and monitor highly reliable automation [8]. A key component related to overtrust of automation is the automated system's reliability [9]. Highly reliable systems tend to encourage less vigilant monitoring of automated systems. User complacency occurs when an automated system operates as expected. In fact, decreasing the reliability of the automated system increase the detection rate [10]. The research implies that one way to manage overtrust is for the machine to fail more often. While this may be possible for automated systems, for autonomous robots such as driverless vehicles occasional system failures to increase user attentiveness are not an option.

Not surprisingly, human-robot interaction researchers are also finding that people tend to overtrust robots [11]. Salem et al. demonstrated that human subjects that are asked to perform intentionally odd behaviors, such as watering a plant with orange juice, mostly complied [12]. Robinette et al. demonstrated that 27 of 29 human subjects faced with evacuating a realistic emergency environment followed the robot's guidance even though the robot failed in a recent prior guidance task [13]. Several of these subjects followed the robot's guidance even when that guidance directed the person to a dark, unlit room blocked by a couch.

Autonomous system related overtrust research is still in its infancy. The results primarily indicate that overtrust is an issue that the community needs to consider and, to the extent possible, develop methods for managing.

3 A Theoretical Treatment of Trust

Our theoretical treatment of trust originates from our desire to develop formal techniques that would allow a robot to reason about when and whether a person is trusting it or it should trust a person. This avenue of research required a precise and grounded definition of trust. Drawing from Mayer's work on interpersonal trust [14] and Lee and See's work on trust in automation [6] we developed an operational definition of human-robot trust that is formulated within a game theory context. We define trust as, "***a belief, held by the trustor, that the trustee will act in a manner that mitigates the trustor's risk in a situation in which the trustor has put its outcomes at risk***" [15]. Our definition highlights the role of three important factors that influence trust: the trustee, the trustor, and the situation (Fig. 1). Consider, for example, the trust fall. The trust fall is a type of game in which the trustor leans backward and the trustee catches the trustor. With respect to the definition above, the trustor decides to lean back if they believe that the trustee will mitigate their risk by catching them. The trustor's decision to lean back is influenced by characteristics related to the trustee. If the trustee appears weak or incapable of catching the trustor, the perceived risk of leaning back increases. Similarly, if the trustee seems unlikely, unmotivated, or unwilling to catch the trustor, the perceived risk of leaning back also increases. On the other hand, if the trustee appears strong and motivated to catch the trustor, then the perceived risk decreases. Characteristics of the trustor are similarly important. If the trustor has been dropped while recently playing the game, then the perceived risk may be greater. If the trustor has back problems, the perceived risk is greater. Finally, characteristics of the situation also determine one's estimation of risk. If the person is leaning over a soft bed of grass then the perceived risk is less than if one is leaning over a bed of rusty nails or broken glass.

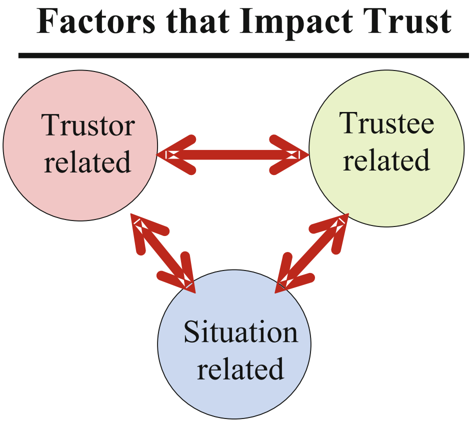


Fig. 1. Three interrelated factors influence trust decisions in our framework.

During the process of the making a trust-based decision, these factors come together to produce a decision. It is not the true numerical valuation that determines this decision, but rather, the trustor's perception of that risk and their evaluation of risk from an egocentric perspective. The saliency of the factors is important and extremely difficult methodologically to control. Individuals may not recognize all of the risks involved with making a decision. Even more likely, individuals may not recognize that they have alternatives available. The lack of recognition of alternatives is an important confounding issue related to trust. Often times a third party observer may evaluate a trustor's decision as an indication of trust, whereas, in reality, the decision was made because the trustor felt that they had no other option. Our view of trust has been that the trustor must recognize that they have an alternative, but shades of gray tend to emerge. For instance, choosing to lean back may feel like the only option, in spite of one's consideration of risk, because the social pressure to do so is great.

We utilize a game theoretic approach to model human autonomous system interaction. For this approach, we apply our definition of trust to game theoretic representations. A social situation demanding trust can be modeled as an extended-form game between one individual named the trustor and another named the trustee (Fig. 2). In an extended-form game the trustor will act before the trustee. The trustor is thus ensured of placing themselves at risk with the expectation that the trustee will mitigate this risk. As depicted in Fig. 2, the trustor decides between action a_i^{tr} and a_j^{tr}. Action a_i^{tr} is described as the *trusting action* because this action places the trustor at risk. Action a_j^{tr} is characterized as the *no-trust action* because this action does not place the trustor at risk. For example, for the trust fall game, the trustor chooses between leaning backward (trusting action) and not leaning back (no-trust action). Similarly, the trustee chooses between actions, a_i^{te} and a_j^{te}. Action a_i^{te} is described as the *maintain trust action* because selecting this action will mitigate the trustor's risk increasing the likelihood of future acceptance of risk by the trustor. Action a_j^{te} is described as the *violate trust action* because selecting this action violates the trustor's belief that the trustee will mitigate their risk.

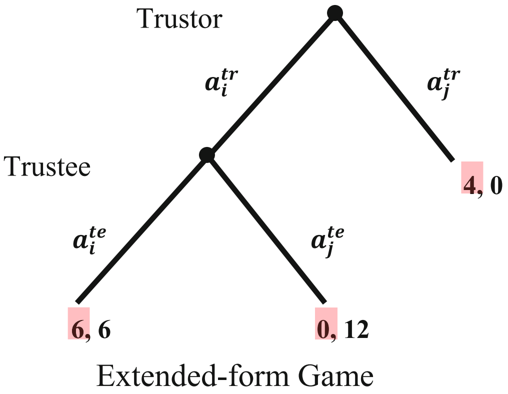


Fig. 2. An extended-form game is depicted representing a trust situation. The highlighted values indicate the reward/cost for the trustor.

If the trustor selects action a_i^{tr} then they risk some outcome or reward equal to $\left\{a_i^{tr}, a_i^{te}\right\} - \left\{a_i^{tr}, a_j^{te}\right\} = o_{i,i}^{tr} - o_{i,j}^{tr}$, where $o_{i,i}^{tr}$ is the outcome for the trustor when the trustor selects the trusting action and the trustee *maintains trust* and $o_{i,j}^{tr}$ is the outcome for the trustor when the trustor selects the trusting action and the trustee *violates trust*. This is the risk associated with leaning back. Once the decision to lean back is made, their fate rests in the hands of the trustee. An established formula for calculating risk is

$$R(x, y) = \sum L(x, y)p(y), \tag{1}$$

where $L(x, y)$ is the loss associated with choosing x when the true value is y and $p(y)$ is the probability of event y occurring [16]. In social decision theory, risk can be similarly calculated as $R\left(\{a_i^{tr}, a_i^{te}\}, \{a_i^{tr}, a_j^{te}\}\right) = \sum L\left(\{a_i^{tr}, a_i^{te}\}, \{a_i^{tr}, a_j^{te}\}\right)p\left(a_j^{te}\right)$ ere the function $L\left(\{a_i^{tr}, a_i^{te}\}, \{a_i^{tr}, a_j^{te}\}\right)$ denotes the loss by choosing $\{a_i^{tr}, a_i^{te}\}$ over $\{a_i^{tr}, a_j^{te}\}$ and $p\left(a_j^{te}\right)$ is the probability that the trustee will choose a_j^{te}. The result is a value $R\left(\{a_i^{tr}, a_i^{te}\}, \{a_i^{tr}, a_j^{te}\}\right) \in \Re$ which can then be compared to the trustor's measure of risk-aversion, θ, a variable representing the trustor's risk-aversion for this type of risk at this moment.

The components of this formulation directly map to the factors discussed in Fig. 1. If the trustor and the trustee select actions simultaneously, then the action selection probability should be represented as the joint probability distribution, $p\left(a_x^{tr}, a_y^{te}\right)$ where the trustor knows whether a_x^{tr} is a_i^{tr} or a_j^{tr}. If the trustor and trustee select actions sequentially then the action selection probability should be represented as a conditional probability, $p\left(a_y^{te} | a_i^{tr}\right)$. In both cases, the action selection probability reflects the trustor's model of the trustee. This model might be highly uncertain, reflecting the trustor's lack of experience with the person or it may be highly certain, a reflection of a long history of interactions with the person and the ability to predict their behavior. For the trust fall, the action selection probability may be based on an estimate of the person's age (too young or old to catch the person) or perceived motivation (laughing or lack of attention). The term $L\left(\{a_i^{tr}, a_i^{te}\}, \{a_i^{tr}, a_j^{te}\}\right)$ reflects the potential loss in a **situation** or during an interaction. There are many different types of losses that can occur. Some examples are financial loss, physical loss in the form of injury, and emotional loss. Multiple forms of loss may also occur during the same interaction. The value θ, is a risk-aversion variable which can be conditioned on a number of different factors, such as previous experiences in similar situations, related to the characteristics and perhaps personality of the trustor.

4 Conceptualizing Overtrust

With respect to automation, Lee and See describe overtrust as, "poor calibration in which trust exceeds system capabilities." [6] As noted by Lee and See, trust in automation differs from interpersonal trust in that automation lacks intentionality, an important facet for trust. Unlike automation, when interacting with an autonomous robot the human tendency to anthropomorphize generates an illusion of intentionality [17]. Hence, overtrust of an autonomous system may relate more closely to interpersonal overtrust than trust in automation.

Proper calibration is a situation in which the individual's estimation of risk, $\acute{R}(x, y)$, is a reasonably accurate reflection of the true underlying risk, $R^*(x, y)$. Formally, is a situation in which an individual makes decisions based a reasonably accurate assessment of risk, $\acute{R}(x, y) \approx R^*(x, y)$. By *reasonably accurate* we mean to convey not that the person has full knowledge of every potential risk, but rather that they are relying on an accurate approximation of the risk associated with a trusting action. From this, a formal definition of overtrust follows as a situation in which the trustor inaccurately believes that the actual underlying risk is less than the true risk, $\acute{R}(x, y) < R^{*}(x, y)$.

Equation (1) suggests that overtrust may occur because (1) the trustor underestimates the loss associated with an action; (2) the trustor underestimates the probability of a trust violation, $p(y)$, occurring or; (3) both. Formally,

$$\begin{aligned} R^{*}(x, y) &> \acute{R}(x, y) \\ R^{*}(x, y) &> \sum \acute{L}(x, y)\acute{p}(y). \end{aligned}$$

We conjecture that these sources of overtrust are qualitatively different. In one case, the trustor underestimates the loss associated with a particular action, $L^{*}(x, y) > \acute{L}(x, y)$. A mischaracterizaon of loss occurs when a person underestimates the physical, emotional, financial loss that will occur in some trust situation. Lack of familiarity with the situation tends to cause loss mischaracterization. Intuitively, the trustor underestimates the loss, $L(x, y)$, associated with the situation related factors. Using the trust fall example, in this case, the trustor evaluates the risk of a fall as less that the true risk because they evaluate the injury resulting from being dropped as less than likely injury.

$$L^{*}(x, y) > \acute{L}(x, y). \tag{2}$$

In the second case, the trustor underestimates the probability of the untrusting action occurring, $p^{*}(y) > \acute{p}(y)$. In this case the trustor accepts too much risk, believing that the robot can perform actions which it cannot, that the robot has knowledge which it does not, that the robot is motivated when it is not, or that the robot's recent prior performance is an indicator of it future performance. With respect to the trust fall the trustor evaluates the likelihood of the trustee dropping the person as smaller than the true likelihood. Intuitively, the source of the trustor's inaccurate risk assessment in this case originates with evaluations of the trustee. For our formulation, the term $p(y)$. is meant to encompass all of the factors associated with misjudging the trustee.

4.1 Importance of Overtrust Evaluations

There is value in attempting to deconstruct overtrust into potential sources of confusion. On one hand, understanding the source of overtrust may allow a robot to better direct its trust recalibration efforts. For instance, if the autonomous system can determine that the trustor is misestimating the ability of the trustee, then it can direct its recalibration effort by reminding the trustor about the trustee's lack of perception, lack of motivation, etc. Alternatively, the autonomous system may need to focus on recalibrating the trustee's estimation of the loss that will occur. Finally, in some situations, the system may need to do both.

In general, the use of stereotypes may contribute to the misjudgment of the trustee [18]. Our work has shown that stereotypes can lead to inaccurate predictions of behavior. Depictions of autonomous systems in the media are likely to have influenced a person's perception of the system by creating unrealistic expectations about how a robot does or should work. Our research has found that human subjects describe autonomous systems as infallible or note that the programmer's intentions related to the system are benevolent and ultimately trustworthy [13]. Hence, moving forward, it will be important to accurately calibrate the person's trust in lieu of the person's preconceived notions of the robot's ability and behavior.

Stereotypical social situations also occur. Certain environments, such as homes, schools, driving one's car, foster a sense a safety. Our work has shown that these types of situations can be cluster into learned groups that the autonomous system can use for prediction purposes [19]. With respect to overtrust, stereotypical situations may encourage people to underestimate the loss that can occur.

As depicted in Fig. 1, aspects of the trustor also influence trust decisions. For instance, the trustor's risk aversion, Θ, varies with respect to that person's history, personality, mood, etc. Yet these factors **are not impacted** by the risk estimate. Hence, a risk-seeking person has an accurate estimation of the risk associated with a particular course of action and rationally chooses to select the trusting action in spite of the high risk. For example, a stunt performer typically does not overtrust their partner during stunt. Rather, they understand the risks and simply choose to accept them. We do not consider these situations to be overtrust because the trustor has accurately estimated the risk of a decision.

5 Empirically Investigating Overtrust

Real-world situations are dynamic and the ideal way to evaluate trust in the real-world is by obtaining complete information of all possible situations. Since this is not possible, it is possible that the trust a person places in a system or an entity may actually be overtrust as the trustor is unaware of the potential modes of failure. This could be in the form of placing their trust in a system that is considered infallible or when certain information related to the environment/entity is overlooked. Both cases could lead to a misevaluation of the potential risk in the trusting action being considered. This dynamic nature of the environment makes evaluation of overtrust difficult.

Games, such as the one in Fig. 2, provide a potential method to investigate overtrust. Due to the ease in computational modeling of such games, these could potentially help investigate the factors relating to overtrust and provide insight into such decisions. The Investment Game, for example, is commonly used to examine trust [20]. During the investment game, the trustor (investor) selects some amount of money to give to a trustee. The money given to trustee appreciates or depreciates according to the market (multiplied by some number between 0–5). The trustee then selects some amount of money to return to the investor.

In this scenario, overtrust occurs when the investor underestimates the risk associated with an investment of some amount of money. The investor may underestimate the loss generated by the market or the probability that the trustee will return the money. By varying these two factors individually it may be possible tease apart the factors we hypothesize independently control overtrust.

Multiple variations of the investment game are possible by changing the game dynamics. While such games can potentially provide a quantitative evaluation of overtrust in terms of risk, augmenting these games with surveys can provide an additional insight into the decision-making process of the trustor.

6 Conclusions

Trust is an important component in human society as it influences how people interact with each other. With numerous examples of human overtrust of robotic and autonomous systems resulting in catastrophic failures, there is a pressing need for understanding trust and how and what drives the mechanisms resulting in overtrust in humans. Game theory and psychology have provided a definition of trust that has given us a way to effectively model, study and predict trust in social interactions. The three factors that impact trust have been identified as trustor related factors, trustee related factors and situation related factors. We have shown that these models inform research directions associated with trust-based decisions.

We have shown that an equation representing the risk associated with a trust-based decision can be used to examine the factors responsible for overtrust. We suggest that the sources of overtrust can be divided into two general categories, situation-based factors related to the loss and trustee-based factors related to the probability of a trust violation. We have suggested an experimental paradigm that could be used to tease these factors apart. The paradigm is a modification of the well-known investment game commonly used in by the behavioral economics community. Future work will focus on using this game to confirm if these factors are in fact independent and, if so, developing methods to prevent overtrust.

References

1. Charette, R.N.: Automated to Death, December 2009. http://spectrum.ieee.org/computing/software/automated-to-death. Accessed 17 Dec 2014
2. Povyakalo, A.A., Alberdi, E., Strigini, L., Ayton, P.: How to discriminate between computer-aided and computer-hindered decisions a case study in mammography. Med. Decis. Making **33**(1), 98–107 (2013)
3. Robinette, P., Wagner, A.R., Howard, A.: The Effect of Robot Performance on Human-Robot Trust in Time-Critical Situation. Technical report: GT-IRIM-HumAns-2015-001, 2015. http://hdl.handle.net/1853/52899
4. Deamer, K.: What the first driverless car fatality means for self-driving tech. Sci. Am., 1 July 2016. http://www.livescience.com/55273-first-self-driving-car-fatality.html. Accessed 10 Nov 2016
5. Czyzewski, A.: Personal flying vehicles project aims to end road congestion, 22 June 2011. https://www.theengineer.co.uk/issues/4-july-2011/personal-flying-vehicles-project-aims-to-end-road-congestion/. Accessed 26 Jan 2017
6. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**, 50–80 (2004)
7. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: an attentional integration. Hum. Factors J. Hum. Factors Ergon. Soc. **52**(3), 381–410 (2010)
8. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consquences of automation-induced "complacency". Int. J. Aviat. Psychol. **3**(1), 1–23 (1993)

9. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Factors **39**(2), 230–253 (1997)
10. Parasuraman, R., Mouloua, M., Molloy, R.: Effects of adaptive task allocation on monitoring of automated systems. J. Hum. Factors Ergon. Soc. **38**(4), 665–679 (1996)
11. Booth, S., Tompkins, J., Pfister, H., Waldo, J., Gajos, K., Nagpal, R.: Piggybacking robots: human-robot overtrust in university dormitory security. In: International Conference on Human-Robot Interaction (2017)
12. Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K.: Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI) (2015)
13. Robinette, P., Allen, R., Li, W., Howard, A., Wagner, A.R.: Overtrust of robots in emergency evacuation scenarios. In: ACM/IEEE International Conference on Human-Robot Interaction (HRI 2016), Christchurch, New Zealand (2016)
14. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Acad. Manag. Rev. **20**(3), 709–734 (1995)
15. Wagner, A.R.: The role of trust and relationships in human-robot social interaction. Ph.D. Diss., School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA (2009)
16. Holton, G.A.: Defining risk. Financ. Anal. J. **60**(6), 19–25 (2004)
17. Heider, F., Simmel, M.: An experimental study of apparent behavior. Am. J. Psychol. **57**, 243–259 (1944)
18. Wagner, A.R.: Robots that stereotype: creating and using categories of people for human-robot interaction. J. Hum. Robot Interact. **4**(2), 97–124 (2015)
19. Doshi, J., Kira, Z., Wagner, A.R.: From deep learning to episodic memories: creating categories of visual experiences. In: Third Annual Conference on Advances in Cognitive Systems ACS (2015)
20. King-Casas, B., Tomlin, D., Anen, C., Camerer, C., Quartz, S., Montague, P.: Getting to know you: reputation and trust in a two-person economic exchange. Science **308**(5718), 78–83 (2005)

The Sky's (Not) the Limit - Influence of Expertise and Privacy Disposition on the Use of Multicopters

Chantal Lidynia(✉), Ralf Philipsen, and Martina Ziefle

Human-Computer Interaction Center (HCIC), RWTH Aachen University,
Campus Boulevard 57, 52074 Aachen, Germany
{lidynia,philipsen,ziefle}@comm.rwth-aachen.de

Abstract. There are a variety of civil usage contexts for "drones" or multicopters, unmanned aerial vehicles or remotely piloted aircraft systems. Probably best known is the so-called delivery drone that many companies are testing and hoping to deploy. But the acceptance of said technology, especially those equipped with cameras, by the public is rarely investigated. Therefore, an empirical survey (N = 228) was conducted in Germany to determine the influence of expertise with aviation as well as privacy disposition on the perception of multicopters. It was found that a regular pilot's license did not impact the attitude towards drones from people with no experience with aviation, drones or otherwise. Concerning drones flying over the own home, not even active drone users would condone a stranger's multicopter to cross over their property unless operated by rescue services. Here the important factor is the need for privacy and perceived risk of it being violated.

Keywords: Drones · Human factors · Privacy · Piloting experience · Technology acceptance · Policy

1 Introduction

Unmanned Aerial Vehicles (UAVs) or multicopters are no longer only of interest for military operations. They offer a multitude of options for civil deployment, such as wildlife monitoring, advertisement, security, or delivery services, to name but a few (cf. [1]). Consequently, the civil usage has gained more and more interest and is the focus of an ever-growing field of research. For the most part, said research can be divided into two main venues. First, the technological feasibility. In this context, the aspects of getting the airborne devices to perform the intended tasks are the main focus, be it to have them carry a specific load, fly on a predetermined route, or detect certain objects, be it wildlife, people, cars, or even structural integrity (e.g., [2–4]). The other main line of research is on the judicial side (e.g., [5–7]). The deployment of multicopters is linked to the adherence to laws and regulations, both of which are, as of yet, in constant flux as the needs of all parties involved are still not clearly defined. What is mostly missing from research so far, is the social aspect with technology acceptance and public perception as the key domains [8]: Although multicopters, commonly referred to as drones, are unmanned, humans are involved in almost every aspect of their deployment. However, it is not only as ground pilots or developers. The largest group exposed to

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_26

drones is the general public, meaning those who are passively affected by this new technology by being flown over or being filmed without their knowledge or consent, for example. Therefore, their opinion and attitude towards this technology is of vital interest to a successful integration of multicopters into civil usage contexts. Nevertheless, research into this human side of civil drone deployment is only slowly gaining track, e.g., [8, 9].

2 Related Work

In a previous study, [8] could show that human factors are a driving force behind the acceptance of drones in civil usage contexts. They examined requirements for civil drones in different usage scenarios and surveyed demands in regard to appearance, routing, and autonomy of the drones. It was found that - especially laypeople -were concerned about the huge potential for privacy losses when using drones in contrast to experts (active drone pilots) who were more concerned about the risks of accidents and damage. Thus, for the general public, the biggest barrier of drone usage lies in the possible violation of privacy.

However, privacy is a rather vague concept that lacks a universal definition (cf. [10–12]) and strongly depends on the respective context. Therefore, it is unsurprising that with every new technology available, the preexisting concept of privacy might or needs to undergo changes. This also holds true for laws and regulations meant to protect the individual's privacy. For more details, see, for example, [6, 7, 13–15]. Other researchers also pointed out the importance of privacy in civil drone deployment (e.g., [16, 17]). Although legislations vary considerably between countries and continents, multicopters and their implications for privacy breaches are also of interest in Europe [6, 18]. A rather detailed overview of issues and privacy concerns is given in [19]. Luppicini and So [20] conducted a qualitative research into ethical and social aspects of commercial drone use, the probably best known instance of civil drone deployment, especially since delivery companies are currently researching the potential of including drones into their fleet (e.g., DHL, Amazon, UPS). For a theoretical insight into the potential of delivery drones, see, e.g., [21]. Pauner, Kamara and Viguri offer a comprehensive overview over privacy protection, regulations, and drone use with a focus on the European Union [18].

Another big sector of drone deployment is that of surveillance and safety. Police forces are using drones for crowd surveillance during demonstrations, to oversee traffic, or to patrol borders. The technology they use includes cameras and therefore also provides visual data that might or might not violate individual's privacy. This was also pointed out in the study conducted by Wang et al. [16]. Vattapparamban et al. [22] give not only a comprehensive summary on privacy issues with drone deployment, but they also introduce a number of commonly available multicopters that might be encountered more frequently in the future. Furthermore, they also detail different ways that drones can be misused, thereby addressing the possible safety issues when dealing with drones. Lidynia et al. [8] also found fear of damage done by multicopter malfunctions a barrier to acceptance. This is congruent with other studies, e.g., [20, 22]. In [23], the authors examined incident and accident reports from around the world involving civil remotely piloted aircraft systems (RPAS). Combining human factors and safety issues, [24] studied the impact of knowledge of rules and regulation on the avoidance of (near) mid-air collisions. Furthermore, they also included the factor of expertise in their study. In their earlier study, [8] discovered significant differences in the evaluation of multicopters based on prior experience with drones.

Based on open questions from this previous research and the development of current issues with a main focus on privacy, this paper aims at offering more insight into the impact of public perception of drones, taking a human factors perspective and exploring the impact of expertise with drones, the level of individual technology self-efficacy, as well as users' privacy "disposition" on the acceptance and evaluation of civil drone use.

3 Methodology

An online questionnaire was distributed via social media and postings in different multicopter and pilot related online forums. Included were items about general evaluation of drones, occupational or private use of them, the possession of an aircraft pilot license. To evaluate technical self-efficacy (SET), items adopted from [25] were included. Furthermore, statements from different privacy scales were added to examine the participants' need for privacy. The reliability of both scales has been calculated with sufficient Cronbach α scores ($\alpha_{SET} = .870$, *n* = *226*, *4 items;* $\alpha_{PRIVACY} = .772$, *n* = *220*, *9 items*).

4 Sample

A total of N = 228 participants completed the questionnaire. The sample included 37 women (16.2%) and 190 men (83.3%). Their age ranged from 14 to 71 years, with an average age of 38.77 years (SD = 13.53). 29.8% (n = 68) of the participants indicated owning a drone, 69 (30.3%) had used one at least once and 10 (4.4%) used multicopters in their occupation. According to users' statements, four different expertise groups were formed: 39.5% (n = 90) have neither drone experience nor own an aircraft piloting license, 29.8% (n = 68) use drones but have no pilot's license, 18.9% (n = 43) possess said license but do not use drones, and 11% (n = 15) fly both drones and regular aircrafts.

The technical self-efficacy, assessed by a 6-point Likert scale (with 0 = very low, and 5 = very high), was found to be rather high with a mean of 4.07 (out of 5 points max.) (SD = .93). The reported need for privacy, also measured with a 6-point Likert scale (0 = very low privacy need, 5 = very high privacy need), reached an average of 3.67 (SD = .74).

5 Results

Data was analyzed using descriptive as well as inferential statistics (ANOVAs, t-tests). To determine correlations between variables, the non-parametric Spearman-Brown rank analysis was used. For all analyses, the level of significance was set at 5% level.

The following chapter is structured as follows: First, the general attitudes towards drones will be presented. Second, the valuations of current and potential future legal regulations will be shown. Last, concerns regarding a general usage, pilot factors and the specific possibility of overflights over own real estate property will be considered.

