

Feasibility of Wearable Fitness Trackers for Adapting Multimodal Communication

Daniel Barber^(✉), Austin Carter, Jonathan Harris, and Lauren Reinerman-Jones

Institute for Simulation and Training (IST), University of Central Florida (UCF),
Orlando, FL 32826, USA

{dbarber, acarter, jharris, lreinerm}@ist.ucf.edu

Abstract. In addition to efforts to increase the intelligence and perception capabilities of robots to enable collaboration with human counterparts, there is also a focus towards improving interaction mechanics. Multimodal communication is one such tool under investigation due to its dynamic ability to select explicit and implicit communication modalities with the aim of facilitating robust exchanges of information. Although there is extensive research in the domain of explicit communication using auditory, visual, and tactile interfaces, investigations into systems that incorporate implicit methods, or actually adapt and select appropriate modalities for reporting data from a robot to human is limited. Furthermore, a missing piece is identifying how and when to trigger these changes. A novel strategy to accomplish adaptation is through identification of teammate's physiological state. From the literature, one can find examples of researchers using high fidelity systems to measure physiological response and predict user workload, but many of these technologies are prohibitively expensive or not suitable for use in many domains of interest for human robot interaction such as dismounted infantry operations. Recent advancements in wearable consumer technologies, specifically fitness trackers supporting integration with third party software, are making it possible for incorporation of low cost systems in a variety of novel applications. A logical extension of these applications being physiological state measurement to drive adaptive automation in the form of multimodal interfaces. This paper describes the results of a study to assess the feasibility of using data from a wearable fitness tracker in an adaptive multimodal interface for squad-level human-robot interaction.

Keywords: Fitness trackers · Physiological assessment · Human-robot interaction · Multimodal communication · Adaptive automation

1 Introduction

A growing body of research and added commercial interest is pushing intelligence and perception capabilities of robots into new areas of collaboration with human counterparts. Through congressional mandate and funded research efforts, robots are no longer seen as remote control tools, but teammates capable of taking on different roles and responsibilities to accomplish a shared objective [1–3]. In addition to making robot teammates smarter, there is also a strong focus towards improving interaction mechanics

for seamless integration with end users within mixed-initiative teams. Within the mixed-initiative paradigm, teams employ flexible interaction strategies where each agent (human or robot) contributes what is best-suited at the most appropriate time [4]. At the root of any interaction between humans and robots is the exchange of information using auditory, visual, and tactile modalities. Appropriately using these modalities is required for effective communication, with interactions tailored to human expectations, demands, and mental models [5]. Multimodal communication is a framework and tool under investigation to meet this need due to its support for the flexible selection of explicit and implicit communication modalities to enable robust exchange of information when compared to single modalities [6–8]. Although there is extensive research in the domain of explicit communication using auditory, visual, and tactile interfaces, investigations into systems that adapt and select appropriate modalities for bi-directional interaction with human teammates is limited.

1.1 Adaptive Automation for Human Robot Interaction

The environments today's soldiers interact within are inherently complex. Working within teams, regardless of the presence of a robot, includes multitasking during which soldiers must pay attention to their own task execution and their teammates. For example, a cordon and search operation, one of the most frequently used tactics dismounted soldier teams use in complex urban environments, requires reconnaissance, enemy isolation and capture, and weapons and material seizures [9]. With the inclusion of robots to assist in cordon and search or other operations, there is potential for an increase in soldier workload due to the superhuman information gathering capabilities of robots. Robot teammates equipped with cameras, LIDAR, SONAR, and other sensors have the ability to capture and aggregate a multitude of data, which could negatively impact a soldier's situational awareness and workload if not delivered appropriately.

Adaptive automation refers to a system capability that enables task sharing between a human operator and a system [10, 11]. With robots and their interfaces becoming more capable and independent, adaptive automation is well suited to enabling mixed-initiative squad-level team concepts. Previous efforts at using adaptive automation in ground robot teaming scenarios have shown performance benefits [10, 12]. Extending this work to adapt multimodal communication is henceforth likely to improve team communication performance. In such a scenario, automation built within human robot interfaces can select the appropriate modality, or combination thereof, to deliver messages to soldiers in a way that does not increase cognitive demand or interfere with tasks using conflicting visual or auditory resources.

1.2 Implicit Communication for Adaptive Strategies

In addition to understanding what single or combination of modalities will result in the most effective exchange of information, a critical piece of the puzzle is identifying how and when to trigger these changes. Situational context when directly interacting with a robot is one method of triggering these changes, but an alternative strategy of interest from the domain of implicit communication is the identification of teammate

physiological states [13]. From the literature, one can find examples of researchers using electroencephalograms (EEG), electrocardiograms (ECG), eye tracking, and other sensors in combination to measure physiological response to classify a participant's level of workload [13–16]. By employing physiological sensors to classify a user's state, automation within an interface can trigger different multimodal communication strategies to maintain a baseline level of performance. For example, when a soldier is experiencing high workload; a multimodal interface may chunk auditory reports together or pair with tactile feedback to ensure messages are received. Teo, et al. [13] developed a closed-loop system that demonstrated this exact concept by using a combination of physiological sensors to trigger automation in a remote supervisory reconnaissance and surveillance mission with a ground robot.

1.3 Wearable Fitness Trackers for Implicit Communication

Although using physiological sensors to measure a person's state is a promising technique, many of the technologies, are prohibitively expensive or not suitable for use in squad level human robot teaming. Within this domain, users are on the move, and current physiological sensing devices would interfere with operations or disrupt their wearing of other equipment. Recent advancements in wearable consumer technologies, specifically fitness trackers supporting integration with third party software, are closing this gap, making it possible for incorporation of low cost, non-disruptive systems in a variety of novel applications. The Microsoft Band 2, Fig. 1, is an example of a wrist-worn device that supports real-time collection of heart rate, inter-beat interval, heart rate variability, skin temperature, ambient temperature, and galvanic skin response (GSR) over a Bluetooth connection on multiple operating systems [17].



Fig. 1. Microsoft band 2 wearable fitness tracker.

These sensors, in particular the optical heart