5.1 General Attitude Towards Drones

Looking at the complete sample (N = 228), the attitudes towards drones was rather positive. 60.5% of the sample stated that they had a positive attitude, whereas 39.5% had a negative view of this technology. Nevertheless, the emerging picture becomes more diverse by taking the different expertise groups into account. As can be seen in Fig. 1, the groups that do not use drones are of two minds whether to rate drones positive or negative. Both the non-pilots and the aircraft pilots that do not use drones show a nearly balanced attitude distribution with a slight tendency to a negative view.

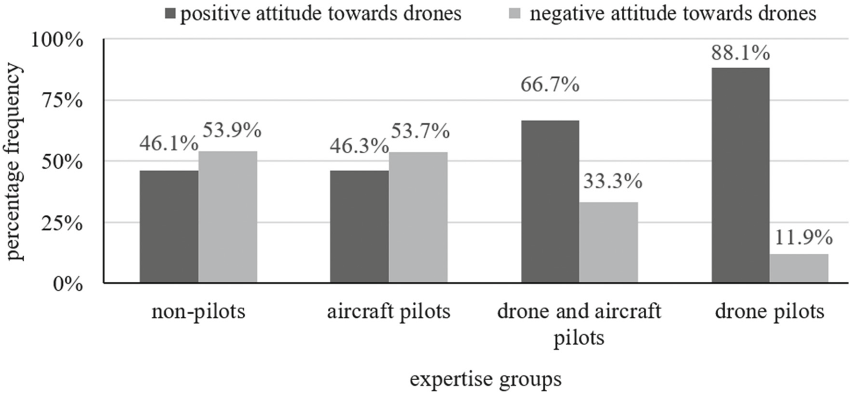


Fig. 1. Percentage frequencies of attitude towards drones in different expertise groups

In contrast, a two-thirds majority of the drone-using pilots had a positive opinion about multicopters. The consent to the technology becomes even clearer when looking at the drone users without a piloting license: Almost 90% reported to have a positive mindset. A closer look into the data revealed that the remaining negative attitudes were predominantly stated by participants who were professional drone users (11.9%)

Furthermore, the participants with opposing attitudes towards drones differed significantly regarding their technical self-efficacy and their privacy disposition: On average, proponents had a higher technical self-efficacy ($M_{SET} = 4.2$, $SD_{SET} = 0.8$) and a lower need for privacy ($M_{PRIVACY} = 3.6$, $SD_{PRIVACY} = 0.7$) than participants with a negative predisposition ($M_{SET} = 3.9$, $SD_{SET} = 1.0$, $M_{PRIVACY} = 3.8$, $SD_{PRIVACY} = 0.7$), as taken from $t_{SET}(221) = 2.099$, $p_{SET} = .037$, $d_{SET} = 0.288$ and $t_{PRIVACY}(221) = -2.574$, $p_{PRIVACY} = .011$, $d_{PRIVACY} = -0.353$.

5.2 Policy Requirements

Looking at policy requirements, first the evaluation of the present legal situation in Germany will be presented. Second, regulations that are currently under public discussion will be showcased.

Current Policy Requirements. Looking at the current legal situation, it becomes clear that none of the current legal restrictions for private drone usage is perceived as overly restrictive by any of the expertise groups (see Fig. 2). Only the limitation to operate drones in sight was rated ambiguously by non-users and drone pilots without an aircraft pilot license.

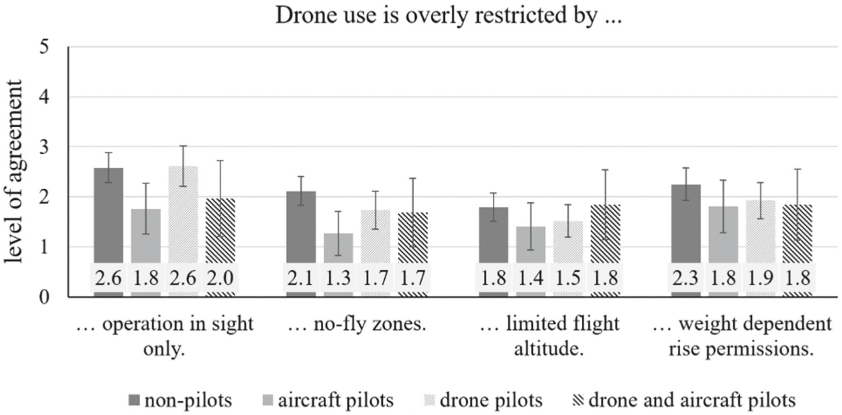


Fig. 2. Average agreement to current legal regulations (min = 0, max = 5) and related 95.0% confidence intervals (CI) depending on expertise groups

While there were no significant main effects of expertise for the restriction regarding flight altitude and rise permissions, expertise significantly influenced the evaluation of operation in sight (F(3,218) = 3.662, p = .013) and no-fly zones (F(3,218) = 3.364, p = . 020). Pairwise comparisons revealed there was a significant difference between aircraft pilots and participants without pilot license regarding the limitation of operation in sight only (p < .05 for all comparisons). Pilots regarded this legal regulation as over-restriction to a minor extent in comparison to non-pilots. Concerning no-fly zones, only pilots who do not use drones themselves differ from the other expertise groups with the lowest approval rate of no-fly zones.

Looking at the impact of participants' technical self-efficacy and their personal need for privacy on the evaluations, it was revealed that the technical self-efficacy negatively correlated with the agreement to no-fly zones (r = −.181, p = .006) and limited flight altitude (r = −.140, p = .037) as too restrictive. Similarly, the reported need for privacy correlated negatively with the restrictions concerning no-fly zones (r = −.190, p = .004) and flight in sight (r = −.185, p = .005). However, all correlations were of small effect size.

Future Policy Requirements. Considering potential future policy requirements, it became clear that all expertise groups agreed to seamless legal specifications for usage (see Fig. 3). However, the agreement was even stronger for participants without drone usage experience. These participants also stated that there should be a compulsory registration for all drones and a mandatory license for all drone pilots. In contrast, the requirements of both drone user groups were rather inconclusive: The average agreement towards drone registrations and drone pilot licenses was almost the same as the mathematical neutral point of the scale used ($M_{NEUTRAL} = 2.5$).

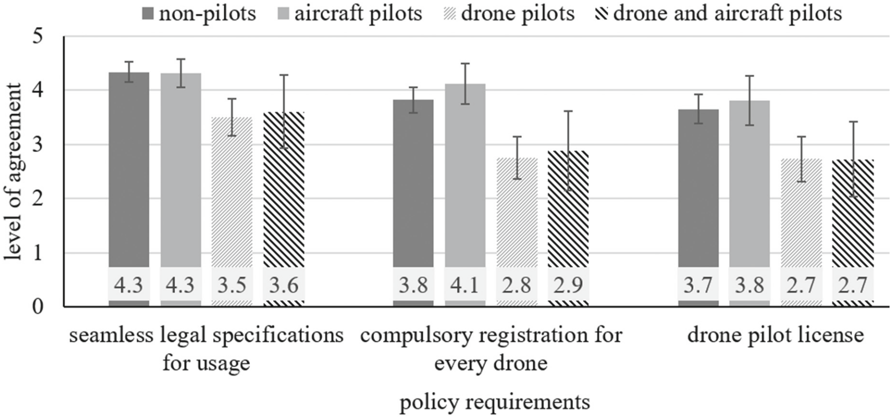


Fig. 3. Average agreement to possible legal requirements (min = 0, max = 5) and related 95.0% CI depending on expertise groups

Pairwise comparisons confirmed the picture emerging from Fig. 3. Whereas the differences between users and non-users of drones were significant for all policy requirements ($p < .05$ for all comparisons), the possession of an aircraft pilot license led to no significant difference.

Furthermore, the participants' technical self-efficacy was slightly negatively correlated with the request for legal specifications for drone usage ($r = -.200$, $p = .003$) and licenses for drone users ($r = -.159$, $p = .017$). The personal need for privacy had no influence on participants' statements about policy requirements.

5.3 Concerns Raised Toward Drone Usage

In the following, a closer look at tribulations of drone usage will be presented. First, general concerns will be outlined. Second, worries about drone pilots' behavior will be introduced.

General Concerns. High approval rates were found for most of the queried concerns (see Fig. 4). All expertise groups reported high concerns about a potential use of drones as weapons, the misuse by criminals and potential espionage application. Likewise, all groups agreed that drones can cause bodily harm. In contrast, the fear that drones would violate privacy was shared only by non-pilots and aircraft pilots, whereas both drone user groups disclosed a more neutral position. Although almost all mentioned concerns got high agreement levels, the no expertise group stated that drones would scare them. Significant main effects depending on the expertise groups were found for the fear that drones can be used by criminals (F(3,220) = 2.943, p = .034), can be used to spy on people (F(3,220) = 10.192, p < .001), violate the privacy (F(3,220) = 11.043, p < .001), and scare in general (F(3,219) = 6.954, p < .001).

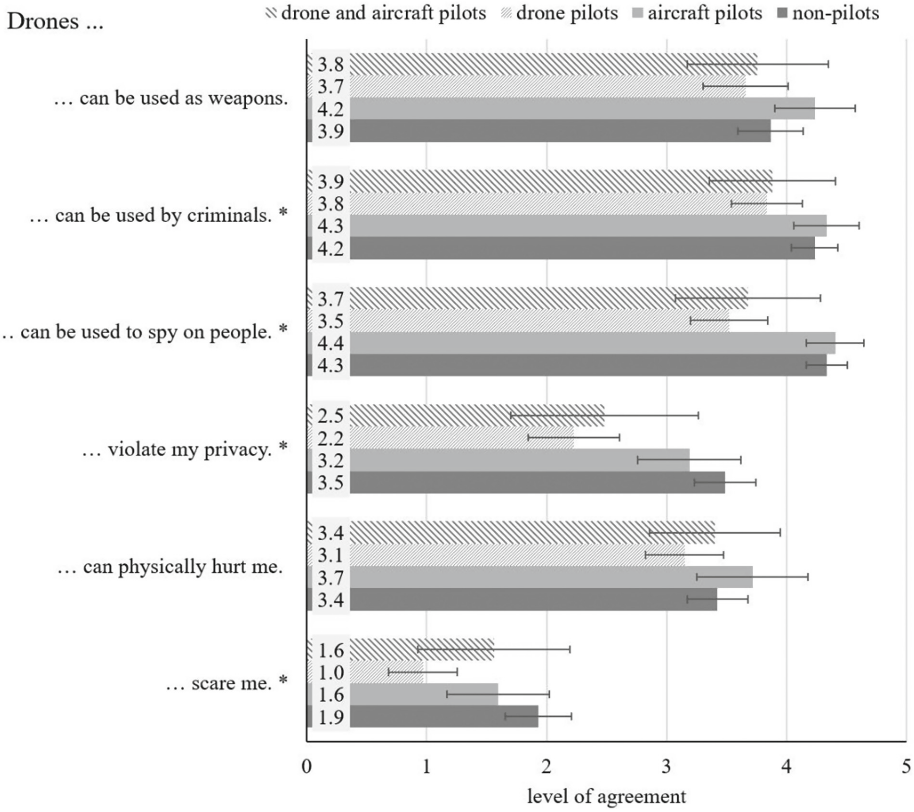


Fig. 4. Average agreement to concerns about drones (min = 0, max = 5) and related 95.0% CI depending on expertise groups (* indicating significant main effects)

Pairwise comparisons revealed that the aforementioned significant effects stem from the participant being a drone user or not and are independent of an aircraft piloting license. In general, agreement to concerns about criminal misuse, spying, and privacy violations was significantly lower for drone-users than non-users. The only exception was the scare potential: Here, aircraft pilots that used drones showed a response behavior comparable to non-users. Only drone-users without a piloting license differed significantly from all other expertise groups and showed almost no fear of drones at all.

Considering technical self-efficacy, slight but significant negative correlations with the fear of privacy violations (r = −.133, p = .045) and the potential scariness of drones (r = −.448, p < .001) were found: The higher the technical self-efficacy the lower the concerns. In contrast, the personal need for privacy was slightly positively correlated with most concerns. The higher the need for privacy the higher the concerns about criminal misuse (r = .199, p = .003), hazards for the own body (r = .226, r = .001), spy usage (r = .196, p = .003), violation of privacy (r = .269, p < .001), and the general fear of multicopters (r = .170, p = .010).

Human Factors Behind Problems. All questions that dealt with problems of drone flight caused by humans and their erroneous behaviors got high and comparable approval rates by all expertise groups. As can be seen in Fig. 5, the average agreement was close to the scale's maximum for all items and most groups. Noticeably, aircraft pilots without drone experience showed the highest agreement ratings out of all groups. However, this difference was only significant regarding the expected violations of legal regulations (p < .025 for all pairwise comparisons).

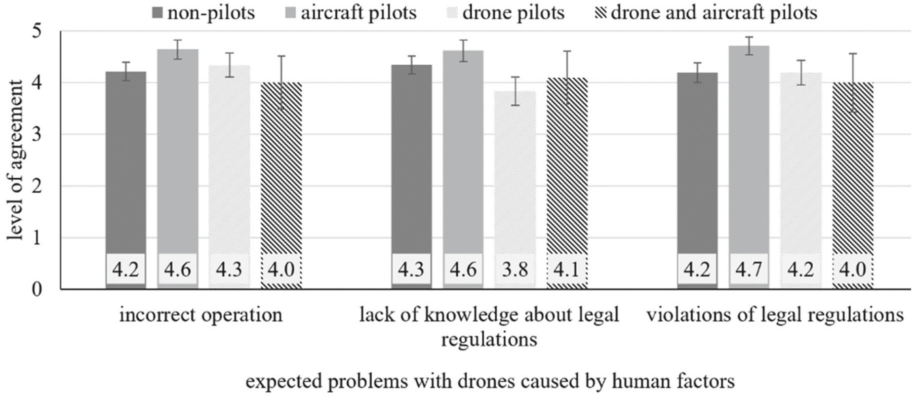


Fig. 5. Average agreement to expected problems with drones caused by human factors (min = 0, max = 5) and related 95.0% CI depending on expertise groups

While there were no significant correlations between the average agreement to the mentioned reasons and technical self-efficacy, the personal need for privacy was slightly positively correlated with both the expected incorrect operation (r = .190, p = .004) and the expected lack of knowledge about legal regulations (r = .164, p = .014).

5.4 Overflight Over Own Real Estate Property

The participants were asked for what reasons they would approve of drones flying over their own real estate property. As shown in Fig. 6, most surveyed causes got rather neutral agreement ratings ranging from slight rejection (e.g., "… if I partake of possible profits." or "… if the drone indicates whether it is filming.") to minor consent (e.g., "… if I know the flight's purpose." or "… if I'm the recipient of a drone delivery."). The operation by rescue services was the only fully accepted reason for drone overflights. In contrast, the operation by police showed only a slight approval, although it got the second highest rates out of all questioned reasons.

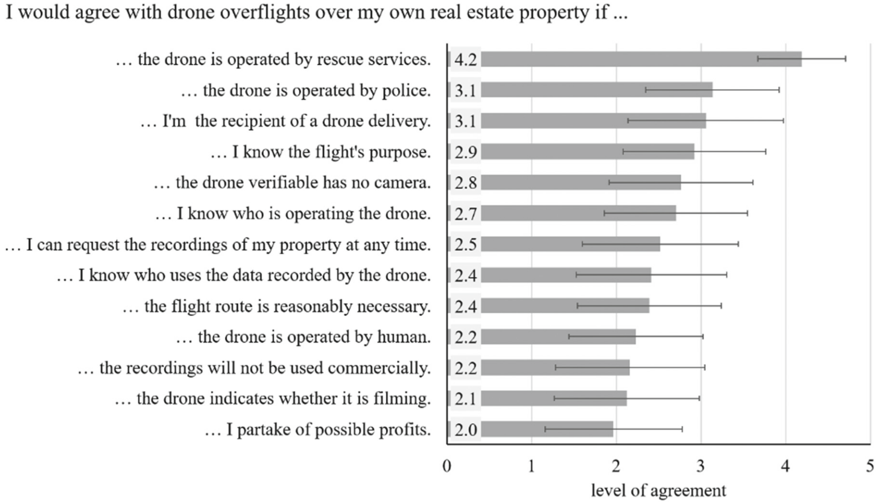


Fig. 6. Average agreement to reasons for accepting overflights (min = 0, max = 5) and related standard deviations

Table 1. Correlations between the need for privacy and reasons for acceptance of overflights (with *p < .01; **p < .001).

I would agree with drone overflights over my own real estate property if …	r
… I partake of possible profits.	−.390**
… the drone indicates whether it is filming.	−.312**
… the recordings will not be used commercially.	−.293**
… the drone is operated by a human.	−.284**
… the flight route is reasonably necessary.	−.324**
… I know who uses the data recorded by the drone.	−.263**
… I can request the recordings of my property at any time.	−.256**
… I know who is operating the drone.	−.296**
… the drone verifiably has no camera.	−.258**
… I know the flight's purpose.	−.253**
… I'm the recipient of a drone delivery.	−.372**
… the drone is operated by police.	−.208*
… the drone is operated by rescue services.	n.s.

Unlike with the surveyed questions mentioned before, neither the experience with usage of drones nor the possession of an aircraft pilot license had any significant influence on the acceptance of drone overflights. The same applies to the participants' technical self-efficacy.

However, a closer look at the personal need for privacy revealed altogether different results. Here, negative correlations could be identified for almost all reasons within the questionnaire to vindicate drones flying over private property (see Table 1). This means that the higher the personal need for privacy is, the lower is the acceptance of multicopter overflights of the own property. The only exception could be found for drones that are operated by rescue services. This is the only deployment context whose evaluation is independent from the individual privacy need.

6 Discussion and Limitations

With the rising interest of using drones in civil usage contexts, the approval of their deployment largely hinges on the public's acceptance. There are many concerns or perceived risks involved with aerial vehicles. These might be especially due to reports in newspapers and TV addressing near collisions between planes and multicopters.

The present study shed some light on the influence of aviation expertise on sharing airspace between manned and unmanned aircrafts. Interestingly, in most issues aircraft pilots did not differ in their opinion of drones from laypeople who had no experience with either multicopters or aircraft piloting.

Even though unexpected, being an aircraft pilot did not alter the opinion or perception of drones from non-pilots of planes or drones. Nevertheless – even though only marginal – aircraft pilots did exhibit the most caution and highest concerns when dealing with drones. Apparently, the experience with aircrafts also evokes a very cautious attitude. One could speculate that those pilots do not fear to crash their plane, but (near) collisions with drones are always troublesome.

Another aspect of drone deployment is the overflight over private property. For the whole group, each individual reason for deployment was evaluated as rather neutral. Still, laypersons with a high need for privacy showed more objection to each single reason for drone overflights. The issue of possible privacy violations was once again a deciding factor on the evaluation of drones and their deployment, especially regarding one's own property. The fact that the overall sample was quite undecided or indifferent towards tolerating drone overflight over private property might be hinged on the type of study. A simple questionnaire, in which the reasons are to be answered without relation to each other, might have led to an answering behavior without extremes. The well-known central tendency bias, an answering pattern centered around the middle of the answering options, might have come into effect here [26]. Thus, a Conjoint Analysis in which the reasons are to be weighed against each other and that allows to identify possible trade-offs between causes of drone overflights could be more appropriate. Future studies should therefore employ different empirical methods such as Conjoint Analysis to find trade-offs between acceptable and inacceptable reasons, especially regarding different usage contexts such as recreational, emergency, and commercial.

Furthermore, the study was conducted in Germany where landlords do not have to express prior consent to the use of airspace above their property. If anything, they can only lodge complaints if drones fly too low and they feel disturbed by the noise.

Another limitation of the present study is the group size within the sample as well as the overall number of aircraft pilots. That group is not represented adequately. Although we presume this did not alter the results significantly, further studies with equally distributed group sizes within the sample should be conducted to validate the results.

Acknowledgements. The authors thank all participants for their patience and openness to share opinions on a novel technology. Furthermore, thanks go to Sascha Krott for his research assistance.

References

1. Boucher, P.: Joint Research Centre, European Commission: Civil Drones in Society—Societal and Ethics Aspects of Remotely Piloted Aircraft Systems. European Commission, Ispra (2014)
2. Gąszczak, A., Breckon, T.P., Han, J.: Real-time people and vehicle detection from UAV imagery. In: Proceeding of SPIE: Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques, pp. 78780B–1–13 (2011)
3. Ștefan, D., Ștefan, M.-M.: The Drones are coming. What to choose? Low and medium altitude aerial archaeology on limes transalutanus. J. Anc. Hist. Archeol. **3**, 25–35 (2016)
4. Ostojić, G., Stankovski, S., Tejić, B., Đukić, N., Tegeltija, S.: Design, control and application of quadcopter. Int. J. Ind. Eng. Manag. **6**, 43–48 (2015)
5. Marin, L., Krajčíková, K.: Deploying Drones in policing southern European borders: constraints and challenges for data protection and human rights. In: Završnik, A. (ed.) Drones and Unmanned Aerial Systems, pp. 101–127. Springer, Cham (2016)
6. Pauner, C., Viguri, J.: A legal approach to civilian use of Drones in Europe. Privacy and personal data protection concerns. Democr. Secur. Rev. **3**, 85–121 (2016)
7. Gorkič, P.: The (F)Utility of privacy laws: the case of Drones? In: Završnik, A. (ed.) Drones and Unmanned Aerial Systems. Legal and Social Implications for Security and Surveillance, pp. 69–81. Springer, Cham (2016)
8. Lidynia, C., Philipsen, R., Ziefle, M.: Droning on about Drones—acceptance of and perceived barriers to Drones in civil usage contexts. In: Savage-Knepshield, P., Chen, J. (eds.) Advances in Human Factors in Robots and Unmanned Systems, pp. 317–329. Springer, Cham (2017)
9. Clothier, R.A., Greer, D.A., Greer, D.G., Mehta, A.M.: Risk perception and the public acceptance of Drones. Risk Anal. **35**, 1167–1183 (2015)
10. Solove, D.J.: A taxonomy of privacy. Univ. Pa. Law Rev. **154**, 477–564 (2006)
11. Solove, D.J.: Privacy: a concept in disarray. In: Solove, D.J. (ed.) Understanding Privacy, pp. 1–11. Harvard University Press, Cambridge (2008)
12. Acquisti, A., Taylor, C., Wagman, L.: The economics of privacy. J. Econ. Lit. **52**, 442–492 (2016)
13. Friedenzohn, D., Mirot, A.: The fear of Drones: privacy and unmanned aircraft. J. Law Enforc. **3** (2014)
14. Gruber, R.H.: Commercial Drones and privacy: can we trust states with "Drone Federalism"? Richmond J. Law Technol. **11**, 14 (2015)

15. Rao, B., Gopi, A.G., Maione, R.: The societal impact of commercial Drones. Technol. Soc. **45**, 83–90 (2016)
16. Wang, Y., Xia, H., Yao, Y., Huang, Y.: Flying eyes and hidden controllers: a qualitative study of people's privacy perceptions of civilian Drones in the US. Proc. Priv. Enhancing Technol. **3**, 172–190 (2016)
17. Schlag, C.: The new privacy battle: how the expanding use of Drones continues to Erode our concept of privacy and privacy rights. Pittsburgh J. Technol. Law Policy **13**, 1–22 (2013)
18. Pauner, C., Kamara, I., Viguri, J.: Drones, current challenges and standardisation solutions in the field of privacy and data protection. In: ITU Kaleidoscope: Trust in the Information Society, pp. 1–7 (2015)
19. Finn, R.L., Wright, D., Jacques, L., De Hert, P., Union, E.: Study on Privacy, Data Protection and Ethical Risks in Civil Remotely Piloted Aircraft Systems Operations. Summary for Industry. European Commission, Brussel (2014)
20. Luppicini, R., So, A.: A technoethical review of commercial Drone use in the context of governance, ethics, and privacy. Technol. Soc. **46**, 109–119 (2016)
21. Balaban, M.A., Mastaglio, T.W., Lynch, C.J.: Analysis of future UAS-based delivery. In: Roeder, T.M.K., Frazier, P.I., Szechtman, R., Zhou, E., Huschka, T., and Chick, S.E. (eds.) Proceedings of the 2016 Winter Simulation Conference, pp. 1595–1606 (2016)
22. Vattapparamban, E., Güvenç, I., Yurekli, A.I., Akkaya, K., Uluagac, S.: Drones for smart cities: issues in cybersecurity, privacy, and public safety. In: International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 216–221. IEEE (2016)
23. Wild, G., Murray, J., Baxter, G.: Exploring civil Drone accidents and incidents to help prevent potential air disasters. Aerospace **3**, 22 (2016)
24. Fontaine, O., Martinetti, A., Michaelides-Mateouc, S.: Remote Pilot Aircraft System (RPAS): just culture, human factors and learnt lessons. Chem. Eng. Trans. **53**, 205–210 (2016)
25. Beier, G.: Kontrollüberzeugungen im Umgang mit Technik [Locus of control when interacting with technology]. Rep. Psychol. **24**, 684–693 (1999)
26. Weisberg, H.F.: Central Tendency and Variability (No. 83). SAGE Publications, Newbury Park (1992)

Users' Trust in Automation: A Cultural Perspective

Hsiao-Ying Huang[1(✉)] and Masooda Bashir[2(✉)]

[1] Illinois Informatics Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA
hhuang65@illinois.edu

[2] School of Information Sciences, University of Illinois at Urbana-Champaign,
Champaign, IL, USA
mnb@illinois.edu

Abstract. With the rapid advancement of artificial intelligence technologies, people have more opportunities to interact and/or collaborate with smart automatic agents than ever before. To establish a successful relationship between humans and automation, trust plays a critical role. Therefore, our understandings of why and how people place and calibrate their trust toward intelligent automation become important but challenging issues. This present paper investigates people's trust in automation from a culture perspective. We conducted an online survey to measure people's general trust in automation and multiple dimensions of cultural tendencies. Our results indicate significant correlations between certain cultural tendencies and trust in automation. For instance, we found that people with more beliefs in horizontal collectivism and individualism also incline to have higher trust in automation. The multiple regression analysis also discovered predictable effects of cultural tendencies on people's trust in automation. These findings are further discussed in this presentation.

Keywords: Human factors · Trust · Human-automation interaction · Collectivism and individualism · Cultural psychology

1 Introduction

Today's technological innovations ranging from robotic surgery to self-driving cars have made automation deeply integrated and indispensible in our daily lives. Intelligent automation can be applied as a tool, teammate, or companion to humans across different settings, cultures, and societies. Users' trust in automation plays a critical role in establishing a successful relationship between humans and automation. For instance, misplaced trust can result in catastrophic consequences. Therefore, design features in automated systems that can help users calibrate adequate levels of trust in automation are fundamentally important in human-automation interaction [1].

The formation of trust in automation is a dynamic process that is affected by system performance and personal factors [2]. Previous studies have identified several personal factors that affect users' trust in automation, including dispositional trust, personality traits, age, gender, and ethnicity/culture [2–4]. Although the literature on human-automation interaction has acknowledged the impacts of personal factors on trust formation, we observe that there is still a lack of empirical evidence about how these factors exactly

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_27

influence users' trust toward automation, especially when it comes to cultural influences. Therefore, to fill this gap in knowledge, this paper aims to examine cultural influences on users' trust in automation from a cultural perspective [5].

A cultural perspective is a shared pattern of attitudes, beliefs, values, self-definitions, and norms, which can be found among people who speak the same language and/or come from the same geographical region or historical background [6]. Prior research has indicated that individualism is positively related to general trust in automation [7]. However, it is uncertain whether collectivism has a similar impact. In this paper, we focused on these two most defined cultural perspectives—individualism and collectivism. Drawing on Triandis [5], we examined the relationship between four dimensions of collectivism-individualism and people's trust in automation by adding horizontal (universalism/benevolence) and vertical (achievement/power) values. A total of four dimensions including horizontal individualism/collectivism and vertical individualism/collectivism were investigated in relation to trust in automation.

2 Background

2.1 Trust in Automation

Trust in automation plays an important role in the success of human-automation interaction. Based on trust, humans decide when intervention is needed to assist or repair an automated technology. In other words, level of trust can determine how much humans rely on a machine or automated system [8]. Inappropriate trust in automation can lead to not only misuse or disuse of a system but also disastrous consequences [1, 9]. Therefore, determining how to assist users in calibrating the appropriate level of trust toward automation is a critical concern.

Human-automation trust has been studied and defined in various ways [2, 10]. Lee and See [1] propose one of the most prevalent definitions of trust in automation. They describe trust as a human mental state of believing that an automated agent will assist an individual to achieve goals in an uncertain and vulnerable situation [1]. However, when automation fails to adjust to unanticipated situations, humans will start altering their trust toward that automated system. The formation of human-automation trust has been described as analogous to interpersonal trust [2, 11]. Fundamentally, both types of trust signify 'situation-specific attitudes' when people encounter uncertainties in a relationship. Nevertheless, human-automation trust and interpersonal trust are formed by different attributes. Human-automation trust develops based on the performance, process, and/or purpose of an automation [12], while interpersonal trust is based on the ability, integrity, and/or benevolence of a trustee [13].

The formation of human-automation trust involves complex factors. Before interacting with automation, human operators have some baseline degree of 'dispositional trust' toward an automated system, which is based on personal characteristics and the reputation of the system [2, 14]. When starting interaction, operators will develop dynamic 'learned trust' toward the system based on its performance and situation [2]. As Hoff and Bashir [2] indicate, trust in automation is a dynamic process depending on the characteristics of the human operator, the performance of the automated system, and the environment. Understanding the impact of human characteristics on trust formation can help us better design automated systems and assist operators in properly calibrating their trust toward the system.

Trust can vary across countries, races, genders, religions, and generations. Previous research has identified effects of personal characteristics on dispositional trust in automation, including age, gender, personality, and culture [2]. Although substantial research on human-automation trust has been done in this field, how these personal characteristics are related to human-automation trust still remains unclear, especially when it comes to cultural perspective [2, 10]. Considering the trend of globally automated systems, culture becomes an important factor because it varies among different individuals across the world. Cultural values can further affect an individual's judgment regarding whether to trust and/or accept an automated system [2, 4]. Therefore, understanding how individuals' cultural perspectives are related to trust in automation is critical and essential.

2.2 A Cultural Perspective: Collectivism and Individualism

Among studies on the diversity of cultural values, one well-known approach is understanding whether cultures tend toward collectivism or individualism [15]. Prior research in cross-cultural psychology defines collectivist societies as interdependent where relationships emphasize common bonds and individualist societies as independent where relationships emphasize consolidating personal freedom [15–17]. In other words, people from collectivist societies incline to prioritize their family, work, and society above individual needs, and their social behaviors are regulated by social norms [5]. On the other hand, people with individualistic backgrounds show less concern about societal responsibilities, and their behaviors are motivated by personal attitudes and joyful experiences [5, 15, 18]. Collectivism frequently manifests in parts of Europe and much of Africa, Asia, and Latin America, whereas individualism usually features in the United States and other western nations [19].

In terms of interpersonal relations, previous work suggests that individualism–collectivism relates to how willing people are to trust others [20–23]. Generally, collectivism exhibits a narrower trust radius and thus, people tend to reserve a given trust level for ingroup members and are less likely to extend the same level of trust to outgroup members [23]. Conversely, in individualist societies, people have a broader trust radius and are prone to express the same trust level toward both ingroup and outgroup members.

2.3 Trust in Automation from a Collectivism-Individualism Perspective

When it comes to people's trust in automation, studies have shown that people from different cultures have dissimilar attitudes toward automation [7, 10, 24–26]. Also, collectivist-individualist characteristics have been found to correlate with level of trust in automated systems. For instance, Chien et al. [7] discovered that people in the U.S. express higher general trust in automation than people in Taiwan and Turkey. They also revealed that people with more individualist tendencies have higher general trust toward automated systems. That is, similar to interpersonal trust, people exhibiting more individualist values seem to have a broader trust range, which can be extended to trust in automation. However, interestingly, another study by Huerta, Glandon, and Petrides [24] found that Mexicans—who generally tend toward values of collectivism—incline to place higher trust in automated decision aids and lower trust in manual decision aids, when compared to their American peers.

These inconsistent findings by previous work suggest a complicated relationship between culture and human-automation trust. As Triandis [5] articulates, collectivism and individualism are cultural tendencies that can coexist in any given individual's mindset. Thus, people's trust toward an automated system can vary due to the complexity of their collectivism-individualism combination. In an era of personalized technologies, we think that a better approach to understanding how collectivist-individualistic cultures are related to trust in automation is based on a micro-individual level, instead of a macro-societal level (e.g., countries, Western-Eastern soceites). Also, comparison among values contained within the concepts of collectivism and individualism is important [5, 27]. Therefore, we not only adopt the collectivism-individualism conceptual categories but also include a horizontal-vertical dimension, which results in four cultural tendencies, including: Horizontal collectivism, Horizontal individualism, Vertical collectivism, and Vertical individualism.

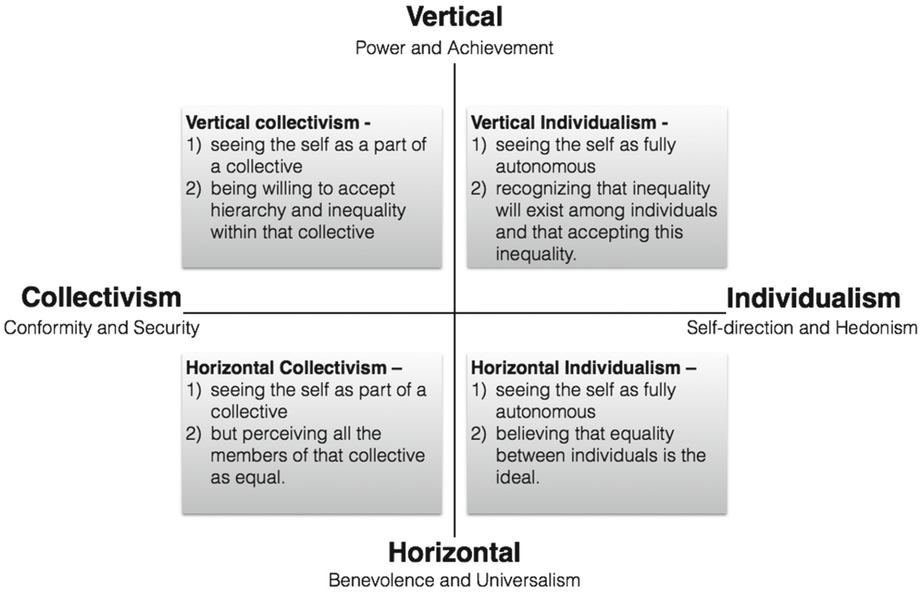

Vertical
Power and Achievement
Vertical collectivism -
1) seeing the self as a part of a collective
2) being willing to accept hierarchy and inequality within that collective
Vertical Individualism -
1) seeing the self as fully autonomous
2) recognizing that inequality will exist among individuals and that accepting this inequality.
Collectivism
Conformity and Security
Individualism
Self-direction and Hedonism
Horizontal Collectivism –
1) seeing the self as part of a collective
2) but perceiving all the members of that collective as equal.
Horizontal Individualism –
1) seeing the self as fully autonomous
2) believing that equality between individuals is the ideal.
Horizontal
Benevolence and Universalism

Fig. 1. Visualization for four cultural dimensions [5, 27]

As Fig. 1 shows, the horizontal value represents measures of 'benevolence and universalism,' and, on the contrary, the vertical value represents measures of 'power and achievement' [5, 27]. The meaning of each dimension is as follows [28]:

1) *Horizontal collectivism*: seeing oneself as a part of a collective while seeing each member of society as equal.
2) *Horizontal individualism*: seeing oneself as an autonomous individual while seeing each member of society as equal
3) *Vertical collectivism*: seeing oneself as a part of a collective while accepting hierarchies and inequalities within that collective.
4) *Vertical individualism*: seeing oneself as an autonomous individual while accepting hierarchies and inequalities between individuals.

In this study, we assessed these four cultural tendencies and their connections with people's trust in automation.

3 Method

We conducted an online survey to examine users' trust in automation from a cultural perspective. An online survey questionnaire was distributed via a crowdsourcing platform in February, 2015. All 156 participants completed the survey, 96 of whom were male. Approximately 43% of participants were between 18–30 years old, 51% were between 30–60 years old, and 6% were older than 60 years old. For education, 21% of participants had a high school degree, 55% completed college, and 24% earned advanced degrees. Furthermore, 80.8% of our participants were from the United States, 18.6% were from India, and 1% were from Sri Lanka.

Participants' general trust in automation and their cultural tendency were assessed. We adopted 12 items from Singh et al. [29] to measure dispositional trust in automation. For cultural tendency, a 16-item scale developed by Triandis and Gelfland [27] was used to assess four cultural dimensions of collectivism and individualism. Note that this study is only a part of our research project in trust in automation. More details and results of research will be published in the near future.

4 Results

According to correlational analysis, we found significant correlations between cultural perspectives and trust in automation. Participants who exhibited a higher tendency in horizontal individualism (r = .317, p < .001), horizontal collectivism (r = .225, p = .005), and vertical collectivism (r = .191, p = .017) showed a higher general trust in automation. Although these three dimensions all have positive correlations with trust in automation, they provide different explanations as to why people incline to place higher trust in automation, which we will discuss further in the next section.

We further conducted multiple regression analysis to examine if users' cultural tendencies have predictable associations with their trust in automation. The result revealed that four cultural dimensions contributed significantly to the regression model (F = 6.55, p < .001) and accounted for 14.8% of variance in participants' trust in automation. As shown in Table 1, two cultural dimensions have positive regression weights, which are horizontal individualism and horizontal collectivism. This suggests that participants who exhibit higher horizontal collectivism and individualism will also show higher trust in automation.

Table 1. Multiple regression analysis of general trust in automation by cultural tendency

	Standardized Beta	t-value
Constant		7.81 ($p < .001$)
Horizontal individualism	.310	3.99 ($p < .001$)
Vertical individualism	−.012	−.16 ($p = .872$)
Horizontal collectivism	.216	2.39 ($p = .018$)
Vertical collectivism	.005	.058 ($p = .954$)
R^2	.148	

5 Discussion

In this study, we present the results of an online survey assessing people's general trust in automation and their four cultural tendencies. The findings indicate significant correlations between certain cultural tendencies and trust in automation. Participants who exhibited more cultural beliefs in horizontal individualism, horizontal collectivism, and vertical collectivism also showed higher general trust in automation. As mentioned before, each of these cultural tendencies has a positive relationship with trust in automation for a different reason.

For people inclining to agree with horizontal individualism, they tend to see automation as a means of enhancing personal autonomy for each individual in society. They are thus prone to placing higher trust in automation in general. Similarly, for people identifying with horizontal collectivism, they seem to see automation as a beneficial tool that can advance the whole community or society, which results in their higher trust in automation. On the other hand, for people agreeing with vertical collectivism, they tend to regard automated systems as machines with more expertise and abilities than humans collectively. This belief in automation's superior power and achievement leads them to place higher trust in automation.

We further confirmed such predictable effects of cultural tendencies on general trust in automation by conducting multiple regression analysis. Interestingly, our results show that both horizontal individualism and collectivism are significantly positive predictors of trust in automation. This finding suggests that people holding more horizontal values, such as benevolence and universalism, will incline to have higher trust in automation, regardless of whether they believe in collectivism or individualism. This further implies that people with horizontal values may also have more positive bias [14] toward automation.

Overall our results provide a different viewpoint from previous studies and show a possibility that the collectivism-individualism perspective may not be sufficient for explaining people's trust in automation. Instead, the added horizontal-vertical dimensions may be able to provide a more comprehensive understanding about how people with diverse cultural values decide whether to place their trust in an automated system. Despite the contribution of this study, we also want to address certain limitations.

Although we found a correlation between trust and cultural tendencies, we can only provide hypothetical explanations of the results. Future studies interested in this area should conduct an in-depth interview or other experiment to gain better understanding about why people holding certain cultural tendencies display more trust in automation. Furthermore, we recruited participants from an online crowdsourcing platform, which limits the generalizability of our findings. We suggest that future studies recruit participants via multiple sources to enhance the diversity of participants.

6 Conclusion

With the rapid advancement of artificial intelligence technologies, we can expect that, in the near future, people will have more opportunities to interact and/or collaborate with smart automatic agents. Achieving a better understanding about why and how people place and calibrate their trust toward these smart automated agents becomes an extremely important but also challenging issue. This study investigates people's trust in automation from a culture perspective and the findings shed light on a new cultural dimensions for understanding this phenomenon. While there is still a long road to arriving at a comprehensive picture of human-automation trust, this study provides a stepping stone towards an understanding of how culture may influence our trust in automation and perhaps technology in general.

References

1. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**, 50–80 (2004)
2. Hoff, K.A., Bashir, M.: Trust in automation integrating empirical evidence on factors that influence trust. Hum. Factors J. Hum. Factors Ergon. Soc. **57**(3), 407–434 (2015)
3. Hancock, P.A., Billings, D.R., Oleson, K.E., Chen, J.Y.C., de Visser, E., Parasuraman, R.: A meta-analysis of factors impacting trust in human-robot interaction. Hum. Factors **53**, 517–527 (2011)
4. Schaefer, K.E., Chen, J.Y., 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. Hum. Factors **58**(3), 377–400 (2016)
5. Triandis, H.C.: The psychological measurement of cultural syndromes. Am. Psychol. **51**(4), 407 (1996)
6. Triandis, H.C.: Cultural syndromes and subjective well-being. Cult. Subject. Well-Being 13–36 (2000)
7. Chien, S.Y., Sycara, K., Liu, J.S., Kumru, A.: Relation between trust attitudes toward automation, Hofstede's cultural dimensions, and big five personality traits. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 60, no. 1, pp. 841–845. SAGE Publications, September 2016
8. 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. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 52, no. 19, pp. 1340–1344. SAGE Publications, Los Angeles, September 2008

9. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Factors **39**, 230–253 (1997)
10. Yerdon, V.A., Marlowe, T.A., Volante, W.G., Li, S., Hancock, P.A.: Investigating cross-cultural differences in trust levels of automotive automation. In: Schatz, S., Hoffman, M. (eds.) Advances in Cross-Cultural Decision Making, pp. 183–194. Springer, Cham (2017)
11. Nass, C., Moon, Y., Carney, P.: Are people polite to computers? Responses to computer-based interviewing systems. J. Appl. Soc. Psychol. **29**, 1093–1109 (1999)
12. Lee, J.D., Moray, N.: Trust, control strategies and allocation of function in human machine systems. Ergonomics **22**, 671–691 (1992)
13. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Acad. Manag. Rev. **20**, 709–734 (1995)
14. Madhavan, P., Wiegmann, D.A.: Effects of information source, pedigree, and reliability on operator interaction with decision support systems. Hum. Factors **49**, 773–785 (2007)
15. Triandis, H.C.: Individualism & collectivism. Westview Press, Boulder (1995)
16. Hofstede, G.: Motivation, leadership, and organization: do American theories apply abroad? Org. Dyn. **9**(1), 42–63 (1980)
17. Matsumoto, D., Weissman, M.D., Preston, K., Brown, B.R., Kupperbusch, C.: Context-specific measurement of individualism-collectivism on the individual level: the Individualism-Collectivism Interpersonal Assessment Inventory. J. Cross Cult. Psychol. **28**(6), 743–767 (1997)
18. Triandis, H.C., Bontempo, R., Betancourt, H., Bond, M., Leung, K., Brenes, A., Georgas, J., Hui, C.H., Marin, G., Setiadi, B., Sinha, J.B.: The measurement of the etic aspects of individualism and collectivism across cultures. Aust J. Psychol. **38**(3), 257–267 (1986)
19. Jiang, H.: Revisiting individualism and collectivism: a multinational examination of pre-service teachers' perceptions on student academic performances. Intercult. Educ. **27**(1), 101–110 (2016)
20. Yamagishi, T., Cook, K.S., Watabe, M.: Uncertainty, trust, and commitment formation in the United States and Japan 1. Am. J. Sociol. **104**(1), AJSv104p165-194 (1998)
21. Realo, A., Allik, J., Greenfield, B.: Radius of trust: social capital in relation to familism and institutional collectivism. J. Cross Cult. Psychol. **39**(4), 447–462 (2008)
22. Huff, L., Kelley, L.: Levels of organizational trust in individualist versus collectivist societies: a seven-nation study. Organ. Sci. **14**(1), 81–90 (2003)
23. Van Hoorn, A.: Individualist–collectivist culture and trust radius: a multilevel approach. J. Cross Cult. Psychol. **46**(2), 269–276 (2015)
24. Huerta, E., Glandon, T., Petrides, Y.: Framing, decision-aid systems, and culture: exploring influences on fraud investigations. Int. J. Acc. Inf. Syst. **13**, 316–333 (2012)
25. Kim, M.K., Heled, Y., Asher, I., Thompson, M.: Comparative analysis of laws on autonomous vehicles in the US and Europe. Ga. Tech. J., 1–12 (2014)
26. Schoettle, B., Sivak, M.: Public opinion about self-driving vehicles in China, India, Japan, the US, the UK, and Australia (2014)
27. Triandis, H.C., Chen, X.P., Chan, D.K.S.: Scenarios for the measurement of collectivism and individualism. J. Cross Cult. Psychol. **29**(2), 275–289 (1998)
28. Individualism and Collectivism Scale. http://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/CollectiveOrientation.pdf. Available on 26 Feb 2017
29. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced "complacency": development of the complacency-potential rating scale. Int. J. Aviat. Psychol. **3**(2), 111–122 (1993)

Human-Robot Teaming

Human Robot Team Development: An Operational and Technical Perspective

Jurriaan van Diggelen(✉), Rosemarijn Looije, Jasper van der Waa, and Mark Neerincx

TNO, Kampweg 5, 3769 DE Soesterberg, The Netherlands
{jurriaan.vandiggelen,rosemarijn.looije,jasper.vanderwaa,
mark.neerincx}@tno.nl

Abstract. Turning a robot into an effective team-player requires continuous adaptation during its lifecycle to human team-members, tasks, and the technological environment. This paper proposes a concept for human-robot team development over longer periods of time and discusses technological and operational implications. From an operational perspective, we discuss the types of adaptations to team behavior that are required in a military house search scenario. From a technological perspective, we explain how teamwork adaptations can be implemented using a teamwork module based on ontologies and policies. The approach is demonstrated in a virtual environment, in which humans and robots collaborate to find objects in a house search.

Keywords: Human robot teaming · Policies · Defense

1 Introduction

Turning a robot into an effective team player cannot be completely realized at design time. This is because many of its behavior requirements only become apparent after the system has been deployed. To illustrate this point, we consider a use case in which robots assist soldiers during a house search for explosive materials. To successfully participate in this mission, the robot must possess a diverse set of communication skills, e.g. for deciding to whom it should report its findings, or whether it can pick up some object without permission. It is highly unlikely that these behaviors have been perfectly pre-programmed by the robot development firm when the robot was delivered [1]. Therefore, they must be adaptable by the end user without the need of changing code.

As outlined in Fig. 1, we distinguish between three functional layers when designing human robot teams (HRT's). The lowest layer concerns robotic, human or joint activities of the team-members. The middle layer (*teamwork*) concerns all coordination activities that are required to enable these activities (e.g. allocating agents to tasks, or ensuring common ground [2]). These activities can be performed by one dedicated member or by all members of the HRT [3]. The upper layer concerns the functionality concerned with HRT development, i.e. monitoring and adapting the coordination activities to better match current circumstances.

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_28

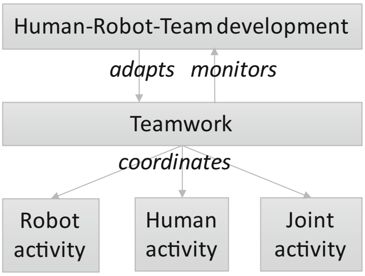


Fig. 1. Functional framework for human robot teamwork

In this paper, the focus is on HRT development which will be discussed from an operational and technological perspective. The operational perspective is illustrated in a military use case in which robots are allowed to participate in a debriefing of a mission. One of the goals of a debriefing is to allow the team to develop by sharing positive and negative experiences. In the case of the robotic participants, this means that explicit working agreements can be established between humans and robots which guides future behavior [4]. In this way, the functioning of the human-robot team is expected to improve as the team becomes more experienced.

The technological requirements are largely driven by this use case. To represent working agreements in a machine readable way, we have adopted a policy-based approach. Policies are a generic way to specify and govern an agent's behavior using rules for permissions and obligations. To make the policy engine applicable to our use case, we have built ontologies and a policy engine implemented in the Drools expert system [5]. The ontologies define the domain specific terms that are needed to specify relevant working agreements. For example, they specify what qualifies as a dangerous object. The policy engine defines rules of a specific format in a way that is computable and easily understandable to non-expert users. For example, using the ontologies and policy engine we can ensure that "the robot is not allowed to pick up dangerous objects" is understood by the robot.

The main contributions of this paper are an operationally relevant scenario containing debriefings for human-robot team (HRT) development, and a technological approach for HRT development based on ontologies and a rule-based policy engine.

We have tested our approach for HRT development with domain experts in the field using an implemented demonstrator. We have implemented the working agreements in a policy engine, and local agent behavior using behavior tree in a virtual environment. We implemented a test environment in which a soldier and robot jointly perform a house search and engage in a debriefing afterwards for HRT development.

The paper is organized as follows. Section 2 discusses related work. Section 3 discusses the operational aspects of HRT development using scenarios and possible working agreements. Section 4 describes our policy language and ontologies that we use to formalize these working agreements. Section 5 describes a demonstrator, followed by a conclusion in Sect. 6.

2 Related Work

The operational benefits of integrating robots in tactical teams (such as Special Weapons and Tactics (SWAT) [6], or Search and Rescue teams [7]) have been widely recognized. The most prominent reasons are enhanced capabilities offered by robots (e.g. using sensors such as radar or laser, or using strong robotic arms), and enhanced human safety by using robots as forward observers and allocating dangerous tasks to robots.

Most current applications of robots in the defense domain still require a high degree of operator involvement (e.g. tele-operated drones), and usually regard robots as a tool used by humans [8]. On the other hand observations and interviews with Explosive Ordinance Disposal (EOD) personnel [9] and industrial workers [10] who work in close cooperation with a robot show that people are inherently social creatures and attribute social features to a robot even when it is not autonomous. The expectation of researchers is that increased autonomous capabilities will increase the attribution of social features to robots. They expect that this will happen to such an extent that robots will be regarded as a teammate of humans, freeing up valuable human resources for other tasks. A practical military application of this idea is known as *manned-unmanned teaming (MUM-T)* where helicopter- and fighter jet pilots collaborate with UAV's during their operations [11].

From the academic community, the idea of human-agent teaming is quite old.

The first papers appeared in the early nineties and focused on capturing team properties in logical formalisms [12]. Based on these formalizations, executional concepts were soon proposed. Examples are agent communication languages such as KQML [13], which allow software agents to communicate on a higher level of abstraction. Another example is SharedPlans [14], which allows software agents to establish a common team plan. Another area of research on teamwork technology focusses on policies or norms as a way to govern and constrain the behavior of autonomous systems such that uncoordinated activities are ruled out [15]. This research area is still very active; a number of current challenges are reported in [16].

This paper builds upon insights from these technical and operational/human factors communities.

3 Operational Perspective

This section explains the steps we have taken to obtain the operational perspective of HRT. Because teamwork may occur in many forms, we have focused on one particular form of working together and described it in a scenario.

Relating to Fig. 1, the scenario should describe:

- The *coordination* that is required from the *teamwork* module. This includes all coordination activities, such as ordering, notifying, prohibiting, from robot to operator and from operator to robot.
- The *adaptation* and *monitoring* that is required from the *HRT development* module. This includes deciding when and how often the teamwork adaptation should take place, and which types of adaptations can be expected.

3.1 Scenario

In interaction with domain experts, we have developed the following scenario.

A house search is ordered for a house where there is a supposition of explosive materials. A tactical site exploitation team is deployed, the team consists of a team leader (TL), assistant team leader, and three search teams each consisting of two actors. The focus in the rest of the scenario will be on one of the search teams which consists of a human soldier (S1) and a robot (R1); search team 1 (ST1).

All search teams receive information from the TL about the situation (no people or boobytraps expected, two story house with stairs after entrance and several rooms on the ground floor, consolidation point for carefully retrieved evidence is outside the house).

Then ST1 receives instructions from the TL. ST1 is the first team to enter the house, it should check the hallway and route to stairs, check if provided layout of ground floor is correct and search the entrance. Inform me (the team leader) when route to stairs is clear.

ST1 walks to door, R1 waits on direct order from S1 as it is prohibited to enter without approval. S1 provides order to enter and check route to stairs. R1 enters hallway and checks route to stairs, informs S1 that situation is clear. S1 also enters and informs TL that route is clear. S1 checks ground floor plan and informs TL that it is according to intelligence. S1 then provides R1 with new shared task; in the middle of the room there is a consolidation point for evidence, you look clockwise, I'll look counterclockwise. R1 provides updates when it sees something (e.g. table, no danger). When room is clear according to R1 it reports this to S1, when S1 is also finished it reports the conclusion to TL. They then proceed to the next room, as ordered by TL, S1 enters R1 follows. S1 again introduces the consolidation point for this room and divides the task in clockwise (R1) and counterclockwise (S1). R1 sees something that might be an explosive, as it does not have the capabilities to investigate this it informs S1. S1 is busy with another object and does not react on R1. R1 then informs TL who sends S2 to the room to support.

After the house search is finished the complete team sits together for a debrief.

In this scenario R1 is prohibited to enter a room and obliged to inform S1 of everything. It is prohibited to inform TL directly unless S1 does not react. Adaptation of these policies can occur during the scenario, e.g. when S1 gets irritated about the updates S1 can say S1 prefers only updates on dangerous objects and not of everything. Another adaptation opportunity is during the debrief, where the decision can be weight between the different team members.

3.2 Working Agreements

After development of the scenario it was formalized in smaller steps (e.g. S1 provides new assignment to R1). These steps were then analyzed further, describing the specific working agreements they might require.

A situation as “before door” requires the following working agreement: If I (the robot) see a closed door or entrance to another room on the route, I will wait for instructions unless my human team member walks on or I already received an assignment for this entrance.

“Can’t do it” requires the working agreement: If I can’t perform my assignment I report to my human team member. Which in its turn requires the working agreement on reporting to S1. A working agreement that is contained in “report to human team member” is a further refinement that says: If my team member does not react I’ll rapport to the team leader.

A prerequisite for all these working agreements is that there is also an ontology present that for example defines *team member*. Furthermore, all working agreements must be described in such a way that they can be adapted, changed or refined. The first working agreement can for instance be changed so that the robot always walks in a room before the human team member. When the robot encounters a situation where it is not able to perform its assignment it could have the refinement of its working agreement that it informs S1 unless S1 is busy and it will then inform TL directly.

4 Technological Perspective

4.1 Hybrid Agent Architecture Model

The Hybrid Agent Architecture Model (HAAM) is shown in Fig. 2. Contrary to the *functional model* shown in Fig. 1, HAAM is a *system model* serving to identify the main components which are relevant when building Human Agent Teamwork.

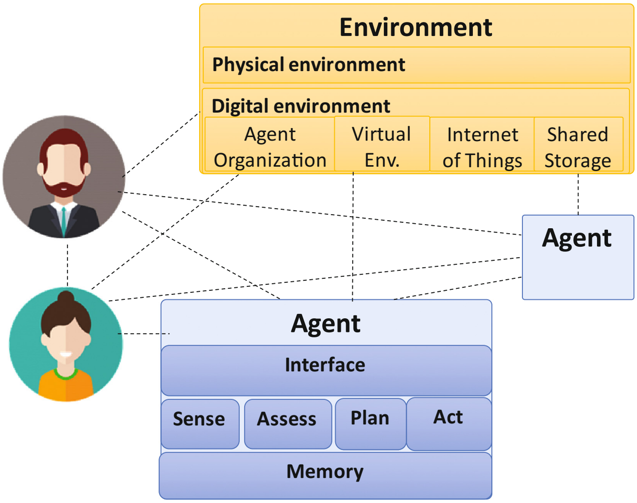


Fig. 2. Hybrid Agent Architecture Model (HAAM)

The main components of HAAM are humans, agents and a digital environment. Interaction may occur between all components (as indicated by the dashed lines). In our case, the humans are the members of the search teams, that collaborate with robots. The agents are embodied as robots (UGV's), and interact with each other and with humans. Agents and humans also interact with their environment, which consists of a physical parts (which we don't have to design, and hence is not specified further in this model), and a digital part (which we can design). Depending on the application, the digital environment specifies various shared components establishing common ground between the agents.

For our application, a shared agent organizational model is implemented as part of the digital environment, to specify the roles, permissions and obligations that are required for agent coordination. These policies can intervene directly in an agent's action selection cycle (sense; assess; plan; act), for example by blocking or enforcing agent's actions. The agent organization model and the agent's internal action selection cycle are discussed in Sects. 4.2 and 4.3 respectively. To test different concepts early in the development cycle, we have adopted a virtual environment (as a component in the digital environment, which will be discussed in Sect. 4.4.

4.2 Agent Organization

The agent organization model coordinates communication and activities. We follow a policy-based approach (similar to [17]). The policy engine uses contextual information and a set of working agreements to intervene in an agent's action selection cycle.

Working agreements (or policies) are formatted as rules in the form "*if <conditions are met> then <create policy decision>*". The conditions of a policy may contains contextual information which is provided by agents or the environment. When the conditions of a policy are met, a policy decision is created that states whether an action is obligated or prohibited. An example of a policy is:

***If** a robot finds a dangerous object and can secure the object **then** the robot must ask its team member if it is allowed to secure that object.*

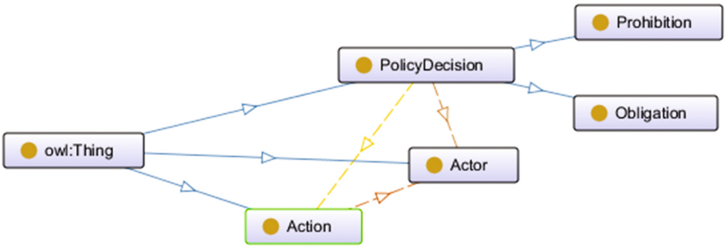


Fig. 3. The core ontology of the agent organization model

The policies are specified in the Drools expert system language. The information itself is structured according to a set of ontologies [18]. These define the common concepts between agents. The policy-engine has a single core ontology that defines the basic concepts of an action, actor and policy decision which can be obligations or prohibitions (see Fig. 3).

Besides the core ontology, which contains the most abstract and generic concepts required for HRT development, several domain specific ontologies specify the concepts that are needed to define the specific types of working agreements discussed in the previous Section.

4.3 Agent Action Selection

In our demonstration we used simulated robots in a virtual environment whose action selection loops are implemented with behavior trees [19]. A behavior tree is a tree-like graph where each node resembles an action, a condition check or a method how to propagate through the tree below it (e.g. breadth-first or depth-first). The agent's behavior is defined by propagating through the tree and perform each node's instruction. The resulting action chain can depend on environmental conditions, on previous actions or entirely external influences.

The agent organization model (i.e. the policy engine described in Sect. 4.2) intervenes in the action nodes; when an action node becomes active a query is send to the model and the policy decision is returned. In the case of a prohibited action, a different node becomes active either due to regular propagation through the tree or when the policy decision is accompanied by an obligated action. For example if a policy prohibits the action of securing a dangerous object, an additional policy may become active that obligates the agent in keeping its distance and send this action to the agent. Figure 4 shows a portion of the implemented behavior tree. The tree visualizes the real-time state of the agent's behavior as the game is played. The highlighted nodes are currently active.

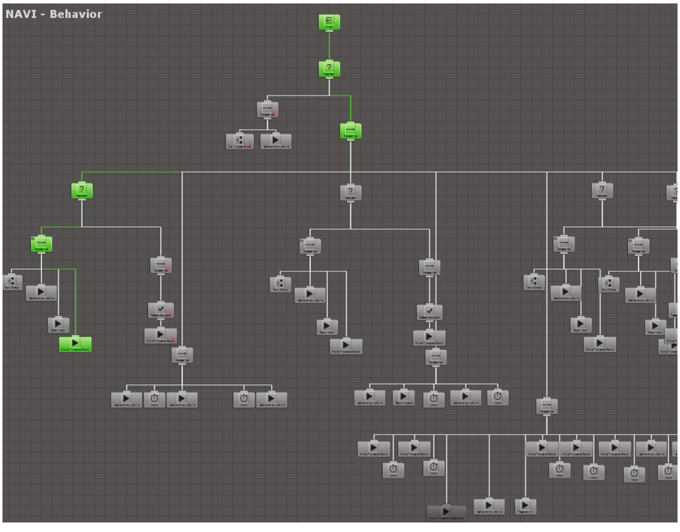


Fig. 4. A portion of the implemented behavior tree

The leaf node of the highlighted sub-tree is the node that is currently being executed. Nodes with a cross are rejected due to prohibiting policy decisions.

4.4 Virtual Environment

To test the human robot teamwork at an early stage of development, we implemented the house search scenario in a virtual environment (i.e. Unity [20]). A human participant can perform a house search form his or her own first person view. An autonomous robot helps the human in performing this task. Both robot and human can communicate with each other. The figure below shows the perspective of the robot. Note that this perspective is not visible by the human performing the house search, but can be generated for demonstration purposes (Fig. 5).

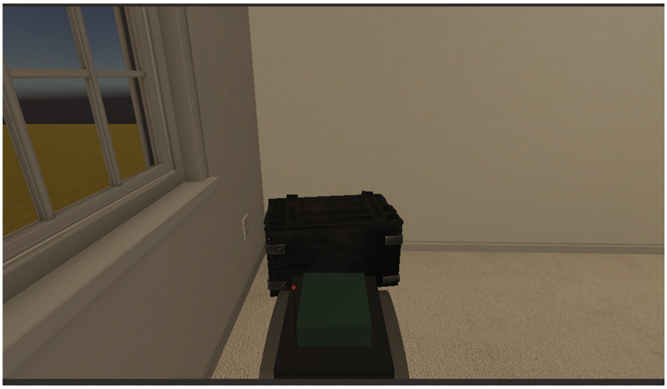
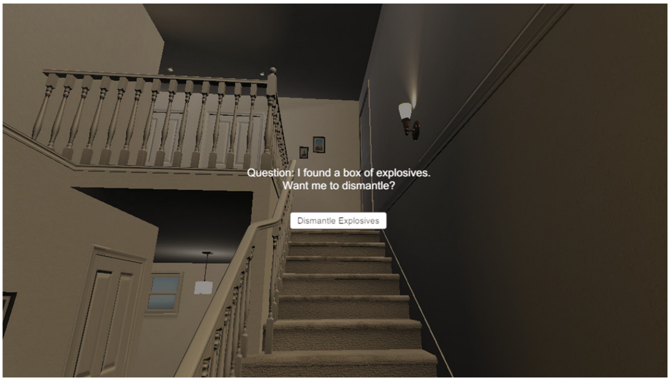


Fig. 5. Robot view of robot perceiving a black box

In the example above, the robot perceives a black box. Upon finding this box, ontology reasoning is used to classify the black box as a dangerous object, after which



Question: I found a box of explosives.
Want me to dismantle?

Fig. 6. Human view of human receiving a request from the robot

the policy engine fires a notification policy (discussed in Sect. 4.2) that this should be communicated to the operator. This is shown in the figure below (Fig. 6).

The human and robot can search the house in this way as a team, engaging in continuous communication. After their assignment is completed, the human and robot have the opportunity to reflect on their cooperation and may choose to adjust the policies that drive their communication, i.e. HRT development.

5 Conclusion

This paper presents an operational and technological concept for HRT development. The operational challenge is to enable operators to invest in improvements in team coordination over the long term, while maintaining their primary focus on the task at hand. We have developed a military house search scenario, and identified several types of human-robot team adaptations that are desirable in this domain. We propose that these adaptations should be pursued during the debriefing. From a technological perspective, we have proposed an ontological policy-based approach, where ontologies and a rule-based policy enforcement mechanisms serve as the primary means to represent and reason with these adaptations. The policies should be understandable by humans and machines. To enable human readability, we propose to use Domain Specific Languages. To enable machine-readability, we propose to use ontologies which are grounded in the underlying behavior mechanism (in our case behavior trees).

In the future, we intend to further investigate HRT development using this approach. We plan to develop the test environment further to also measure team performance. This enables measuring the effects of HRT development over longer periods of time.

References

1. Bradshaw, J.M., Hoffman, R.R., Woods, D.D., Johnson, M.: The Seven deadly myths of “autonomous systems”. IEEE Intell. Syst. **28**(3), 54–61 (2013)
2. 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**(6), 91–95 (2004)
3. Van Diggelen, J., Bradshaw, J.M., Johnson, M., Uszok, A., Feltovich, P.J.: Implementing collective obligations in human-agent teams using KAoS policies. In: Padget, J., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V.T., Matson, E., Polleres, A. (eds.) Coordination, Organizations, Institutions and Norms in Agent Systems V, pp. 36–52. Springer, Berlin (2010)
4. Neerincx, M.A., van Diggelen, J., van Breda, L.: Interaction design patterns for adaptive human-agent-robot teamwork in high-risk domains. In: Harris, D. (ed.) International Conference on Engineering Psychology and Cognitive Ergonomics, pp. 211–220. Springer, Cham (2016)
5. Drools: https://www.drools.org/
6. Asif, K.I., Bethel, C.L., Carruth, D.W.: Iterative interface design for robot integration with tactical teams. In: Savage-Knepshield, P., Chen, J. (eds.) Advances in Human Factors in Robots and Unmanned Systems, pp. 3–16. Springer, Cham (2017)

7. Kruijff, G.J.M., Janíček, M., Keshavdas, S., Larochelle, B., Zender, H., Smets, N.J., Mioch, T., Neerincx, M.A., et al.: Experience in system design for human-robot teaming in urban search and rescue. In: Yoshida, K., Tadokoro, S. (eds.) Field and Service Robotics, pp. 111–125. Springer, Berlin (2014)
8. Murphy, R.: Introduction to AI Robotics. MIT Press, Cambridge (2000)
9. Carpenter, J.: Culture and Human-Robot Interaction in Militarized Spaces: A War Story. Routledge, Abingdon (2016)
10. Sauppé, A., Mutlu, B.: The social impact of a robot co-worker in industrial settings. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3613–3622. ACM (2015)
11. 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. In: AIAA Guidance, Navigation, and Control Conference, p. 6237 (2011)
12. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artif. Intell. **42**(2–3), 213–261 (1990)
13. Finin, T., Fritzson, R., McKay, D., McEntire, R.: KQML as an agent communication language. In: Proceedings of the Third International Conference on Information and Knowledge Management, pp. 456–463. ACM (1994)
14. Tambe, M.: Towards flexible teamwork. J. Artif. Intell. Res. **7**, 83–124 (1997)
15. Bradshaw, J. M., Feltovich, P. J., Johnson, M. J., Bunch, L., Breedy, M. R., Eskridge, T., Jung, H., Lott, J., et al.: Coordination in human-agent-robot teamwork. In: International Symposium on Collaborative Technologies and Systems, 2008. CTS 2008, pp. 467–476. IEEE (2008)
16. Bradshaw, J.M., Montanari, R., Uszok, A.: Policy-based governance of complex distributed systems: what past trends can teach us about future requirements. In: Adaptive, Dynamic, and Resilient Systems, pp. 259–284. Auerbach Publications (2014)
17. Uszok, A., Bradshaw, J.M., Johnson, M., Jeffers, R., Tate, A., Dalton, J., Aitken, S.: KAoS policy management for semantic web services. IEEE Intell. Syst. **19**(4), 32–41 (2004)
18. Guarino, N.: Formal ontology, conceptual analysis and knowledge representation. Int. J. Hum. Comput. Stud. **43**(5–6), 625–640 (1995)
19. Marzinotto, A., Colledanchise, M., Smith, C., Ögren, P.: Towards a unified behavior trees framework for robot control. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 5420–5427. IEEE (2014)
20. Unity: https://unity3d.com/

Team Synchrony in Human-Autonomy Teaming

Mustafa Demir(✉), Nathan J. McNeese, and Nancy J. Cooke

Department of Human Systems Engineering, Arizona State University, Mesa, AZ, USA
{mdemir,nmcneese,ncooke}@asu.edu

Abstract. In Human-Autonomy Teaming (HAT), the development of a highly autonomous agent as a team member is difficult. Similar to human-human teaming, there are multiple dimensions of social behaviors that occur during HAT that must be accounted for within the development of the agent or system. One of these dimensions is team synchrony. In general, team synchrony is, when two systems (or two individuals in a team) are synchronized, resulting in their recurrences being dependent on each other. In order for a human-autonomy team to be synchronous the agent must communicate effectively (i.e., synchronize effectively) with its human team members. Thus, in this paper we present a conceptual discussion on what team synchrony is, how it occurs, and how to better develop it in HAT. To ground our discussion, we use our recent studies in which we empirically looked at team behaviors and team synchrony of HAT.

Keywords: Human-autonomy teaming · Synthetic agent · Teamwork · Synchrony

1 Introduction

Teams are complex, adaptive, and dynamic social systems that are driven by interactions among the team members [1]. A complex system's behavior emerges as a result of self-organization among the interacting parts of which the system is comprised [2]. Therefore, teams are the self-organizing results of individuals cooperating with each other, over time, towards team-level goals. According to this perspective, teamwork can be described in terms of the interactions among team members and between team members and their environment [3, 4].

Over the last couple of decades, with recent technological enhancements, teaming has changed. Increasingly, human team members work involves interactions with highly automated systems (e.g., synthetic agents and robots) with a trend toward autonomy taking on the role of a teammate [5]. Traditionally, teaming consisted of only human-human teams, but advancements in robotics and advanced machine learning have led the way for Human–Autonomy Teaming (HAT). This has led to a research interest in HAT [6]. Autonomy may be advantageous for teams in terms of reducing workload, but it can also adversely increase the cognitive demands on human teammates due to the autonomy's limited verbal behavior [6–8]. In addition, humans are familiar with interacting with other humans (i.e., human-human teaming) in this type of dynamic environment. Therefore, in HAT, the development of a highly autonomous agent as a team

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_29

member is difficult. Similar to human-human teaming, there are multiple dimensions of social behaviors that occur during HAT that must be accounted for within the development of the agent or system.

One of these dimensions is team synchrony. In general, team synchrony is, when two systems (or two individuals in a team) are synchronized, resulting in their recurring behaviors being dependent on each other [9]. In order for a team to be effectively synchronous, the agent must communicate and coordinate effectively with its human team members. Given the importance of team synchrony in human-human teaming, it is important to conceptually outline the role of team synchrony in HAT.

In our most recent experiment, we used a full-fledged synthetic teammate as a team member that was able to communicate and coordinate with other human teammates [10]. Based on our experiments, we have identified multiple empirical findings that we feel are important for advancing HAT, but more specifically team synchrony during HAT. These conceptual understandings based on empiricism help to identify what team synchrony means within the context of HAT, and how to empirically study team synchrony in HAT. Thus, at a high level, we discuss whether synchrony in human-human teaming can be transferable to human-autonomy teaming (HAT). Included is a discussion on what team synchrony is, how it occurs, and how to better develop it in HAT. To ground our discussion, we will use our recent studies in which we empirically looked at team behaviors and team synchrony of HAT.

2 Human-Autonomy Teaming

Human team members works increasingly involves interaction with highly automated systems (e.g., synthetic agents or robots) in highly dynamic environments, such as Command-and-Control or surgical rooms. This increasing role of highly automated systems as a team member has started a paradigm shift from human-human teaming to HAT. Autonomy is an independent system of human control [11], and it can be defined as systems that can function—at least partially in a self-directed manner—outside of the sorts of situations that they were designed for by using some intelligence-based capabilities [12].

There are several studies which have examined the use of autonomy or intelligent systems as a teammate [13–17]. For instance, one of these studies [14] considers intelligent systems as teammates, equivalent in status to humans, and therefore, they define 'team' as an "actor-agent community" (p. 35), and in their study, they underline that intelligent systems may take the initiative and give orders to fellow teammates. Another study [15], again considers this actor-agent community as a HAT and define it as "the dynamic, interdependent coupling between one or more human operators and one or more automated systems requiring collaboration and coordination to achieve successful task completion" (p. B64). Therefore, without outside intervention, an autonomous system can independently achieve goals and maintain good performance in highly dynamic environments by interacting with other humans or other agents [12, 18, 19]. Although these studies consider autonomy as a team member, one of the studies conducted by [20] underlines that the autonomous system's lack of intelligence is a large obstacle on its path to becoming a team member, and they also suggest ten steps to solve these obstacles.

Even if autonomy is advantageous for teams in terms of reducing the workload, it can also increase the cognitive demands that are placed on human teammates, because the autonomy has limited interaction behaviors. Thus, understanding dynamic interaction in HAT and how it differs from human-human teams needs to be taken more seriously in team science.

3 Synthetic Teammate Project

3.1 Synthetic Teammate

One goal of the synthetic teammate project is to create a synthetic teammate capable of human-like behavior in order to interact with other human team members. In the study, one of the team members was a synthetic teammate developed using the ACT-R cognitive modeling architecture [21], which is composed of five components [10]: (1) language analysis to accommodate a variety of English constructions, (2) language generation to choose possible utterances, (3) dialog modelling to recognize when communication is obligatory, (4) situation modelling for situation awareness, and (5) agent-environment interaction to fly an Unmanned Aerial Vehicle (UAV) between destinations.

At the beginning of the study, participants were informed that one of their teammates, i.e., the pilot, was a synthetic teammate and that it had limited communication capabilities. Therefore, in order to interact (i.e., communicate and coordinate) with the synthetic teammate effectively, the human team members needed to send messages in a constructive way - without any cryptic language or misspellings.

3.2 Synthetic Task Environment and Experiment

In the synthetic Unmanned Aerial System (UAS) environment, heterogeneous teams of three members (i.e., pilot, navigator, and photographer) are required to take good photos of critical target waypoints during simulated missions by interacting with each other via text-chat. In this task, three special-kinds of communication-coordination events can occur during the missions: Information-Negotiation-Feedback (INF) [22]. The interaction among team members at critical target waypoints is as follows: (1) the navigator produces a dynamic flight plan with speed and altitude restrictions of waypoints, and sends this *information* to the pilot; (2) the pilot controls the UAV in terms of heading, altitude and airspeed, and then *negotiates* with the photographer about altitude and airspeed; (3) the photographer adjusts camera settings based on the current altitude and airspeed, and photographs ground targets, and after that sends *feedback* to other team members [23] (see Fig. 1).

In this task, there were three heterogeneous team members that communicated via a text-based communications system. In addition to that there were three conditions (by manipulating the pilot role): (1) the Synthetic - the synthetic teammate was the pilot; (2)

the Control - an inexperienced human participant was a pilot; and (3) the Experimenter - one of the experimenters, who was experienced with the task, was the pilot.

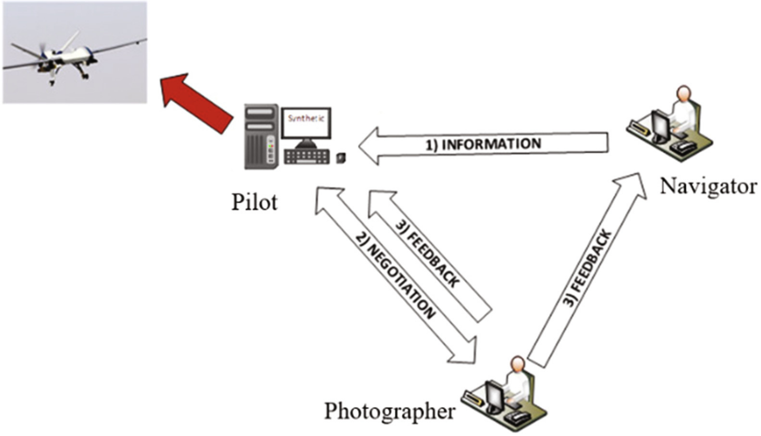


Fig. 1. Team Coordination (Information-Negotiation-Feedback) in HAT (Modified from [7]).

4 Team Synchrony

4.1 What Is Team Synchrony?

Synchrony/ synchronization is a complex dynamical process "wherein more than one dynamical systems are coupled or forced (periodically or noisy) in order to realize a collective or synchronous behavior" [24, 9, p. 160]. In this case, synchronization can be considered from two perspectives: (1) synchronization is a "process fundamental for the embodied grounding of *communication"* [25]; and *coordination*, which is the core aspect of synchronization, is behavioral synchronization among two or more individuals in space and time and, thus, those behavioral patterns can be represented as intentional, dynamic patterns [1]. Synchronization occurs when a unit of individuals, interacting over time, jointly engage in behavioral and affective monitoring. This serves an adaptive function that works by creating equality among the groups' neural, affective, and behavioral patterns [25].

In teams (which are adaptive, dynamic, and complex systems), individual team members are structural elements which temporarily act as one unit; this is also known as "synergy." [26]. Synergy is a team-level process that results in coherent, emergent team behaviors through the interactions of the individual team members. Therefore, we can view spatial-temporal synchronization between team members within a team as a functional synergy [26].

It is also important to explain how the heterogeneous team members' actions during the task can be synchronized within the team behaviors. In order to capture synchronization (which is the basis for team behaviors) among team members, we must first understand how repeated interactions among team members can scale to team behaviors [26].

4.2 Measuring Team Synchrony via Team Communication

Team communication is a subset of team synchrony. Therefore, communication as a purposeful social interaction is one of the ways to explain team synchronization, because it is a cognitive process in which representations are shared among the individuals [25]. One of the measurement methods for synchrony is recurrence plots. When two systems are synchronized then their recurrences are dependent on each other.

Recurrence Plots (RP) was originally introduced by [27] in order to visualize complex system dynamics, i.e., the behavior of trajectories of dynamical systems in phase space [28, 29]. After it was shown to be a strong predictor of data that captures key features of nonlinear systems, recurrence analysis emerged as a sophisticated way to illustrate system properties. For instance, a recurrence plot can be constructed to represent the dynamics of a single system across a delaminated period of time; from such a plot, researchers can develop metrics to represent properties of the system. This extension of the recurrence plot is called Recurrence Quantification Analysis (RQA).

Later on, extensions of RQA were put forward to look at more than one system and their dynamics: a bivariate version of RQA called Cross Recurrence Plots (CRP) and its analysis Cross Recurrence Quantification Analysis (CRQA), and, likewise, a multivariate version of RQA called Joint Recurrence Plots (JRP) and also its analysis Joint Recurrence Quantification Analysis (JRQA). By comparing the states of two different systems, CRPs reveal dependencies between the systems and show the progressions of two different phase space trajectories. Researchers can benefit from CRP by using it to discover recurrent structures between the behaviors of two individuals (be they in a single or multivariate state space) without having to make assumptions about data structure [30].

Another extension of RQA is JRQA (a nonlinear data analysis method) which is a mix of RQA and CRQA. This analysis is especially useful for assessing synchronization between interacting systems or assessing the systems that can jointly influence one another. In this method, first, the recurrences of the systems are plotted separately, then, the two separate recurrence matrices are combined to find times of simultaneous recurrence [29].

4.3 Using Joint Recurrence Plots to Examine Team Synchrony

One of the approaches for investigating team interaction patterns and their change over time involves looking at communication flow using JRQA that quantifies how many recurrences (and their length) are present by phase space trajectory in a dynamical system [31]. Within the team concept, JRQA can be applied to examine how and why several teams differ from one another in their dynamics. Accordingly, it shows the degree to which team members synchronize their activities during their interaction via text or voice communication.

In the synthetic teammate project, we applied JRQA on each of the UAV missions' communication dataset (time-series data sets), and extracted several measures. The following two measures are commonly used for our reports: Recurrence Rate – RR (ratio of all recurrent points in the upper triangle to size of recurrence plots), and Determinism – Det (ratio of all diagonally adjacent recurrent points to all recurrent points in the upper triangle) [31]. The Determinism measure, on joint recurrence plots, states of the system's epochs of similar time evaluation are the diagonal lines [32]. From this definition, chaotic (i.e., unpredictable) processes will have no (or very short) diagonals, whereas predictable (i.e., deterministic) processes will have longer diagonals and fewer single, isolated recurrence points [31]. This determinism ratio can take values between 0% (no repeats in the time series) and 100% (i.e., the time series repeats perfectly) (see Fig. 2). For instance, Fig. 2 shows three teams' joint recurrence plots from the synthetic, control, and experimenter conditions. These three examples of joint recurrence plots demonstrate three different synergies among three heterogeneous team members (i.e., navigator, pilot, and photographer) during the UAV task. In Fig. 2, the synthetic team shows more predictable team communication behavior with longer diagonals (Determinism: 69%) than the control (Determinism: 34%) and experimenter teams (Determinism: 38%). Figure 2 also indicates that the synthetic team also had a higher Recurrence Rate (RR= 38%) than the control (RR= 16%) and the experimenter teams (RR= 22%; see Fig. 2). In this example, the synthetic team had more synchrony than the other two teams. However, having more synchrony within the team does not equate to having higher team performance. In this example, the synthetic team's performance score was lower than that of the other two teams. Therefore, an argument can be made that the quality and effectiveness of the synchrony is more important than the quantity or frequency of synchrony.

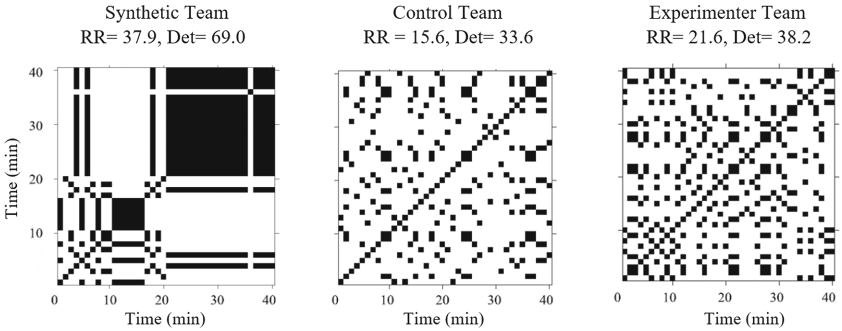


Fig. 2. Example Joint Recurrence Plots from three UAV teams' interactions (length 40 minutes)

5 Team Synchrony with an Autonomous Synthetic Agent

Team communication and coordination are subsets of team synchrony within the perspective of previous scholars' study [1, 25, 26]. To assess whether teams are effectively synchronous or not, we can look at the team interaction (i.e., communication and coordination) in two ways: from the human team member who interacts with an autonomous agent and from the autonomous agent that interacts with human team members.

An autonomous agent's interaction with its team members and task environment is crucial for good team performance in a dynamical task environment. The recent studies conducted for the synthetic teammate project [6, 7, 33] underline the importance of team communication behaviors and, in turn, team coordination. From the human team members' perspective, one of those studies indicated that adding a faux-synthetic teammate (played by a human) as a team member changed the human team members' communication behavior [33], and they exerted more control on the "synthetic team member". Later, another study [6] showed that due to the nature of limited communication behaviors of the synthetic teammate, it is required that human team members need to communicate with the synthetic teammate in stricter ways that it would find interpretable (i.e., no cryptic or misspelled language) or else resulted in intricacies in coordination that might have led to a failed mission. Another empirical study [7] showed that, even when human team members communicated properly with synthetic agent, it did not improve their team performance because of limited interaction capabilities within the team.

This limited interaction capability of the synthetic agent, and in turn among the team members in the HAT, lead to poor team adaptation in the dynamic environment, especially during roadblocks. Findings from one of the studies [8] highlights that the anticipation of other team members' behaviors and information requirements in synthetic teams was lower than the all-human teams. Therefore, developing team interaction mechanisms within HAT is needed for effective team situation awareness and in turn team performance [8].

Overall these recent findings from our studies address the limited team synchrony within HATs, because of poor temporal-spatial synchrony in behavior among the team members (which is distinct from coordination). In order to overcome the autonomous agent's communication and coordination limitations, the synthetic agent needs to have interaction patterns that are synchronous with humans in both temporal and spatial states in terms of *what* (communication), *when* (coordination), and *how* (communication and coordination) [6, 25]. This will also help to improve the synthetic agent's poor team situational awareness. To maintain effective team synchrony in a HAT requires continuous joint behavioral interactions among the team members by establishing neural, affective, and behavioral patterns. Thus, the initial state of effective team synchrony in a HAT starts with its subsets: effective communication and effective coordination among the team members.

6 Conclusion

As we continue to advance teaming from human-human to HAT there are many critical aspects of teaming that must be considered both during the conceptualization and application of HAT. Two such considerations are team communication and team coordination, culminating in the concept of team synchronization. If we are to create and motivate effective HAT, we must understand the impact of team synchrony and plan for improving synchronous aspects of communication and coordination. In this paper, we have presented both a conceptual understanding of the importance of team synchrony during HAT and provided details on how to study team synchrony in a HAT context.

Acknowledgments. This research was partially supported by ONR Award N000141110844 (Program Managers: Marc Steinberg, Paul Bello).

References

1. Arrow, H., McGrath, J.E., Berdahl, J.L.: Small Groups as Complex Systems: Formation, Coordination, Development, and Adaptation, 1st edn. SAGE Publications Inc, Thousand Oaks, CA (2000)
2. Kelso, J.A.S.: Dynamic Patterns: The Self-organization of Brain and Behavior. MIT Press, Cambridge (1997)
3. Cooke, N.J., Gorman, J.C., Rowe, L.J.: Team effectiveness in complex organizations: crossdisciplinary perspectives and approaches. In: An Ecological Perspective on Team Cognition, pp. 157–182. Taylor & Francis, Abington (2009)
4. Stevens, R.H., Gorman, J.C.: Mapping cognitive attractors onto the dynamic landscapes of teamwork. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) Foundations of Augmented Cognition. Directing the Future of Adaptive Systems, pp. 366–375. Springer, Berlin (2011)
5. Fiore, S.M., Wiltshire, T.J.: Technology as teammate: examining the role of external cognition in support of team cognitive processes. Front. Psychol. **7**, 1531 (2016)
6. Demir, M., McNeese, N.J., Cooke, N.J., Ball, J.T., Myers, C., Freiman, M.: Synthetic teammate communication and coordination with humans. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting vol. 59, no. 1, pp. 951–955 (2015)
7. Demir, M., McNeese, N.J., Cooke, N.J.: Team communication behaviors of the human-automation teaming. In: 2016 IEEE International Multi-disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), pp. 28–34 (2016)
8. Demir, M., McNeese, N.J., Cooke, N.J.: Team situation awareness within the context of human-autonomy teaming. Cogn. Syst. Res. (2016)
9. Dang, T.S., Palit, S.K., Mukherjee, S., Hoang, T.M., Banerjee, S.: Complexity and synchronization in stochastic chaotic systems. Eur. Phys. J. Spec. Top. **225**(1), 159–170 (2016)
10. Ball, J., et al.: The synthetic teammate project. Comput. Math. Organ. Theory **16**(3), 271–299 (2010)
11. Vagia, M., Transeth, A.A., Fjerdingen, S.A.: A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? Appl. Ergon. **53**(Part A), 190–202 (2016)
12. Endsley, M.R.: Autonomous Horizons: System Autonomy in the Air Force—A Path to the Future. Department of the Air Force Headquarters of the Air Force, Washington DC, Autonomous Horizons AF/ST TR 15-01, June 2015
13. Sycara, K., Lewis, M.: Integrating intelligent agents into human teams. In: Salas, E., Fiore, S.M. (eds.) Team Cognition: Understanding the Factors that Drive Process and Performance, pp. 203–231. American Psychological Association, Washington, DC (2004)
14. Wijngaards, N., Kempen, M., Smit, A., Nieuwenhuis, K.: Towards sustained team effectiveness. In: Boissier, O., Padget, J., Dignum, V., Lindemann, G., Matson, E., Ossowski, S., Sichman, J.S., Vázquez-Salceda, J. (eds.) Coordination, Organizations, Institutions, and Norms in Multi-agent Systems, pp. 35–47. Springer, Berlin (2006)

15. Cuevas, H.M., Fiore, S.M., Caldwell, B.S., Strater, L.: Augmenting Team Cognition in Human-Automation Teams Performing in Complex Operational Environments. Aviat. Space Environ. Med. **78**(5), B63–B70 (2007)
16. Langan-Fox, J., Canty, J.M., Sankey, M.J.: Human–automation teams and adaptable control for future air traffic management. Int. J. Ind. Ergon. **39**(5), 894–903 (2009)
17. Schulte, A., Donath, D., Lange, D.S.: Design patterns for human-cognitive agent teaming. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics, pp. 231–243. Springer, Cham (2016)
18. Schooley, L.C., Zeigler, B.P., Cellier, F.E., Wang, F.Y.: High-autonomy control of space resource processing plants. IEEE Control Syst. **13**(3), 29–39 (1993)
19. Krogmann, U.: From Automation to Autonomy-Trends Towards Autonomous Combat Systems. NATO Science and Technology Organization, France, Unclassified RTO MP-44 (1999)
20. 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**(6), 91–95 (2004)
21. Anderson, J.R.: How can the human mind occur in the physical universe? Oxford University Press, Oxford (2007)
22. Cooke, N.J., Gorman, J.C.: Interaction-based measures of cognitive systems. J. Cognit. Eng. Decis. Mak. **3**(1), 27–46 (2009)
23. Cooke, N.J., Shope, S.M.: Designing a synthetic task environment. In: Schiflett, L.R.E., Salas, E., Coovert, M.D. (eds.) Scaled Worlds: Development, Validation, and Application, pp. 263–278. Ashgate Publishing, Surrey (2004)
24. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization: A Universal Concept in Nonlinear Sciences, 1st edn. Cambridge University Press, Cambridge (2003)
25. Semin, G.R.: Grounding communication: synchrony. In: Kruglanski, A.W., Higgins, E.T. (eds.) Social Psychology: Handbook of Basic Principles, 2nd edn, pp. 630–649. Guilford Press, New York, NY (2007)
26. Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P., Richardson, M.J.: Competing together: Assessing the dynamics of team–team and player–team synchrony in professional association football. Hum. Mov. Sci. **32**(4), 555–566 (2013)
27. Eckmann, J.-P., Kamphorst, S.O., Ruelle, D.: Recurrence plots of dynamical systems. EPL Europhys. Lett. **4**(9), 973 (1987)
28. Blasco, R., Carmen, M.: Synchronization analysis by means of recurrences in phase space (2004)
29. Knight, A.P., Kennedy, D.M., McComb, S.A.: Using recurrence analysis to examine group dynamics. Group Dyn. Theory Res. Pract. **20**(3), 223–241 (2016)
30. Romero, V., Fitzpatrick, P., Schmidt, R.C., Richardson, M.J.: Using cross-recurrence quantification analysis to understand social motor coordination motor coordination in children with autism spectrum disorder autism spectrum disorder. In: Webber Jr., C.L., Ioana, C., Marwan, N. (eds.) Recurrence Plots and Their Quantifications: Expanding Horizons, pp. 227–240. Springer, Cham (2016)
31. Marwan, N., Carmen Romano, M., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex systems. Phys. Rep. **438**(5–6), 237–329 (2007)
32. Rizzi, M., Frigerio, F., Iori, V.: The early phases of epileptogenesis induced by status epilepticus are characterized by persistent dynamical regime of intermittency type. In: Webber Jr., C.L., Ioana, C., Marwan, N. (eds.) Recurrence Plots and Their Quantifications: Expanding Horizons, pp. 185–208. Springer, Cham (2016)

33. Demir, M., Cooke, N.J.: Human teaming changes driven by expectations of a synthetic teammate. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 58, no. 1, pp. 16–20 (2014)
34. McNeese, N.J., Demir, M., Cooke, N.J., Myers, C.W.: Teaming with a synthetic teammate: insights into human-autonomy teaming. J. Hum. Fact. Ergon. Soc. (submitted)

Toward an “Equal-Footing” Human-Robot Interaction for Fully Autonomous Vehicles

Theocharis Amanatidis(✉), Patrick Langdon, and P. John Clarkson

Engineering Design Centre, Cambridge University Engineering Department,
Trumpington Street, Cambridge CB2 1PZ, UK
{ta323,pml24,pjc10}@cam.ac.uk

Abstract. Fully autonomous vehicles can be classified as robots. In this paper we propose to approach the development of autonomous vehicle user interfaces from a human-robot interaction perspective, based on two principles. First, different robots require different user interfaces depending on their level of automation. Second, as the level of robot automation increases so should the automation of the interface itself; creating a spectrum ranging from a conventional “master-slave” level interaction to a fully intelligent “equal-footing” level interaction. Two research questions arise: where along the spectrum described above should autonomous vehicle user interfaces be, and what technological advance would have the greatest impact in enabling those interfaces. This paper presents the theoretical foundation of our research at the intersection of three previously unconnected fields: autonomous vehicles, human-robot interaction and affective computing. We then outline an experimental framework for developing a prototype interface based on our findings.

Keywords: Autonomous vehicles · Human-robot interaction · Affective computing

1 Introduction

The need for autonomous transportation is thought to arise from a combination of two societal factors [1]. The first is economic: population size of metropolitan areas is expanding, leading to increased congestion and cost of private vehicle ownership. The second is demographic: the median age of citizens in western societies is rising, causing reduced functional ability in an ever larger part of society. It would therefore be beneficial to develop measures to reduce the dependency of transportation on manual operation; one of the sort is the user interface of current automobiles [2]. One possible way to achieve this is through partial but progressively increasing autonomous control, all the way to fully autonomous vehicles [3]. Autonomous vehicles could allow the vehicle to be shared between a larger number of users with minimal inconvenience, collecting customers from and delivering customers to different locations [2]. This would reduce the number of vehicles needed per capita, reducing ownership costs and congestion [1]. Moreover, it will provide customers with reduced capabilities a means of transportation

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_30

not previously available to them [4]. Yet this shift is dependent on developing an appropriate user interface that would replace traditional automotive controls.

The Traveller Needs and UK Capability Study identifies two groups of users that would immediately benefit from the shift described above: progressive metropolites (tech-savvy young professionals in urban areas) and dependant passengers (young, elderly or people with impairments who usually don't hold a driver's licence) [2].

Our research goal is thus to develop a user interface for self-driving vehicles. In this paper, we will set the theoretical foundation of our research to achieve this goal using the following three-step process (Fig: 1): understand the *context* of autonomous vehicles, use the *principles* of human-robot interaction and implement some *technological advances* of affective computing. Each step is discussed in a section below.

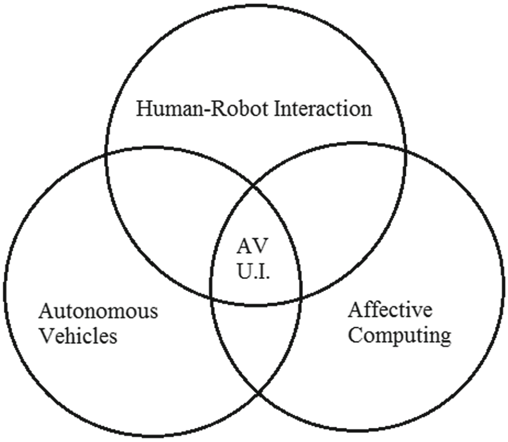


Fig. 1. Visualisation of our investigation for autonomous vehicle (AV) user interface development at the intersection of autonomous vehicle, human-robot interaction and affective computing.

2 Autonomous Vehicle User Interfaces

Autonomous vehicle research has progressed rapidly, to the stage that public trials have begun or are ready to begin in the coming months [5]. While extensive research has already been conducted on autonomous vehicle control systems, the development of their user interface is still in its infancy. In order to take advantage of the potential benefits of fully self-driving vehicles, a natural, inclusively designed user interface is needed. This interface should be designed to take into consideration the diverse range of capabilities and needs of different members of society and traveller types [2].

2.1 Levels of Autonomy in Self-driving Vehicles

To begin the process of investigating the design of such interface, it is important to understand the classification of different levels of autonomous vehicles. Each level requires a different type of interaction from the user, as discussed below. For the purposes of this paper the classification of different levels of autonomy in autonomous vehicles is based on the SAE J3016 standard [6], summarised in Table 1. Fully autonomous vehicles are defined as Level 5 automation on the SAE standard, and could be designed without any traditional manual controls such as a steering wheel or pedals. Vehicles in this highest level of automation do not require any user interaction to operate, even in emergency situations, other than selecting a destination and reporting progress. As such, the user may be out-of-the-loop for the entirety of the journey, allowing for minimal situational awareness and hence for a fully immersive interface if desired [7]. It is hence possible that users of these vehicles will not have to have a driving licence or be otherwise unable to drive due to age, temporary or permanent capability loss. Level 5 vehicles have thus the opportunity to achieve the goal of providing the freedom of the automobile to members of society that were previously excluded by capability [4]. The present paper focus on this level of automation.

Table 1. Summary of SAE J3016 automation levels and their description [6]. The dashed line signifies the transition from a lower to a higher degree of automation between the different levels.

Automation Level and Name	Execution of Control	Environment Monitoring	Emergency Response	System Capability
0 – No Automation	Human	Human	Human	Some Scenarios
1 – Driver Assistance	Shared	Human	Human	Some Scenarios
2 – Partial Automation	System	Human	Human	Some Scenarios
3 – Conditional Automation	System	System	Human	Some Scenarios
4 – High Automation	System	System	System	Some Scenarios
5 – Full Automation	System	System	System	All Scenarios

2.2 Related Work

A considerable amount of research has been undertaken on the user interface of partially and conditionally autonomous vehicles (Levels 2 and 3 on the SAE scale). The main focus is on designing the handover of control between periods of automated and manual driving – for a summary see [8]. Yet, as described in Sect. 2.1 above, in the context of Level 5 fully autonomous vehicles there will be no handover of control and hence potential designs are free to take a more creative approach to user interface design.

Surprisingly little research has focused on the issues specific to fully autonomous vehicle interfaces. One of the earliest studies was part of the ARGO autonomous vehicle project [9]. While the first prototype was based on a conventional vehicle dashboard, the second used a handheld Personal Digital Assistance computer (PDA) to control the vehicle. This would theoretically enable vehicle control from different seats or being passed on between different users in the car. This approach seems to be copied by the majority of recent projects using the modern equivalents of a PDA: a smartphone or tablet, such in concept cars by Rinspeed [10] or Mercedes [11]. Furthermore, it restricts the number of modalities used to visual, auditory and perhaps haptic, is not adaptable to user's capabilities and requires some level of prior knowledge or training.

3 Human-Robot Interaction for Autonomous Vehicles

Both historically [12] and by definition, fully autonomous vehicles can be classified as robots. For instance, the Cambridge Dictionary defines a robot as "a machine controlled by a computer that is used to perform jobs automatically" [13]. Therefore, we propose to investigate developing a user interface from a human-robot interaction foundation.

3.1 Spectrum of Autonomy in Human-Robot Interaction

Thrun argues that human-robot interaction cannot be studied without taking into account the level of autonomy of the robot [14]. This is because the level of autonomy will, along with other factors such as environment, determine the kind of tasks a robot can perform and the expectations humans will have of it. He defines two types of human robot interaction based on the two extremes of robot automation: what he calls indirect and direct interaction. In indirect interaction, as can be found in industrial robots, the user operates the robot as a "master" and the robot executes the command as a "slave". In direct interaction, as could be imagined with an artificial intelligence (AI) agent, the user and robot interact with each other on an "equal-footing". Thus, Thrun proposes a framework for human-robot interaction based on two principles. First, that different robots require different user interfaces depending on their level of automation. Second, as the level of robot automation increases so should the automation of the interface itself. While Thrun argues there are only two distinct types of interaction described above, Yanco argues that there is a continuum of robot autonomy and that the amount of intervention varies [15]. Based on these two assessments we propose there is a spectrum of autonomy in human-robot interaction which reflects the autonomy of the robot. A visual representation of this is shown in Fig.2 below.

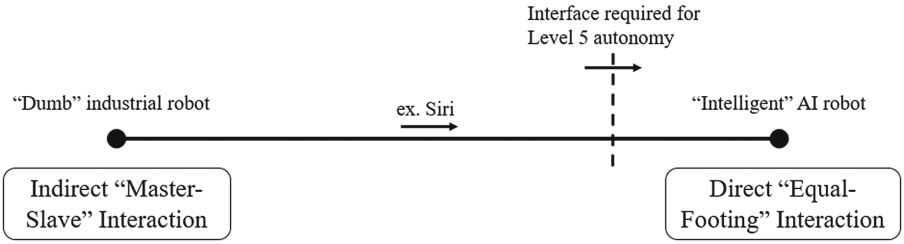


Fig. 2. Proposed spectrum of autonomy in human-robot interaction, based on the work by Thrun [14] and Yanco and Drury [15]. We aim to find where the dashed line is located along this spectrum.

3.2 Autonomous Vehicle Interfaces Based on the Spectrum of Autonomy

The spectrum of autonomy in human-robot interaction is of importance when trying to understand what kind of interface would be beneficial for an autonomous vehicle user interface. Two research questions arise from its study: where along the spectrum described above should our user interface be? And, what technological advance would have the greatest impact in enabling that interface?

Some researchers argue that the ultimate goal is to reach the end of the spectrum and an “equal-footing” interaction that would mimic human-human interactions, such as for instance between taxi driver and user. Norman argues that robots should adapt to humans rather than the other way around, and believes humans are much better at interactions at the “equal-footing” versus “master-slave” level [16]. This is reinforced by Stanton et al. [17], who argue that “the problem is not that automation is too powerful, the problem is that it’s not powerful enough” and that feedback and “conversation” is required. On the other hand, Thrun is unsure users would want to interact with robots the same way they do with humans or animals [14]. Finally, Pritchett and Feary warn about using findings based on human-human interaction to human-automation and human-robot interactions [18].

Given the above lack of agreement between researchers we believe that these questions may be answered experimentally, and we propose a plan of how to do so in a later section. Furthermore, the next section will present some possibilities that affective computing can bring to help move towards “equal-footing” interaction in fully autonomous vehicles.

4 Affective Computing in Autonomous Vehicle User Interfaces

Picard defines affective computing as: “computing that relates to, arises from or influences emotions” [19]. Based on this definition and her eponymous work, she argues that affective computing is not necessarily creating a machine with emotions but making a machine understand and/or convey emotions is sufficient to qualify as an affective machine. In this section we try to understand the motivation for using affective computing technologies in autonomous vehicle user interfaces and investigate some technological advances that may help develop an affective user interface.

Scheutz argues there are three benefits to affective computing: emotions make agents or robots more believable, recognising emotions is crucial to adapt to user needs and emotions are an integral part of control of complex agents or robots [20]. Similarly, Jaimes and Sebe argue that human-robot interaction that can sense the affective states of humans and accordingly adapt their behaviour are “likely to be perceived as more natural, efficacious and trustworthy” [21]. In the same text Jaimes and Sebe also conduct a thorough literature review of emotion recognition from facial expressions and audio. Their main findings are that these systems can differentiate between a small number of basic emotions relatively accurately but are not context-sensitive and do not analyse emotions on long enough timescales to infer mood or attitude. Finally, they suggest there may be unexplored potential in multi-modal emotion recognition. The two shortcomings discussed and the recommendation on multi-modal recognition are of importance for autonomous vehicle interaction and therefore need to be addressed.

We believe it is important to conclude with the motivation for affective, natural user interfaces in autonomous vehicles. This motivation is fourfold: (1) Improve user experience, (2) Enable customisation/personalisation, (3) Reinforce the brand

and (4) Increase inclusivity. We propose a system with facial and voice/tonal emotion recognition as discussed above as input and a virtual agent in the form of pilot or driver as output. Such a system could increase the levels of trust for new users and would be a step towards providing a chauffeur in every vehicle. The level of customisation and personalisation would vastly increase compared to existing systems, could adapt depending on situation or scenario and would constantly keep up to date with the latest trends. This system would also reinforce brand identity, be the differentiating factor between brands and be the face of the company not just in the vehicle, but potentially also in advertisements and showrooms. Finally, it would reduce any required training and enable new higher levels of inclusivity by adapting to each user's needs and capabilities.

5 Future Work and Conclusions

Based on the theoretical background presented in this paper, we plan to answer our research questions experimentally and develop a prototype interface based on our findings. Our process will be to first perform experiments in the low fidelity environment of a driving simulator and progressively increase fidelity. Initial experiments would use Wizard of Oz techniques for facial emotion recognition and natural language dialog. Experiments of increasing fidelity will use software tools for facial and voice recognition that are publically available in order to determine what technological advances may have the most impact.

In conclusion, this paper presented the theoretical foundation of our research at the intersection of three previously unconnected fields: autonomous vehicles, human-robot interaction and affective computing. We first discussed the different levels of autonomous vehicles and presented previous work on autonomous vehicle user interfaces within the fully autonomous context. We then discussed two principles of human robot interaction and proposed a spectrum of autonomy in human-robot interaction that we used to elicit our research questions. Finally, we presented our motivation for the adoption of affective computing technologies in autonomous vehicle interfaces and discussed the benefits and shortcoming of some of them. We will use this approach as a useful guide in developing and evaluating user interfaces for fully autonomous vehicles.

References

1. Mitchell, W.J., Borroni-Bird, C.E., Burns, L.D.: Reinventing the Automobile—Personal Urban Mobility for the 21st Century. The MIT Press, Cambridge (2010)
2. Transportation Systems Catapult Traveller Needs and UK Capability Study: https://ts.catapult.org.uk/wp-content/uploads/2016/04/Traveller-Needs-Study-1.pdf
3. Flemisch, F., et al.: Design of human computer interfaces for highly automated vehicles in the EU-project HAVEit. UAHCI, vol. 6767, pp. 270–279 (2011)
4. Jeon, M. et al: Towards Life-Long Mobility: Accessible Transportation with Automation. AutoUI Adjunct, pp. 203–208 (2016)
5. UK Autodrive: www.ukautodrive.com
6. SAE J3016 Standard: www.sae.org/misc/pdfs/automated_driving.pdf

7. Endsley, M.R.: Designing for Situation Awareness: An Approach to User-Centered Design. CRC Press, Boca Raton (2016)
8. McCall, R. et al: Towards A Taxonomy of Autonomous Vehicle Handover Situations. AutoUI, pp. 193–200 (2016)
9. Cellario, M.: Human-centered intelligent vehicles: toward multimodal interface integration. IEEE Intell. Syst. **16**(4), 78–81 (2001)
10. Rinspeed: http://www.rinspeed.eu/aktuelles.php?aid=17
11. Mercedes: www.mercedes-benz.com/en/mercedes-benz/innovation/research-vehicle-f-015-luxury-in-motion/
12. Moravec, H.P.: The stanford cart and the CMU rover. IEEE **71**(7), 872–884 (1983)
13. Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/robot
14. Thrun, S.: Toward a framework for human-robot interaction. Hum.-Comput. Interact. **19**(1–2), 9–24 (2004). http://www.tandfonline.com/doi/abs/10.1080/07370024.2004.9667338
15. Yanco, H.A., Drury, J.L.: A taxonomy for human-robot interaction. In: AAAI Fall Symposium on Human-Robot Interaction, pp. 111–119 (2002)
16. Norman, D.A.: The Design of Everyday Things: Revised and Expanded Edition. Basic Books, New York City (2013)
17. Stanton, N.A., et al.: The psychology of driving automation: a discussion with Professor Don Norman. IJVD **45**(3), 289–306 (2007)
18. Pritchett, A., Feary, M.: Designing human-automation interaction. In: Boy, G.A. (ed.) The Handbook of Human-Machine Interaction: A Human-Centered Design Approach. Ashgate, Farnham (2011)
19. Picard, R.: Affective Computing. The MIT Press, Cambridge (1997)
20. Scheutz, M.: Artificial emotions and machine consciousness. In: Frankish, K., Ramsey, W.M. (eds.) The Cambridge Handbook of Artificial Intelligence. Cambridge University Press, Cambridge (2014)
21. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: a survey. Comput. Vis. Image Underst. **108**, 116–134 (2007)

Virtual Reality Technologies for Training Applications and C2 of Unmanned Systems

Automated Collision Avoidance Developed Within a Mixed Reality System

Josh Wilkerson[1(✉)], Ryan McGee[1,2], Brian Reitz[1], Roberto Bellido[1],
Katia Estabridis[1], Gary Hewer[1], and Ryan Erb[1]

[1] Naval Air Warfare Center Weapons Division, 1 Administration Circle,
China Lake, CA 93555, USA
{joshua.l.wilkerson,brian.reitz,robert.bellido,
katia.estabridis,gary.hewer,ryan.erb}@navy.mil
[2] eScience Institute, University of Washington, Seattle, WA 98195, USA
rsmcgee@uw.edu

Abstract. As the use of unmanned systems becomes more prevalent in both commercial and military applications, increasing performance requirements have led to a greater demand for automation. The ability for an autonomous unmanned system to perform basic tasks reliably reduces the operator's cognitive tasks and workload, which could facilitate the control of multiple systems by a single operator. A key component for autonomous systems in many applications is the ability to perform reliable collision avoidance. A critical part of collision avoidance, often overlooked in existing algorithms, is that only a small set of velocities are actually reachable by the platform due to physical limitations or environmental factors. This paper presents the 3D Automated Velocity Obstacle Collision Avoidance (AVOCA) algorithm; a velocity obstacle based collision avoidance system that uses Kinematic Velocity Constraints (KVCs) to bind the velocity selection process. Results for AVOCA are presented from both simulation and experimentation using physical and virtual platforms in a Mixed Reality (MR) environment.

Keywords: 3-D collision avoidance · Mixed Reality · Human Machine Interaction

1 Introduction

The demand for unmanned systems is growing steadily in conjunction with their capabilities. The future of unmanned systems might include teams of Unmanned Vehicles (UVs) working harmoniously to accomplish a variety of tasks. This type of vision requires the platforms to have basic levels of autonomy to allow a team of agents to coordinate their actions in an unsupervised mode with the goal of reducing the operator's cognitive and workload tasks.

Autonomy research can assist with this challenge by allowing the operator to focus on high-level instructions for a swarm of UVs while relying on the autonomy algorithms to manage basics tasks such as trajectory planning and collision avoidance.

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_31

This paper presents a novel collision avoidance algorithm (called AVOCA), which has been developed and tested in a Mixed Reality (MR) environment system called the Autonomy Research Arena (AURA). AVOCA is a multi-vehicle collision avoidance algorithm that enables control of multi-agent teams by a single operator in a seamless manner. The algorithm has been demonstrated in AURA with physical and virtual agents. The goal of AURA is to support the development of autonomy technologies by providing a high-fidelity, flexible testing environment that is applicable to all stages of autonomy research and development. The MR environment also supports the development of software based autonomy algorithms (e.g., trajectory planners, collision avoidance) and human interface technology (e.g., gesture control, virtual reality).

This paper is organized as follows: Sect. 2 provides details on the AVOCA algorithm, Sect. 3 presents details on the operation and functionality of AURA (instrumental in the development of AVOCA), and Sect. 4 discusses the details of AVOCA experimentation.

2 The AVOCA Algorithm

The Automated Velocity Obstacle Collision Avoidance (AVOCA) system is a decentralized collision avoidance algorithm, capable of operating in both 2D and 3D applications. Key characteristics of the AVOCA algorithm include:

- Low inter-agent communication requirements
- Agnostic to agent swarm composition (i.e., supports heterogeneous and homogenous agent swarms)
- Algorithmic awareness of platform capabilities and restrictions
- Capable of operating with both cooperative and non-cooperative agents present
- Algorithmic awareness of obstacles present

The AVOCA algorithm utilizes VOs to perform collision avoidance, thus requiring position and velocity of all agents at each time step. AVOCA is agnostic to the source of this data, but in most cases this information comes from sensors, from communication among agents, or from a central agent management system (typical in simulations). Position and velocity of other agents is the only data required at every time step for AVOCA to compute collision free velocities.

AVOCA uses Kinematic Velocity Constraints (KVCs) to provide only command velocities that are expected to be realizable by the platform. Details on KVCs and their construction are provided in Sect. 2.2. The data needed by AVOCA to build the KVCs must be provided at least once at the start of a run; after that, AVOCA will accept new KVC data at any point. This allows the user to determine appropriate KVC data beforehand, employ the use of platform dynamics calculations at run-time to generate the data, or anything in between these cases.

Figure 1 gives an overview of the AVOCA algorithm. AVOCA uses relevant agent and obstacle data to construct the problem space for the current time step. After constructing the problem space, the algorithm begins to search for the reachable, collision free velocity that is closest to preferred velocity. This process is iterative, but is any time in nature; i.e., after completing a single iteration a solution has been found that can be used as a final command velocity. Additional iterations get this solution closer to optimal.

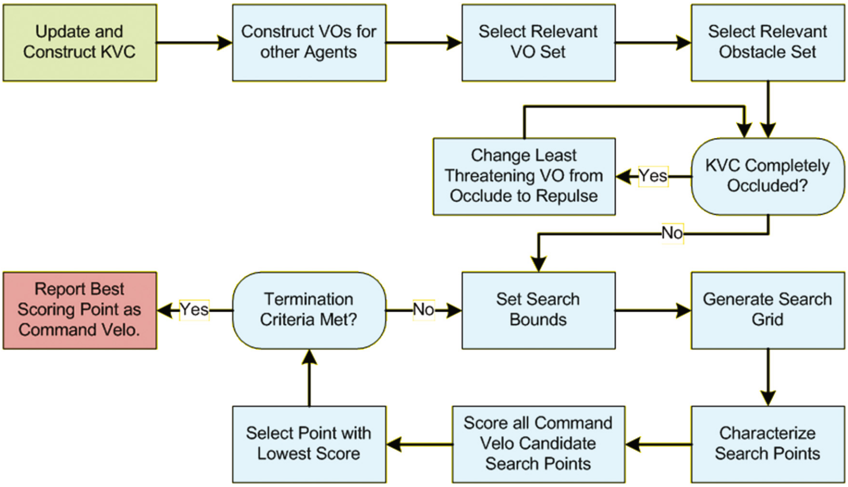

Update and Construct KVC
Construct VOs for other Agents
Select Relevant VO Set
Select Relevant Obstacle Set
KVC Completely Occluded?
Yes
Change Least Threatening VO from Occlude to Repulse
No
Set Search Bounds
Generate Search Grid
Characterize Search Points
Score all Command Velo Candidate Search Points
Select Point with Lowest Score
Termination Criteria Met?
No
Yes
Report Best Scoring Point as Command Velo.

Fig. 1. AVOCA overview

2.1 AVOCA Agents

In AVOCA, agents are the abstract representation of a dynamic entity in the problem space. The first step in the AVOCA algorithm is to define the set of agents and their characteristics in the problem space. This includes:

- An avoidance region, indicating the region in space that no other avoidance regions should overlap. AVOCA supports circular and convex polygon regions for 2D applications and spherical and ovoid regions for 3D
- A position for the agent (the centroid of the avoidance region is placed at this point)
- The current velocity of the agent

Based on the agents and their characteristics, AVOCA issues command velocities that are bound by a set of kinematic constraints (discussed in Sect. 2.2). Information about avoidance regions and kinematic data is generated beforehand and provided at initialization. This information can then be updated during the mission as needed. Agent positions and velocities are then provided/updated via on board sensors or from external sources.

2.2 Kinematic Velocity Constraints

All platforms have limitations on what they are capable of doing in a given period. These limitations can come from a wide variety of sources, e.g., controllable degrees of freedom (fixed wing versus rotary wing aircraft), platform weight and engine capabilities, and operational rules/constraints (rules of the road). Additionally, platform capabilities can be impacted by environmental conditions, e.g., slick surfaces or wind. Regardless of the source, limiting command velocities to achievable values is critical to a collision avoidance algorithm since the vast majority of collision-free velocities are non-realizable in a given time step (even for high-agility platforms). This is especially true when agents are operating in a dense formation, at high speeds, and/or with low maneuverability (i.e., at times when collisions are most likely).

The AVOCA algorithm uses constructs called KVCs to represent the aggregate set of velocity limitations at each time step. The better AVOCA understands its own vehicle and its current operational environment, the better its performance will be. The determination of active velocity limitations is outside the scope of AVOCA. KVCs are linked to the AVOCA Application Program Interface (API), allowing external algorithms to provide data on current limitations. This data can be provided a priori or updated as needed during run time.

2.3 Collision Free Command Velocity Selection

AVOCA uses VOs to select collision free velocities [1]. A VO is a geometric shape in velocity space that represents a set of significant candidate velocities. VOs are very flexible in concept, and the significance of the candidate velocities is dictated by the application. In the context of collision avoidance, VOs are typically unbounded triangles in 2D and unbounded cones in 3D that represent the set of velocities that will result in a collision with another agent. A given agent creates one VO for each of the other agents in the scene. The combined set of velocities contained within the resulting VOs represents all of those that would result in a collision if the agents held course. Naturally, it is unreasonable to expect the other agents to hold course. However, if collision free velocities are selected and applied at a sufficiently high frequency (application specific), a collision free path will be the result as demonstrated in [2].

VOs in AVOCA are always 3D. Dimensionality restrictions are applied via the KVCs to simplify the construction process (discussed in Sect. 2.2). The VO construction methodology is similar to that shown in [3]. The goal of AVOCA is to find the collision free velocity that is realizable and as close as possible (based on Euclidean distance) to the provided preferred velocity. This means that AVOCA must select a velocity that falls inside the KVC region, outside the VOs and obstacles, and is as close to the preferred velocity as it can be. The result of this process is issued as the agent command velocity for the current time step.

3 AURA

The AURA system hosts an MR environment intended to enable incremental testing of autonomy algorithms throughout their development. This is accomplished by providing a flexible environment composed of both virtual and physical participating agents seamlessly interacting with each other. Hönig et al. [4] elaborate on the advantages of a MR robotics framework that enables bidirectional interaction among arbitrary physical and virtual components. These benefits include modifications of robotic platforms (e.g., addition of virtual sensors) and scalability of swarms that would not be feasible in the real world. In addition, Hönig et al. illustrate these advantages in several demonstrations that involve multiple robots and robotics simulators.

Initial testing of an algorithm, for example, could be done with only virtual agents present. This provides benefits for early testing such as limiting external factors that could influence the algorithm, removing risk of platform damage, and allowing faster testing. Intermediate testing could include a single physical agent interacting with one or more virtual agents. This would allow developers to see what external factors may affect the algorithm while still reducing risk in the event of a failure. Advanced testing could involve two or more physical agents in a controlled simulation environment to investigate the interactions between physical agents.

Figure 2 depicts key components of the AURA architecture as well as their interactions. The Human Machine Interface (HMI) portion provides high-level instruction to the agents in the system. These instructions can be provided via conventional interfaces, using gesture control, or virtual reality technologies. Additionally, agents are provided with current details from the known obstacle map (if, for example, obstacle sensing is not part of the intended test). Users can exploit the immersive nature of virtual reality to interact with obstacles in the virtual scene.

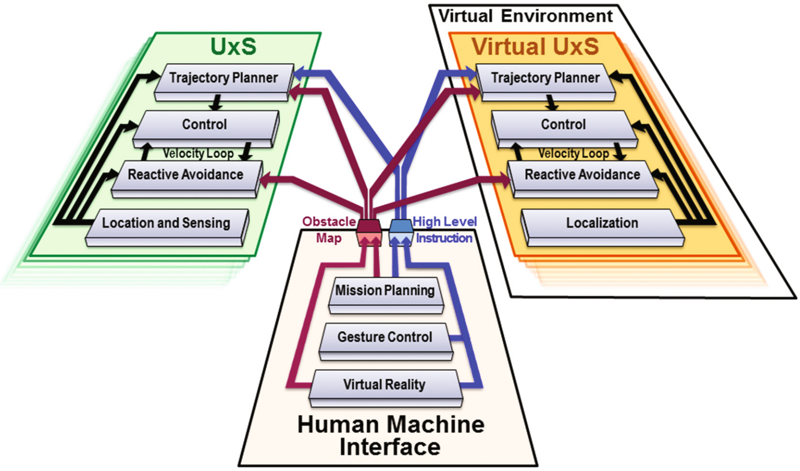


Fig. 2. Architecture for AURA

Virtual and physical agents are handled in a nearly identical manner. Preferred velocities are generated using a trajectory planner. These are provided to a control loop, which is also provided localization information and sensing data, in the case of physical agents.

Figure 3 provides an overview of the AURA system implementation. The required components for basic operation of AURA are outlined in the upper left and upper right panels (i.e., agent ecosystem manager and localization system). The remaining components are available, depending on the needs of the testers. The bottom right-most panel of Fig. 3 indicates the HMI capabilities of the system. AURA is capable of rendering the virtual environment for the tester conventionally or using virtual/mixed reality interfaces. This allows the tester to immersively view and interact with the simulation as needed. These components are described in more detail in the following sections.

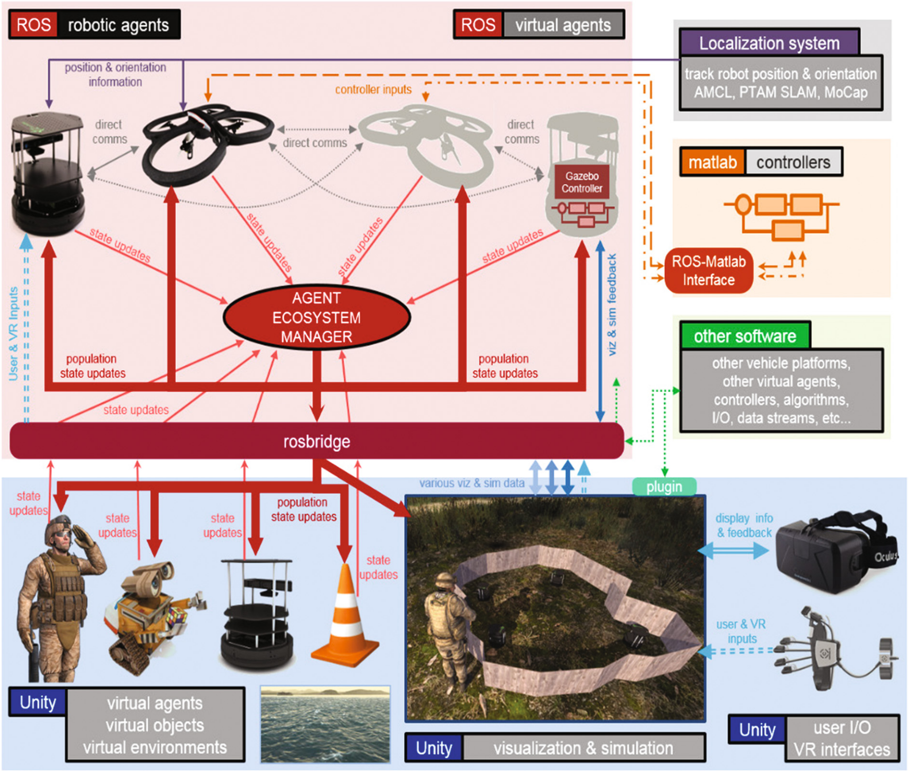


Fig. 3. AURA overview

3.1 Infrastructure

A key element of indoor robotic simulation is agent localization. AURA is designed to be compatible with self-localization methods as well as external localization sources. The experimentation presented here used an 80-camera Vicon motion capture arena to provide localization for the physical agents [5]. High precision agent pose data is provided to AURA from the motion capture system, transformed from real-world coordinates to the virtual simulation coordinate frame, and finally applied to the associated AURA agents.

AURA uses the Gazebo simulator for management of virtual agents [6]. Virtual agents are modeled and controlled in Gazebo, which uses a physics engine to provide realistic agent behavior. Gazebo provides pose information to the virtual agents in AURA, which is transformed into the appropriate virtual world coordinate frame and applied to the associated agents.

AURA supports terminal based control as well as a textual GUI built using ROS functionality. The Unity game engine [7] is used to render the AURA environment and provides the tester with virtual and mixed reality interfaces. The AURA Unity world is the single point where the full virtual environment is visible to the user, rendering all agents/objects known to the system (both physical and virtual) as well as algorithm data generated by the agents. The tester can interact with the simulation via the Unity world as well, such as modifying virtual obstacles in the simulation, issuing high-level commands to the agent swarm as a whole, etc. This functionality both provides unique perspective into the simulation at run time as well as allows for studies into the roles of emergent virtual/mixed reality technologies in autonomy. AURA currently interfaces with Leap Motion [8] and Oculus Rift [9], and integration with Microsoft Hololens [10] is in progress.

3.2 AURA Operation

AURA is built using the Robot Operating System (ROS) framework [11]. Key elements provided by the ROS system include:

- Framework for communication among system components
- Modular design paradigm
- Tools for viewing and interacting with system data both at run-time and after

The central component of the AURA is the agent ecosystem manger, shown in Fig. 3. All AURA agents run separately, requesting membership from the ecosystem manager. In turn, the ecosystem manager provides information on other elements of the ecosystem to the agents at regular intervals. Both the ecosystem manager and agents consist of multiple ROS-based components handling various aspects of their operation. The ROS framework provides flexible means for swapping these components for new ones while enabling communication locally as well as with non-ROS components in the system. External components of the ROS system utilize rosbridge, which provides an easily accessible and flexible means of communication. The rosbridge suite is a ROS package that allows non-ROS software to communicate with the ROS framework. AURA agents have three modes of possible operation, which are physical centralized, physical distributed, and virtual.

4 Experimentation

4.1 Unity World Experimentation

Initial AVOCA experimentation was performed in a Unity virtual world simulator designed to provide an operational environment in which there are no physics engine or vehicle dynamics. The goal of these tests was to observe AVOCA operating with as little outside influence as possible. The scenario used in these tests had four agents in a square pattern, moving to the corner diagonal from their starting location. Once all agents reached their goal locations (termed as completing a crossing), they returned to their starting location. Each agent used a spherical avoidance region. The radius of these regions was set to 25% more than the minimum radius needed to contain the game model used in Unity. Overlaps observed by the testing system were reported as percentage of minimum separation distance and were calculated using Eq. 1.

$$\frac{radius_1 + radius_2 - sepdist}{radius_1 + radius_2} \tag{1}$$

Based on the physical radius of the model, any overlap over 20% represents a collision. This rough approximation assumes the model is a sphere and neglects any difference between width and height. The scenario was run 20 times, with each run consisting of 10 crossings by the agents. The results of these runs are summarized in Table 1.

Table 1. Summary of results from unity experiments

Runs	20	Avg. number of observed overlaps per run	2.25
Crossings per run	10	Avg. duration of overlap (in time steps)	1.25
Total crossings	200	Avg. overlap percentage	3.67%
Total observed overlaps	45	Avg. max overlap	5.01%
Overlaps exceeding collision threshold	0	Max observed overlap	12.69%

There were no overlaps exceeding the collision threshold in the Unity runs; however, there was still approximately two overlaps observed per run of 3.6% on average. The majority of the relatively small overlaps are likely due to inconsistencies between the numerical precision used internally by AVOCA and that of the testing system and Unity. This will be investigated in future work.

The worst observed overlap in the runs was 12.69%, slightly over halfway to the collision threshold. Examining this run showed that AVOCA was struggling with a reciprocal dance situation, which was resolved before a collision occurred. This overlap was brief, lasting only a single time step, indicating that AVOCA reacted quickly to the escalating situation. There was only one other recorded overlap above 8%, implying that the reciprocal dance issue rarely occurs. Ongoing studies will investigate methods to mitigate the reciprocal dance earlier or avoid them altogether by making more conservative decisions before getting in near-collision states.

4.2 AURA Experimentation

Two studies were conducted using AURA: one with all virtual agents and the other with all physical agents. The experimentation scenario consisted of two AscTec Hummingbird quadrotors [12] and two AscTec Pelican quadrotors [13]. The scenario used the same agent orientation and goals as discussed in Sect. 4.1. The radius for the Hummingbird was set to 0.4 m and the radius for the Pelican was set to 0.5 m. These values are both 40% larger than the physical size of the quadrotors based on the maximum measured radius in the X-Y plane.

The KVC values used for AVOCA were held constant for the duration of the experiments conducted. We speculate that holding the KVC values constant gave AVOCA an incomplete understanding of the platform. This situation potentially resulted in more overlaps than if we had calculated KVC values in real time. Nevertheless, these experiments still establish a performance baseline for the algorithm in real world environments.

As data propagates through AURA, there is latency introduced that is unavoidable for the current system configuration. To investigate the impact of this latency, experiments were conducted with AVOCA using slightly increased avoidance regions, under the rationale that these increases would account for the uncertainty introduced by this latency. Three configurations were used in which the avoidance region was increased by 0%, 5%, and 10% inside AVOCA. The presented results use the baseline radii for computing overlap percentages. Each configuration was run 15 times with the virtual agent setup and 5 times with the physical agent setup, with each run consisting of 10 crossings; meaning that the total number of crossings considered per AVOCA configuration was 150 for the virtual agent setup and 50 for the physical agent setup. Based on the physical radii of the models, any overlap over 28.5% represents a collision. This rough approximation again assumes each quadrotor is a sphere. The results of the runs are summarized in Table 2.

Table 2. Summary of AURA experiments

		Average:			
All Virtual	AVOCA Increase %	Overlaps	Overlap %	Max Overlap %	Overlap Duration (sec)
	0	21.67	4.25	22.09	0.19
	5	3.27	7.83	14.61	0.30
	10	1.33	13.46	23.09	0.56
All Physical		Average:			
	AVOCA Increase %	Overlaps	Overlap %	Max Overlap %	Overlap Duration (sec)
	0	44.20	6.40	26.88	0.36
	5	20.80	7.70	47.37	0.36
	10	15.60	9.22	33.93	0.41

As expected, the average number of overlaps and average overlap percentages both increased relative to the experiments discussed in Sect. 4.1. The 0% increase configuration provides the most direct comparison to the experiments presented there. While there were many more overlaps seen in the AURA experiments, the average overlap percentages were still well below the collision threshold value of 28.5%, lasting less than half a second in the vast majority of cases. This indicates that even with the difficulties introduced by vehicle dynamics, physics, and other real world factors, AVOCA was able to successfully provide collision-free velocities in the vast majority of cases and quickly mitigate near-collision situations.

Initially, it may appear that AVOCA performed worse as the avoidance region was increased based on the metrics in Table 2. However, for the virtual agent experiments the average number of overlaps dropped by 85% with a 5% increase and by 94% with a

10% increase. Similarly, in the physical agent experiments average overlaps dropped by 53% in the 5% increase configuration and by 65% in the 10% configuration. We believe the reason the other metrics increased is that, while these configurations coarsely account for latencies in the system, they also remove sensitivity to numeric precision, data inaccuracies, and other issues that could result in relatively small overlaps. Essentially, the avoidance region increases distill out the instances where AVOCA truly struggled to provide collision free velocities, resulting in fewer overlaps overall, but the overlaps tended to be more severe.

None of the overlaps for the 0% increase configuration exceeded the collision threshold. The 5% increase configuration contained two overlaps (1.9% of the total overlaps) that exceeded the collision threshold, one of which resulted in an actual collision between the two Pelican quadrotors. During the five flights conducted for the 5% increase configuration, the overlaps resulting in collision threshold exceedance occurred during flights 1 and 3. There were six overlaps (7.7% of the total overlaps) that exceeded the collision threshold during the 10% increase flights; four occurred during flight 3 (one of which resulted in an actual collision), one occurred during flight 4, and the final occurred during flight 5.

After examining data from the flights containing overlaps that exceeded the collision threshold but did not result in a physical collision, it was determined that the severe overlaps all occurred in one of the upper/lower quadrants of the approximating sphere for the Pelicans or at the top/bottom for the Hummingbirds. Since the Hummingbirds are far wider than they are tall (almost 6.5 times wider), there is ample space above and below the physical vehicle to contain an overlap greater than the collision threshold. While the Pelicans are taller than the Hummingbirds, they are still far from spherical in shape with a width-to-height ratio of 2.7.

Considered overall, the results obtained from the AURA runs with the physical agents resulted in only two physical collisions out of 150 crossings, which indicates a 98.7% success rate for the tested scenario. Future work will include exploration of more complex scenarios including spatial changes as well as additional agents to test the robustness of AVOCA. These experiments will be conducted with real and virtual agents to quantify the percentage overlap of the bounding regions as well as actual (virtual and physical) collisions.

4.3 Future Work

As already mentioned, in its current configuration AURA introduces delays in the system (particularly in the communication network), which result in localization inaccuracies. Ongoing AURA work includes characterizing the system delays so that appropriate mediation strategies can be designed and implemented for each component in the system. Similarly, the use of uncertainty metrics in AVOCA is being examined to account for data delays and other similar issues.

Though rare, the severe overlaps observed are likely due to AVOCA being unable to resolve a reciprocal dance situation. These situations often have common characteristics. AVOCA could be expanded to detect the occurrence of these characteristics and, when detected, steps could be taken to avoid/resolve the situation earlier (e.g., falling back on a basic rules of the road or priority system).

During the tests, it was observed that Pelican platforms were unable to reach the commands issued by AVOCA in a given time step. The two reported collisions were between the slower and bigger Pelican platforms. The Hummingbirds are more agile and can respond quicker to the demands of AVOCA in order to avoid collisions for the tested scenario. Increasing the accuracy of AVOCA KVC data could potentially account for these issues. Future investigations will focus on understanding scenario demands and platform capability responses in relation to a given mission.

The use of spherical platform approximations can lead to misleading results, as discussed earlier. Future studies will include the use of more true-to-life approximation regions in both AVOCA and in the performance metrics used in AURA.

5 Conclusion

This paper introduced the AVOCA algorithm, which is being developed and tested within a Mixed Reality system. In addition, the AURA system was presented as an autonomy-testing framework, capable of incrementally exposing new algorithms/ systems to some of the real world idiosyncrasies that must be overcome by real-time systems. AURA provides the means to assess (with statistical relevance) the performance of new algorithms.

Current results indicate that AVOCA may provide the much-needed autonomous collision avoidance component enabling command and control of UV teams by a single operator to command and control teams of UVs in non-deterministic scenarios through only high-level commands. AVOCA may also provide the needed coordination for other swarm configurations to conduct autonomous missions.

Acknowledgements. Distribution A: Approved for public release: distribution unlimited.

References

1. Fiorini, P., Shillert, Z.: Motion planning in dynamic environments using velocity obstacles. Int. J. Robot. Res. **17**(7), 760–772 (1998)
2. Guy, S.J., Chhugani, J., Kim, C., Satish, N., Lin, M., Manocha, D., Dubey, P.: ClearPath: highly parallel collision avoidance for multi-agent simulation. In: ACM SIGGRAPH/ Eurographics Symposium on Computer Animation (SCA), August 2009
3. Snape, J., van den Berg, J., Guy, S.J., Manocha, D.: The hybrid reciprocal velocity obstacle. IEEE Trans. Robot. **27**(4), 696–706 (2011)
4. Hönig, W., Milanes, C., Scaria, L., Phan, T., Bolas, M., Ayanian, N.: Mixed reality for robotics. In: IEEE/RSJ International Conference Intelligent Robots and Systems, Hamburg, Germany, pp. 5382–5387 (2015)
5. Vicon motion systems ltd. http://www.vicon.com
6. Gazebo. http://gazebosim.com
7. Unity 3d. https://www.unity3d.com
8. Leap Motion. https://www.leapmotion.com

9. Oculus Rift Developer Kit 2. https://www.oculus.com/en-us/dk2
10. HoloLens. https://www.microsoft.com/microsoft-hololens/en-us
11. Robot Operating System. http://www.ros.org
12. AscTec Hummingbird. http://www.asctec.de/en/uav-uas-drones-rpas-roav/asctec-hummingbird
13. AscTec Pelican. http://www.asctec.de/en/uav-uas-drones-rpas-roav/asctec-pelican

Multimodality Evaluation Metrics for Human-Robot Interaction Needed: A Case Study in Immersive Telerobotics

Iina Aaltonen[1(✉)], Susanna Aromaa[2], Kaj Helin[2], and Ali Muhammad[2]

[1] VTT Technical Research Centre of Finland Ltd., Vuorimiehentie 3,
P.O. Box 1000, 02044 Espoo, Finland
{iina.aaltonen,susanna.aromaa,kaj.helin,ali.muhammad}@vtt.fi
[2] VTT Technical Research Centre of Finland Ltd., Tekniikankatu 1,
P.O. Box 1300, 33101 Tampere, Finland

Abstract. Multimodal, wearable technologies have the potential to enable a completely immersive teleoperation experience, which can be beneficial for a number of teleoperated robotic applications. To gain the full benefit of these technologies, understanding the user perspective of human-robot interaction (HRI) is of special relevance for highly advanced telerobotic systems in the future. In telerobotics research, however, the complex nature of multimodal interaction has not attracted much attention. We studied HRI with a wearable multimodal control system used for teleoperating a mobile robot, and recognized a need for evaluation metrics for multimodality. In the case study, questionnaires, interviews, observations and video analysis were used to evaluate usability, ergonomics, immersion, and the nature of multimodal interaction. Although the technical setup was challenging, our findings provide insights to the design and evaluation of user interaction of future immersive teleoperation systems. We propose new HRI evaluation metrics: Type of multimodal interaction and Wearability.

Keywords: Human-robot interaction · Metrics · Multimodal · Wearable · Telerobotics · Immersion · User studies

1 Introduction

Immersive telerobotics, where the user can experience being present at the site of a teleoperated robot, has great potential in many domains, for example, in operating planetary rovers [1] and other tasks in the space [2], mining [3], nuclear power plants [4], high-pressure ocean missions [5], and robotic surgery [6]. From the user perspective, the combination of a teleoperated robot and wearable multimodal interfaces is fascinating. When users are required to interact with the environment via a robot using these interfaces, it is no longer trivial to evaluate and understand the nature and quality of human-robot interaction (HRI).

Many teleoperated systems with wearable, multimodal interfaces have been developed, but most often the user experience is considered very briefly and comprehensive user evaluation methods have not been utilized. Using quantitative performance metrics

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_32

[7] or established questionnaires (e.g., [8, 9]) alone is not enough to account for the nuances of multimodal interaction and the effects of virtual displays. Qualitative methods, such as interviews and observations, are needed to capture the user experiences and the way the multimodal interfaces are actually used.

In this paper, we suggest two evaluation metrics that should be considered in designing and evaluating HRI of immersive telerobotics systems: Type of multimodal interaction and Wearability. The need for the metrics was recognized in a case study, which is also reported in this paper. Based on existing literature and our findings, we also suggest methods for evaluating them. In Sect. 2, the human aspects of multimodal interaction and existing evaluation methods, and related research in telerobotics, are introduced. Section 3 describes the case study, the used evaluation procedure and results. In Sect. 4, the results are discussed and metrics for evaluating wearable and multimodal interfaces are proposed.

2 Related Work

2.1 Evaluation of Multimodal Interaction

In general, in multimodal interaction, the user interacts with a system using two or more modalities, which can refer to sensory modalities (e.g., visual, auditory, [10]) or input modes (e.g., speech, touch, gesture, [11]). Multimodal displays, and also controls, have been suggested for HRI to decrease task difficulty, promote sense of immersion, and mitigate operator workload (see [12]). Several evaluation methods for multimodal interaction have been used in the human-computer interaction domain.

PROMISE [13] is a framework for multimodal dialogue system evaluation which includes several quality and quantity measures. Although PROMISE was developed for multimodal dialogue systems, some of the measures are applicable to HRI, such as user/system turns and semantics. SUXES [14] is a method for capturing and comparing both user expectations and user experiences. The statements can be used for evaluating the overall system and for different input and output modalities. Kühnel et al. [15] compared available usability questionnaires and found that methods AttrakDiff, SUS, and USE are suitable for the usability evaluation of systems with multimodal interfaces, but the selection of questionnaire depends on the purpose of the evaluation. Ramsay et al. [16] evaluated a multimodal mobile phone map application using a variety of methods: log data, field notes (recordings and observations), and interviews. Wechsung [17] has developed a taxonomy for describing multimodal quality aspects of interaction and designed a psychometrically validated MultiModal Quality Questionnaire (MMQQ); The basis was in questionnaires designed for unimodal systems, which were found inapplicable for usability evaluation of multimodal systems.

2.2 User Evaluation of HRI

The metrics for evaluating HRI have been divided into human, robot, and system components [18]. The human component includes seven items: accuracy of mental models, degree of mental computation, human reliability, productive time vs. overhead time, situation awareness, trust and workload. In general, five primary methods for user evaluations in HRI have been suggested: self-assessments, interviews, behavioural and psychophysiology measures, and task performance measures (e.g., time to complete a task) [19]. The use of three or more methods is recommended.

2.3 Wearable and Multimodal Interfaces in Telerobotics

A number of studies mention user experiments of telerobotic systems with similar properties to ours: wearable and multimodal control of a field robot [20, 21]; multimodal control of mobile robots [22, 23]; a haptically controlled robot [24, 25], also in robotic surgery [6]; wearable [26, 27], tangible [28] and traditional [29] user interfaces; and head-mounted displays (HMDs) [30, 31]. Typically, the experiments have involved quantitative performance evaluations, whereas user-related measures are mentioned very briefly and methodological details are often omitted. None of the papers found have elaborated on the multimodal interaction from the user perspective. The following user experiments provide, however, a representative sample of diverse methods used in the evaluation of the human component of HRI.

Kechavarzi et al. [29] evaluated three user interfaces for a teleoperated mobile robot. The methods used were a survey of participants' perceptions of robots and immersive tendencies; performance measures (e.g., time to perform); questionnaires (e.g., satisfaction, immersion, intuitiveness, comfort); and an interview. Fernandes et al. [26] tested three interfaces, including a wearable arm-mounted one, for operating a robotic manipulator in a pick-and-place task. Both objective (e.g., outcome of task) and subjective (a survey concerning ease of use) measures were used. Zareinia et al. [6] tested three haptic hand-controllers in a robot-assisted surgical system. Ten performance measures (e.g., operator effort) and a questionnaire (e.g., easiness to learn and use the system, and comfort) were used. Livatino et al. [31] evaluated different screen and display types, including an HMD, in a virtual medical endoscopic teleoperation task. Both quantitative (collision rate etc.), and qualitative variables (questionnaire, e.g., presence and comfort) were used.

3 Case Study

3.1 Setup

Robot System. The robot included three main components: (1) a four-wheel drive remote controlled car, operated by a 12-volt battery, provided forward, backward and turning manoeuvrings, (2) a robotic arm (Lynx al5d) included three main links and actuator (effector), and (3) a non-stereoscopic pan tilt camera (Tenvis JPT3815W), mounted on an aluminium construction 40 cm over the rover (Fig. 1 – Left). Similar systems have been described in [20, 21].

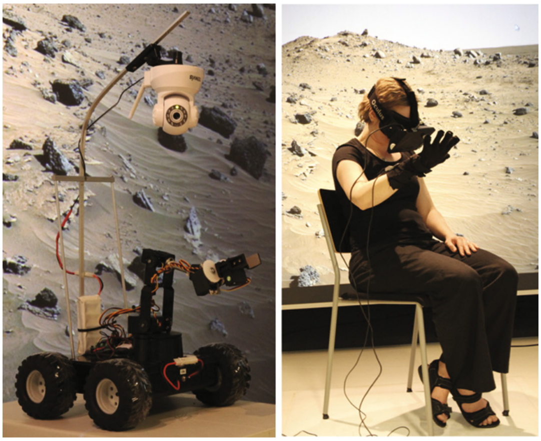


Fig. 1. Left: Robot system. Right: User with HMD and data glove in a mixed reality laboratory. The hand posture is IDLE.

Gesture-Based Control System. The control system was composed of four subsystems. (1) A main computer connected with wireless adapter (Xbee), communicating with the rover. This computer was also connected with (2) an HMD (Oculus Rift DK1) which provided mono-camera feed to users. The tracking position of the HMD was used for the pan tilt camera control. (3) The user also wore a right-handed data glove (5DT-5 Ultra) and (4) the arm of the user was constantly tracked with a Kinect depth camera (Fig. 1 – Right).

The user interacted with the robot system in two ways. First, by tilting the HMD, the user controlled the camera placed on the rover and received visual feed of the robot's surroundings. The same image was displayed for both eyes (Fig. 2 – Left). Second, using the arm with the data glove, the user drove the rover and controlled the robot arm and the gripper. Only one control mode could be active at the time.

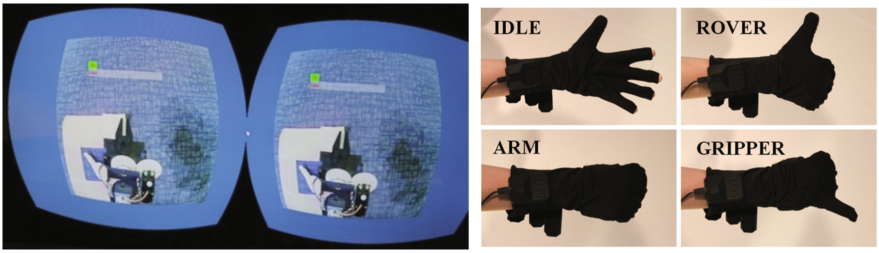


Fig. 2. Left: The HMD view looking down at the gripper and the target. The mode is IDLE, and the tracking is working (*indicated by the green square in the upper left corner*). Right: The data glove and the hand postures for switching the control modes.

The control mode was switched using the glove with different hand postures (Fig. 2 – Right). The modes (postures) were: ROVER (a fist, thumb pointing left), robot ARM (thumb under the fist), and GRIPPER (thumb under fist, 1–4 fingers outstretched, e.g., pinkie, depending on the user). In addition, holding all fingers outstretched put the system to IDLE mode. In the upper left corner of the HMD view, the user was shown which mode was active. A red or green rectangle also indicated if the system could track the user's arm position.

The user's arm position controlled the robot's action. In the ROVER mode, the robot could be moved forward (arm away from torso), backward (towards shoulder), or turn (to either side). In the ARM mode, the position of the user's hand defined the absolute position of the robot's gripper. For example, if the users moved their fist up, the robot lifted its arm/gripper to a corresponding position. In the GRIPPER mode, moving the hand to the right (left) would close (open) the gripper.

3.2 Test Procedure

Participants. Nine volunteers (5 males and 4 females) participated in the final user test, preceded by a debugging session with three testers. The participants were aged 29–57 (average 36) and they were all right-handed. They were recruited via a research organization. Three participants reported using virtual reality technology frequently and four rarely. Two did not have any experience of 3D technologies; others had seen 3D movies or played games. In addition, three participants had prior experience of telerobotics using a haptic control device and a virtual display, and one had teleoperated a farming crane using joysticks.

Task. The task was to teleoperate a robot on Mars to collect a sample ("a rock") while being seated in an orbiter (cf. [32]). The test took place in a mixed reality laboratory where the participant was seated on a chair next to the robot arena. Prior to test, the participant could see the robot's surroundings and practice operating the robot also without mounting the HMD. There were five phases in the task: (1) Drive the robot next to the rock. (2) Move the robot arm next to the rock. (3) Pick up the rock with the gripper. (4) Drive the robot next to a box. (5) Put the rock into the box.

Data Collection. Prior to the test, each participant filled a consent form, a background information form, and two questionnaires: simulation sickness questionnaire (SSQ) and a bodymap. All questionnaires are described in the next subsection. A researcher took notes on how the training phase (duration 10–16 min) went, noting any difficulties in training or technical adjustments.

During the test, the participant's and robot's performance was videoed from three different angles, a side and front view of the participant, and an overall view showing the robot, the rock, the box and the participant in the background; additionally, the view from the HMD was saved. Researchers filled an observation form for each participant and took notes on performance, timing, technical issues, use of control modes, ergonomics, and participants' effort and frustration. Furthermore, researchers noted instructions and other help that were given. The test time was limited to 15 min.

After the test, the participant answered to SSQ, NASA-TLX, bodymap and usability questionnaires, and was interviewed. The audio-recorded interviews were structured to cover themes of training, task performance, user interfaces, and the control concept. In addition to the user evaluations, two human factors researchers used the robot system and made a heuristic evaluation based on Nielsen's heuristics [33].

Questionnaires. An adapted SSQ [9] using 30 items was used to collect simulation sickness symptoms. The experience of discomfort in a certain area of the body was assessed using a bodymap of the upper body (7-point scale: no discomfort–severe discomfort). An unweighted NASA-TLX [8] was used to collect the experience of subjective workload. The usability questionnaire consisted of 40 statements (5-point Likert scale: completely disagree–completely agree). The statements were formulated using a combination of several approaches: systems usability framework [34], Multi-criteria Assessment of Usability for Virtual Environments (MAUVE) system [35] (especially statements originally by R.S. Kalawsky, J.L. Gabbard and D. Hix), and usability factors and goals [36]. The final 40 statements concerned wayfinding, navigation, object selection and manipulation, visual output, presence, immersion, comfort, aftereffects, and the operating concept in general.

Data Analysis. Basic descriptive statistics were calculated from the questionnaire data. The interview answers were grouped regarding pros, cons and improvements of the interfaces as well as any comments on the control modes and multimodal interaction. The three video views and the HMD view were replayed synchronously, and a video analysis was performed. Each test was watched 1–3 times, and the following aspects were noted during each task phase: general description of the nature of interaction, simultaneous use of the HMD and the glove, keeping gaze on target in the ROVER mode, ergonomics issues, mistakes, errors and help given. These findings were combined with the written notes.

3.3 Results

Usability, Ergonomics and Immersion. Based on the heuristic evaluation, the control system was not stable and mature enough to provide a required usability level for users, and therefore most questionnaire results are not reported in detail. In general, the gesture-based control system was natural to use but too insensitive and not always responding to gestures. Identified issues regarding the visual display were image lag and narrow field of view (FOV). Furthermore, the camera image drifted with abrupt head movements, and it had to be reset on several occasions.

Five participants successfully completed all five phases of the task. In three cases, technical problems affected the rover control and some test phases were skipped; and one participant ran out of time. A researcher helped each participant during the test, e.g., by giving verbal instructions to move the robot to a better position if the participants lost their sense of orientation. Three participants mentioned more training would have been useful, for example, to better estimate the mapping between their arm position and the robot speed. The NASA-TLX results and interviews indicated that the participants experienced the test as frustrating and mentally demanding, because the system missed their gestures. Compared with others, the three participants with prior experience of telerobotics had evaluated the task less demanding on all accounts.

The best benefits of the wearable control were mentioned to be the feeling of presence and immersion in the task, and that the wearable interface would be a natural way to operate. Most participants thought they could act naturally with the wearable devices, supported by the questionnaire data ("The HMD did not feel clumsy to wear.", mean and standard deviation 4.1 ± 1.1, scale 1–5 disagree–agree; and "The glove allowed me to move my hand naturally.", 4.3 ± 1.0) and user comments. There were, however, several comments on the hand postures, mostly about difficulties in remembering the control modes but also some on the formation of the posture. The mode changing in general seemed to work better when the IDLE mode was activated between the active modes. Table 1 lists the pros and cons of the devices that were brought up in the interviews.

Table 1. Users' evaluations of wearable devices.

Device	Pros	Cons
HMD	Comfortable, not heavy Nice fit Realistic transmission delay Adequate resolution Clear view to surroundings Camera orientation with respect to body	Neck pain due to posture Not securely attached Long delay Poor image quality Small FOV, perspective Depth vision missing Drift
Data glove	Light-weight, soft Easy to move with Operating without a medium	Sweaty Loose fit No feedback

According to SSQ, there was a minor increase in general discomfort due to the HMD, but otherwise no negative symptoms were mentioned. Regarding ergonomics measured using the bodymap, the main finding was that the participants felt more discomfort in their right shoulder after performing the test (discussed in more detail in S. Aromaa et al. (in review)). Some participants mentioned the required arm trajectory was too wide and therefore uncomfortable. Furthermore, there were comments on awkward postures when the participant's head was "in the right armpit" and the right arm was stretched to upper left. Our observations support these comments.

The participants suggested many improvements. Several comments were made on improving the perspective and FOV of the HMD to show the robot arm at all times. Stereo image, or alternately a depth indicator and the ability to zoom were also wanted. An indicator was also suggested for showing the position of the robot arm. Furthermore, the image should follow gaze more smoothly and with less delay. There were suggestions to include haptic feedback to the glove, especially for object manipulation. The sensitivity of the glove should also be improved. Suggestions were also made to make the required arm movements smaller, and provide elbow support and a physical "knob" to hold onto for position estimation. Regarding the changing of the control modes, one suggested that the left hand could be used for that purpose, or the glove be replaced by a keyboard. Another participant suggested using a joystick for driving the rover and using the glove for the robot arm. For operating in the GRIPPER mode, a pinch-like hand movement was suggested.

Multimodal Interaction. Multimodal interaction was assessed mostly based on the video material. Although individual aspects of the robot control were considered difficult due to technical issues (and thus not reported further), five of nine participants—including all participants with prior teleoperation experience—agreed the concept of operation was logical ("The system was operated in a way I would expect it to be operated.", mean and s.d. 3.2 ± 1.3).

The HMD and the hand gestures were used simultaneously by all participants in the ROVER mode. The most common working strategy was to keep the target (the rock or the box) in the visual field while driving straight forward. If the target was located slightly toward either side, however, the participant did not always realize that the robot heading was not that shown in the centre of the HMD, and the robot passed by the target. One participant was noticed to manoeuvre the rover while turning and simultaneously keeping the target in the view.

We could also observe that some participants turned their head to the direction of their extended arm; either slightly toward a side when turning, or up and down when accelerating and decelerating. When the robot was mobile, the head movements were small and slow with the exception of one participant who moved his head with bigger, jerkier movements. Bigger, searching head movements were clearly done in the IDLE mode. In the ARM and the GRIPPER modes, the HMD view was hardly altered. Typically, the HMD was moved 1–2 times to get a new viewing angle while the arm was kept still. Similar to the ROVER mode, the participant's head tended to follow the arm, but the movements were very small.

4 Discussion

The tests showed that the concept of the multimodal control system was workable and the participants could, despite short training time, teleoperate the robot, although the overall usability could not be evaluated due to low maturity level. Both the HMD and the data glove were felt comfortable for the most part, which could be expected as they are commercial products, and the participants could move naturally while wearing them. However, uncomfortable body postures (especially those related to using both devices) were observed and reported by the participants.

The participants intuitively used the HMD and the hand gestures simultaneously, even though they had not been specifically instructed to do so—which could have caused bias towards using the modalities in a certain way [37]. Regarding the simultaneous use, more training, and perhaps an instructive video, might be useful in making operators aware of typical human actions. For example, when you are learning to drive a car, you need to put a conscious effort not to turn the steering wheel when you check over your shoulder. A similar coupling was observed in the user study. Likewise, it seemed easy to forget that the view on the HMD did not necessarily show the heading direction of the robot—many times the robot or its parts were not visible in the HMD. Adding a heading indicator on the visual display could be helpful.

Originally, the research focus of the case study was on developing the technology of the robot system, and therefore the selection of gestures was done in a very late stage. Ideally, the gestures would have been iteratively designed and tested with users; our participants made valid suggestions for improvement, which can be accounted for in the future. Many of the questionnaire responses reflected the users' frustration on the technical issues, and therefore the interviews and videos proved very valuable in evaluating the user interaction. In the future, the questionnaire and interview items should probe deeper into the nature of multimodal interaction, and some of the user tests could be performed using a simulated robot to overcome the problems related to the robot technology—preferably as a part of iterative design of the user interfaces.

The benefit of doing the case study—regardless of the technical difficulties—was the realization that the multitude of questionnaires and interviews did not cover the multimodal interaction, which was observable in the videos. Furthermore, the human aspects of multimodal interaction are also neglected in the HRI literature, with the exception of multimodal dialogue research. We think multimodal interaction should be studied more rigorously because it will affect the human performance. In addition, we also feel that the wearability of different devices is essential in immersive telerobotics and suggest it should also be used as a HRI metric.

4.1 Suggested Metrics for Wearable and Multimodal Systems

Some of the existing HRI metrics [7] can be useful in evaluating the human aspects of multimodal devices, for example, "Accuracy of mental models of device operation" and "Degree of mental computation". In addition, mental and physical workload and situation awareness apply to multimodal interaction in any domain, and also to wearable interfaces. For ensuring good HRI in telerobotics with wearable and multimodal interfaces, two new metrics are suggested: Type of multimodal interaction and Wearability.

Type of Multimodal Interaction. By this metric we mean to cover the multimodal interaction that (1) the user engages in and (2) the system is capable of. This is close to the quality measure "ways of interaction 'n-way communication (several modalities possible at the same time?)'" used in PROMISE [13]. In addition, four categories (Exclusive, Alternate, Concurrent, Synergistic) have been used for describing multimodal interaction along two axes: use of modalities (sequential or parallel) and data fusion of different modalities (combined or independent) [38]. These categories can be useful in describing interaction for both the system and the user.

More importantly, user tests are needed to evaluate if the users can and will use all the multimodal capabilities the system offers, and how. In practice, we suggest using several methods to evaluate the quality of multimodal interaction. First of all, the experimental task should be designed so that there are possibilities for using the modalities individually and in parallel to facilitate evaluation.

To tackle the multimodality and parallel use, questionnaire statements such as those introduced in MMQQ [17], could be used, e.g., "The different input modalities are blocking/complementing each other.” In addition, evaluating single modalities on their ease of use and learnability is important, as well as the logic on the system level, e.g., “The system was operated in a way I would expect it to be operated.”, and “The way the system was operated was convincing and suits professional use.” [34].

In interviews, these issues can be elaborated further: what aspects were easy or difficult; how natural did the users experience the combination of modalities; did the users have a conscious strategy to use the modalities sequentially or in parallel, and how did this strategy evolve. Furthermore, observations—preferably complemented with videos—are needed to evaluate how the users actually used the modalities, e.g., preferences, disuse or mistakes, and changes in behaviour or speed. Some information can also be deduced from performance measures and log files, if available.

Finally, when evaluating multimodal interaction involving multiple sensory modalities, the psychological effects related to multimodal processing need to be considered [11, 39]. The evaluation gets more complicated when the system output (feedback to user) is multimodal, and it cannot be directly observed which modalities affected the user’s actions—one possible solution is to test the system using combinations of the available modalities.

Wearability. In the telerobotics context, we consider wearability, or “the interaction between the human body and the wearable object” [40], to be characterized by comfort, ergonomics, freedom of movement, and intuitiveness of learning and using the system. Intuitive control, especially when the designed physical representation feels natural, has been associated with improved performance [6, 28], and is closely related to immersion [29]. Intuitiveness is also indispensable if the users have very little training or when attention cannot be allocated to secondary tasks (e.g., [20]).

Regarding HMDs, wearability also involves issues related to virtual displays in general, such as simulation sickness, immersion, situation awareness, sense of direction, and quality of display (2D vs. 3D, resolution, delay; e.g., [29, 30]). In addition, when combined with other wearable devices, it is important to observe if the users adopt awkward body postures without noticing.

In practice, many of the wearability aspects can be measured using customized usability questionnaires such as those used in the case study. The guidelines [40] and methods [41] for wearable computers can be used as well. Additionally, performance measures (e.g., training time) can be useful in determining the intuitiveness of use. User comments and observations are needed to complement the quantitative data.

4.2 Conclusion

In telerobotics research, the human aspects of multimodal interaction have not attracted much attention. We studied human-robot interaction in the context of teleoperating a mobile robot using a wearable multimodal control system. In the case study, we noticed the complex nature of the multimodal interaction and realized there is a need for user evaluation metrics for immersive telerobotics. Two metrics were introduced: Type of multimodal interaction and Wearability, along with methods for measuring them. In future work, the metrics and methods should be researched further. The metrics can help in the design and evaluation of HRI in immersive telerobotics, and also in other teleoperation tasks such as in crane operation and mining.

Acknowledgements. We thank test participants. Research funded under the European Commission's Seventh Framework Aeronautics and Air Transport (AAT) programme in the project VR-HYPERSPACE (AAT-285681) 'The innovative use of virtual and mixed reality to increase human comfort by changing the perception of self and space'. The work of I. Aaltonen was supported by Wihuri Foundation.

References

1. Bualat, M. et al.: Surface telerobotics: development and testing of a crew controlled planetary rover system. In: Proceedings of AIAA SPACE Conference and Exposition, pp. 1–10, San Diego, California, USA (2013)
2. Brooks, T.L.: Operator vision aids for telerobotic assembly and servicing in space. In: Proceedings of the 1992 IEEE International Conference on Robotics and Automation, pp. 886–891. Nice, France (1992)
3. Varadarajan, K.M., Vincze, M.: Augmented virtuality based immersive telepresence for control of mining robots. In: Proceedings of ISCIII 2011 - 5th International Symposium on Computational Intelligence and Intelligent Informatics, pp. 133–138. Floriana, Malta (2011)
4. Eickelpasch, N., et al.: Remote techniques for the underwater dismantling of reactor internals at the nuclear power plant Gundremmingen unit A. Nucl. Energy **36**(1), 49–54 (1997)
5. Yuh, J.: Design and control of autonomous underwater robots: a survey. Auton. Robots **8**, 7–24 (2000)
6. Zareinia, K., et al.: Performance evaluation of haptic hand-controllers in a robot-assisted surgical system. Int. J. Med. Robot. Comput. Assist. Surg. **11**, 486–501 (2015)
7. Steinfeld, A. et al.: Common metrics for human-robot interaction. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI 2006), pp. 33–40. Salt Lake City, Utah, USA (2006)
8. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. **52**(C), 139–183 (1988)
9. Kennedy, R.S., et al.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**(3), 203–220 (1993)
10. Möller, S. et al.: A taxonomy of quality of service and quality of experience of multimodal human-machine interaction. In: IEEE International Workshop on Quality of Multimedia Experience (QoMEx 2009), pp. 7–12. IEEE, San Diego, California, USA (2009)
11. Dumas, B., et al.: Multimodal interfaces: a survey of principles, models and frameworks. In: Lalanne, D., Kohlas, J. (eds.) Human Machine Interaction, LNCS, vol. 5440, pp. 3–26. Springer, Heidelberg (2009)
12. Chen, J.Y.C., et al.: Human performance issues and user interface design for teleoperated robots. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**(6), 1231–1245 (2007)
13. Beringer, N. et al.: PROMISE – a procedure for multimodal interactive system evaluation. In: Proceedings of LREC Workshop on Multimodal Resources and Multimodal Systems Evaluation, pp. 77–80. Las Palmas, Canary Islands, Spain (2002)
14. Turunen, M. et al.: SUXES—user experience evaluation method for spoken and multimodal interaction. In: Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH), pp. 2567–2570. Brighton, UK (2009)

15. Kühnel, C. et al.: Evaluating multimodal systems. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI'10, pp. 286–294. Reykjavik, Iceland (2010)
16. Ramsay, A., et al.: Tilt and go: exploring multimodal mobile maps in the field. J. Multimodal User Interfaces **3**(3), 167–177 (2010)
17. Wechsung, I.: An Evaluation Framework for Multimodal Interaction: Determining Quality Aspects and Modality Choice. Springer, Cham (2014)
18. Murphy, R.R., Schreckenghost, D.: Survey of metrics for human-robot interaction. In: ACM/IEEE International Conference on Human-Robot Interaction, pp. 197–198. IEEE (2013)
19. Bethel, C.L., Murphy, R.R.: Review of human studies methods in HRI and recommendations. Int. J. Soc. Robot. **2**(4), 347–359 (2010)
20. Ryu, D. et al.: Wearable haptic-based multi-modal teleoperation of field mobile manipulator for explosive ordnance disposal. In: Proceedings of the 2005 IEEE International Workshop on Safety, Security and Rescue Robotics, pp. 75–80. IEEE, Kobe, Japan (2005)
21. Jankowski, J., Grabowski, A.: Usability evaluation of VR interface for mobile robot teleoperation. Int. J. Hum. Comput. Interact. **31**(12), 882–889 (2015)
22. Yang, H.S. et al.: Wearable computing based on multimodal communication for effective teleoperation with humanoids. In: Proceedings of the 14th International Conference on Artificial Reality and Telexistence (ICAT 2004). Seoul, Korea (2004)
23. Brice, B., et al.: Towards multimodal interface for interactive robots: challenges and robotic systems description. In: Abdellatif, H. (ed.) Robotics 2010 Current and Future Challenges, pp. 369–380. InTech, Croatia (2010)
24. Horan, B. et al.: Multi-point multi-hand haptic teleoperation of a mobile robot. In: Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, pp. 1112–1118. IEEE, Toyama, Japan (2009)
25. Pham, C.D., et al.: Evaluation of subjective and objective performance metrics for haptically controlled robotic systems. Model. Identif. Control **35**(3), 147–157 (2014)
26. Fernandes, V.B.P. et al.: A wearable interface for intuitive control of robotic manipulators without user training. In: Proceedings of the ASME 2014 12th Biennial Conference on Engineering Systems Design and Analysis, ESDA2014. American Society of Mechanical Engineers, Copenhagen, Denmark (2014)
27. Boudoin, P. et al.: Towards multimodal human-robot interaction in large scale virtual environment. In: Proceedings of the 3rd International Conference on Human Robot Interaction - HRI 2008, p. 359 (2008)
28. Randelli, G. et al.: Evaluating tangible paradigms for ground robot teleoperation. In: Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication - 2011 RO-MAN, pp. 389–394. IEEE, Atlanta, GA, USA (2011)
29. Kechavarzi, B.D. et al.: Evaluation of control factors affecting the operator's immersion and performance in robotic teleoperation. In: Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (2012 RO-MAN), pp. 608–613. IEEE, Paris, France (2012)
30. Martins, H. et al.: Design and evaluation of a head-mounted display for immersive 3D teleoperation of field robots. Robotica **33**(10), 2166–2185 (2014)
31. Livatino, S., et al.: Stereoscopic visualization and 3-D technologies in medical endoscopic teleoperation. IEEE Trans. Ind. Electron. **62**(1), 525–535 (2015)
32. METERON (Multi-Purpose End-To-End Robotic Operation Network) project. http://esa-telerobotics.net/meteron
33. Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods, pp. 25–62. John Wiley & Sons, Inc., New York (1994)

34. Savioja, P., Norros, L.: Systems usability framework for evaluating tools in safety–critical work. Cogn. Technol. Work **15**(3), 255–275 (2013)
35. Stanney, K.M., et al.: Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. Int. J. Hum Comput Stud. **58**(4), 447–481 (2003)
36. Kalawsky, R.S., et al.: Human factors evaluation techniques to aid understanding of virtual interfaces. BT Technol. J. **17**, 1 (1999)
37. Lisowska, A. et al.: Minimizing modality bias when exploring input preferences for multimodal systems in new domains: the archivus case study. In: CHI 2007 Extended Abstracts on Human Factors in Computing Systems, pp. 1805–1810. ACM, New York, NY, USA, San Jose, CA, USA (2007)
38. Nigay, L., Coutaz, J.: A design space for multimodal systems: concurrent processing and data fusion. In: Proceedings of the INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems (CHI 1993), pp. 172–178. ACM, Amsterdam, Netherlands (1993)
39. Wickens, C.D.: Multiple resources and mental workload. Hum. Factors **50**(3), 449–455 (2008)
40. Gemperle, F. et al.: Design for wearability. In: Proceedings of Second International Symposium on Wearable Computers. Digest of Papers, pp. 116–122. IEEE, Pittsburgh, PA, USA (1998)
41. Knight, J.F. et al.: Assessing the wearability of wearable computers. In: Proceedings of the 10th IEEE International Symposium on Wearable Computers, pp. 75–82. IEEE, Montreux, Switzerland (2006)

VR Environment for the Study of Collocated Interaction Between Small UAVs and Humans

Christopher Widdowson[1(✉)], Hyung-Jin Yoon[2], Venanzio Cichella[2], Ranxiao Frances Wang[1], and Naira Hovakimyan[2]

[1] Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820, USA
widdwsn2@illinois.edu

[2] Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 1206 W. Green Street, Urbana, IL 61801, USA

Abstract. Two issues that are crucial to the integration of flying robotic systems into human populated environments include: how humans perceive autonomous flying robots, and how to design and control flying robots to improve the level of comfort and perceived safety for collocated others. This work represents a comprehensive virtual reality test environment to explore scripted and unscripted interactions with flying robots. We employ a multimethod approach by incorporating behavioral measures, self-report questionnaires, and physiological data to characterize human arousal during a variety of predetermined and real-time scenarios in both indoor and outdoor environments. By combining complementary methodological techniques, we can converge on a data-driven model of social etiquette for flying robots; this model can then be reparametrized in terms of planning and control solutions to govern the robot's behavior in a real-world context.

Keywords: Virtual reality · Human-robot interaction · Deep learning · Electrodermal activity · Quadrotor robot

1 Introduction

The co-existence of autonomous vehicles and humans is the vision of a not-so-distant future, in which ground robots and flying machines populate our roads and skies. Some of the major obstacles to the integration of autonomous vehicles into human populated environments include: the complexity of the environment, interactions between robotic systems and human activities, and the unpredictable nature of human beings. For this reason, our research group has begun developing solutions to the motion planning and control problems for multiple autonomous robots that consider mission objectives, data about the environment, as well as human behavior and perception.

To do this, we employ an interdisciplinary approach to create guidance and control algorithms that improve the acceptability of aerial robots in essential social environments, by identifying factors that influence perceived safety and comfort. This involves

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1_33

assessments of human behavior around flying robots while varying features of the robot, e.g. velocity, acceleration, angle of approach, and time to collision (TTC).

To do this in a safe, low-cost, and time efficient manner, we developed a scalable virtual reality (VR) test environment to explore human-robot interaction for a variety of experimental paradigms. The experimental setup facilitates the study of human perception of non-humanoid virtual robots in controlled social environments, and provides a platform for graduate education and research to supplement less accessible research infrastructure, such as flight arenas. The present article describes the VR test environment and data acquisition toolchain we have implemented, and discusses plans for future research.

2 Methods

The current work characterizes human arousal as a logical antecedent to perceived safety during human-robot interactions in VR. As such, perceived safety is operationalized in terms of multiple outcome variables, including electrodermal activity (EDA), photoplethysmography (PPG), and head motion; however, the system can be expanded to include additional sensing. Aspects of the user experience are also assessed via self-report measures, such as simulator sickness, presence, and demographic information.

2.1 Apparatus and Stimuli

Head-Mounted Display (HMD). The HTC Vive VR headset is used to display the virtual environment. The HMD consists of two low-persistence AMOLED displays (90 Hz refresh rate) with a combined resolution of 2160 × 1200 (1080 × 1200 per eye) and approximately 110° horizontal field of view. The system achieves 6DoF (360°) head-tracking by fusing sensor data from a pair of infrared laser emitters - positioned diagonally and in opposite corners of a 3 m × 3 m × 3 m tracking volume - with onboard IMU and laser position sensors embedded in the headset.

Virtual Environment (VE). The VE is generated in the Unity game engine (Unity 5.4.1f1) and presented on a Windows 10 computer (i7-5820K, 3.3 GHz CPU; 32 GB RAM; NVIDIA GeForce GTX 980 Ti graphics card) at approximately 90 frames per second. Spatial ambisonics are simulated in-engine using a head-related transfer function (HRTF) that modulates several sound sources to complement the virtual scene, e.g. ambient wind, crickets, freeway overpass, etc. Distance modeling for sound sources is simulated using a six-sided rectangular volume - in which the user is located at the center - to generate early reflections and late reverberations for local scene geometry.

The simulation is capable of reproducing quadrotor flight behavior at both high and low levels of precision, to meet the needs of a given experimental scenario. For high precision simulations, quadrotor flight dynamics are modeled on a separate machine using Simulink (MathWorks), and streamed real-time via UDP to control the motion of a virtual quadrotor. Simulink receives the position data from the Unity engine and calculates motor outputs that will correct the position of the quadrotor. Simulink uses these motor commands to simulate the dynamics of the vehicle and sends the velocities and orientation to Unity. For low precision simulations, flight behavior is modeled in-engine using a path following technique based on Catmull-Rom spline interpolation. In this case, flight dynamics and attitude control are simulated by injecting sinusoidal noise into the rigid body parameters of the quadrotor (to mimic the bob and wobble effects of lift), and by programmatically adjusting pitch and roll based on the acceleration and velocity profile of the UAV.

Different aspects of the VE are recorded concurrently throughout each experiment and written to a file as output variables for subsequent analysis. Common output variables include UAV position $[x, y, z]^{\mathrm{T}}$, UAV velocity $[v_x, v_{y,} v_z]^{\mathrm{T}}$, user head position, distance to user, and visibility estimates. Here, visibility is determined by calculating whether a target object (e.g. quadrotor) is occluded by an obstacle and whether the target object is within the user's field of view. The occlusion check is performed by raycasting from the user's camera forward vector to the target object's position and checking for intersections with that ray. The field of view check is calculated by determining whether a target object is inside (or intersects with) a plane array that defines the geometric bounds of the camera view frustum. If the object is not occluded, and is within the field of view, then the object is visible (Fig. 1).

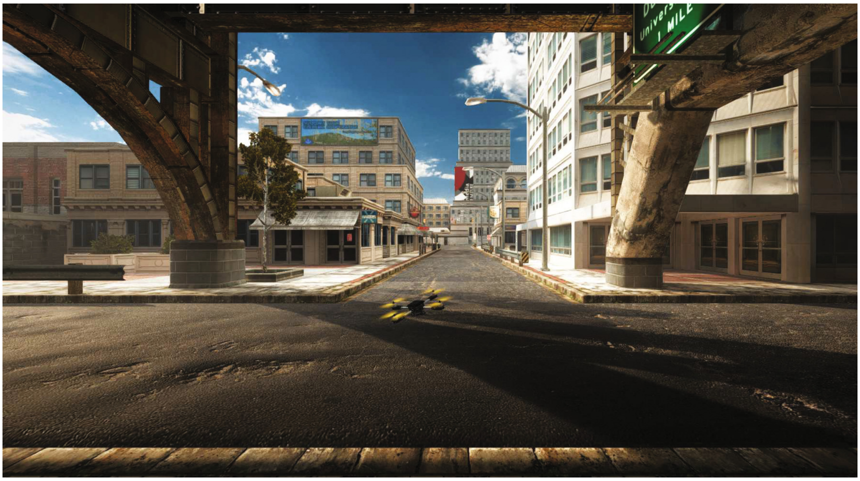


Fig. 1. Quadrotor robot flight in the virtual environment.

Physiological Recordings. EDA and PPG data are recorded using the Shimmer3 Wireless GSR+ unit developed by Shimmer Sensing. The device consists of a 16-bit (24 MHz) ultra-low power microcontroller and Class 2 Bluetooth radio (10 m; 33 ft), capable of resolving skin resistance levels from 10 k to 4.7 M (100 uS to 0.2 uS) for a frequency range of 15.9 Hz. The device supports additional hardware by means of a 4-position 3.5 mm jack interface, including two reusable Velcro strap Ag/AgCl electrodes and an optical pulse sensor, attached by 9 in. and 1 m. lead wires, respectively. The optical pulse sensor contains electronics housed within an ear-clip, including a super-bright LED and an ambient light sensor. EDA and PPG signals are output by the board as voltage and converted by the Shimmer3 ADC to a 12-bit number representing the external skin conductance and PPG signal (pulse rate is derived from the PPG signal).

Processing and interpretation of EDA data presupposes methodological constraints to the design of HRI experiments. For example, it must be decided a priori what constitutes a stimulus event to correctly interpret event-related skin conductance responses. Because participants will likely show increased arousal due to mere exposure to the virtual world, procedures must be designed to include baseline periods (usually at the beginning of a session), which can be subtracted from event-related epochs to control for individual differences in resting state arousal.

EDA analysis consists of decomposing the signal into low-frequency (tonic) and high frequency (phasic) components, which are operationalized in terms of skin conductance level (SCL) and event-related skin conductance responses (ER-SCRs), respectively. In our test environment, EDA data is batch processed with Ledalab, to separate and extract the detrended phasic driver signal from the tonic component [1] (Fig. 2).

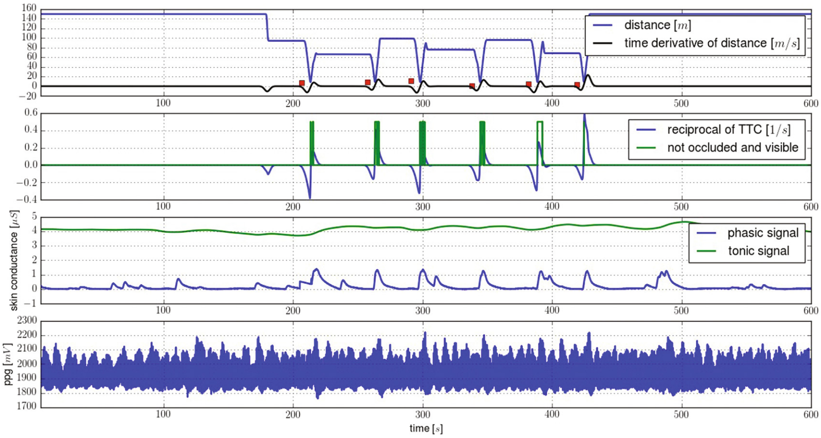


Fig. 2. Example simulation output and physiological signal.

Simulator Sickness Questionnaire (SSQ). The SSQ is a 16-item inventory used to assess simulator sickness. Responses are given according to a 4-point Likert scale (0 to 3). Factor analysis reveals three components: oculomotor discomfort, disorientation, and nausea [2]. The SSQ generates a total severity score and a score for each component. Scale scores for each item are computed by multiplying the reported value for each item by a weight and then summing across items for that component; weighted scale scores for each component can be found by multiplying each component by a unique weight. The total severity score can be computed by summing scale scores across the three components and multiplying by a weight.

Igroup Presence Questionnaire (IPQ). The IPQ is a scale for measuring the sense of presence experienced in a VE [3]. The IPQ consists of three subscales and one general item: Spatial Presence (sense of being physically present in the VE), Involvement (attention devoted to the VE and the involvement experienced), and Experienced Realism (subjective experience of realism in the VE). The general item assesses a general "sense of being there" and correlates highly with all three factors, especially Spatial Presence. Items are rated on a 7-point Likert scale (0 to 6). Means for each factor can be computed by averaging across items loading on that factor; three items contain reverse wording and must be reverse-scored before calculating factor means.

Real-time Data Acquisition. The real-time data acquisition is implemented using the Robot Operating System (ROS) platform [4]. Using cycle time in ROS makes it possible to collect data from the VR simulation and Shimmer device simultaneously, and in a synchronized manner (Fig. 3).

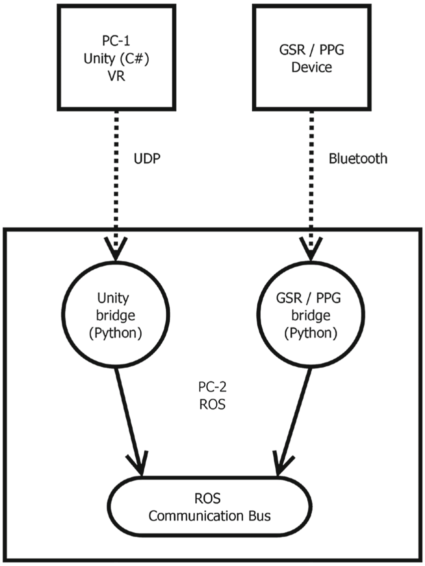


Fig. 3. ROS back-end data acquisition system diagram.

3 Results

Our method of testing human perception of flying robots provides synchronized measurements of the robot's trajectory and its concomitant effect on physiological arousal. Although there have been empirical studies to determine the relevant features of quadrotor motion for human perception [5], most investigations have focused on only a few parameters to represent entire trajectories, which may not be sufficient, e.g. there are infinitely many possible trajectories with the same average speed.

Deep learning techniques have been popularly applied to many areas, including speech recognition [6], image recognition [7], and natural language processing [8]. The present work employs a deep neural network model known as a recurrent neural network (RNN). The RNN has proven to be a powerful tool to approximate a large class of nonlinear dynamic systems [9]. The following section describes a preliminary result of applying a RNN to model physiological arousal induced by a flying robot in VR.

3.1 A Recurrent Neural Network Model

A RNN is used to approximate the dynamic behavior of human's arousal by the flying robot in VR. For example, past trajectories of the robot will be reflected in the human's arousal state, and if the robot is not moving for a long enough period, then the human's arousal state should decrease. To predict the human's arousal signal given the trajectory, a set of test data consisting of trajectories and arousal signals are used to train the RNN.

This section introduces the RNN employed in this research project. The robots state, $\mathrm{x_t} = [x(t), y(t), z(t), v_\mathrm{x}(t), v_\mathrm{y}(t), v_\mathrm{z}(t)]^\mathrm{T}$, is a vector containing the position and velocity of the quadrotor at time t, while $\mathrm{y_t}$ is the human's arousal state measured by a skin conductance sensor. The present work considers the Gated Recurrent Unit (GRU) [10], a type of RNN, with architecture expressed by the following discrete dynamic equations:

$$\mathrm{z_t} = \sigma_\mathrm{g}(\mathrm{W_z x_t} + \mathrm{U_z h_{t-1}} + \mathrm{b_z}) \tag{1}$$

$$\mathrm{r_t} = \sigma_\mathrm{g}(\mathrm{W_r x_t} + \mathrm{U_r h_{t-1}} + \mathrm{b_r}) \tag{2}$$

$$\mathrm{h_t} = \mathrm{z_t} \circ \mathrm{h_{t-1}} + (1 - \mathrm{z_t}) \circ \sigma_\mathrm{h}(\mathrm{W_h x_t} + \mathrm{U_h h_{t-1}} + \mathrm{b_h}) \tag{3}$$

$$\hat{y}_t = \sigma_\mathrm{y}(\mathrm{W_y h_t} + \mathrm{b_y}), \tag{4}$$

where $\sigma_\mathrm{g}(\cdot)$ and $\sigma_\mathrm{y}(\cdot)$ are sigmoid functions, and $\sigma_\mathrm{h}(\cdot)$ is a hyperbolic tangent function. The parameters, W, U, b of the model are determined by an optimization algorithm to minimize the mean square error of the prediction. As can be seen in Fig. 4, the RNN predicts changes in the values of arousal when the robot gets closer to the human. Despite the intuitive prediction result, the RNN has the potential to model complex behavior due to its ability to approximate nonlinear models. Using the entire trajectory as input to the model makes it possible to consider all features of the trajectory: speed, distance, approach angle, etc.

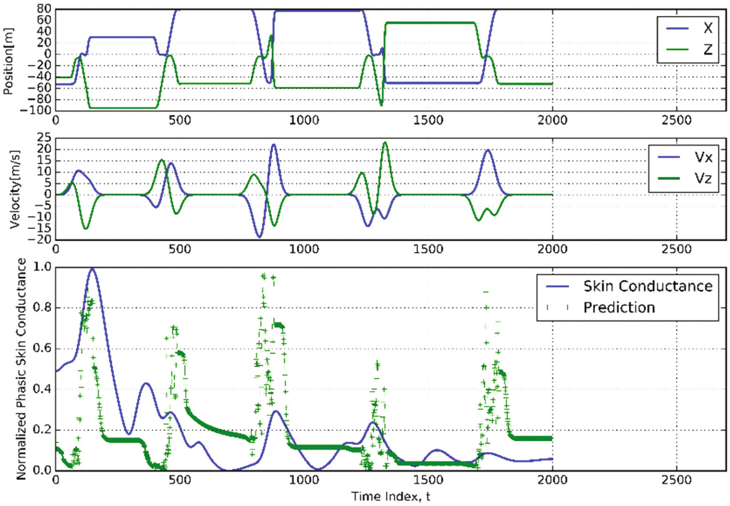


Fig. 4. Normalized phasic response and predicted phasic response. Given the position and time signal, the RNN model predicts the arousal measured in terms of skin conductance.

4 Discussion

The primary objective for this research was to generate a scalable VR test environment to support the investigation of human-robot interaction. To do this, we synthesized a data-driven workflow that enables high (and low) precision simulations of quadrotor flight behavior for a variety of realistic operating conditions. The back-end for this implementation enables the concurrent recording and processing of real-time data from the VR simulation and external sensing devices. This data is later analyzed using statistical methods, but also used as training data for deep learning. By combining these techniques we can converge on a model of human arousal in the presence of collocated UAVs, before extending our research into real-world human-robot interactions. Future research will attempt to validate the results from our VR experiments in the real world, and explore the relationship between actual and perceived safety when interacting with autonomous UAVs.

Acknowledgements. This work has been supported in part by the National Science Foundation through the National Robotics Initiative grant number 1528036 and EAGER grant number 1548409.

References

1. Benedek, M., Kaernbach, C.: A continuous measure of phasic electrodermal activity. J. Neurosci. Methods **190**(1), 80–91 (2010)
2. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**(3), 203–220 (1993)
3. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. Presence Teleoperators Virtual Environ. **10**(3), 266–281 (2001)
4. Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.Y.: Ros: an open-source robot operating system. In: ICRA Workshop on Open Source Software, vol. 3, p. 5. Kobe (2009)
5. Szafir, D., Mutlu, B., Fong, T.: Communication of intent in assistive free flyers. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, pp. 358–365. ACM (2014)
6. Graves, A., Mohamed, A.R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference, pp. 6645–6649. IEEE (2013)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
8. Mikolov, T., Kombrink, S., Burget, L., Černocký, J., Khudanpur, S.: Extensions of recurrent neural network language model. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5528–5531. IEEE (2011)
9. Funahashi, K.I., Nakamura, Y.: Approximation of dynamical systems by continuous time recurrent neural networks. Neural Netw. **6**(6), 801–806 (1993)
10. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)

Author Index

J. Chen (ed.), *Advances in Human Factors in Robots and Unmanned Systems*,
Advances in Intelligent Systems and Computing 595, DOI 10.1007/978-3-319-60384-1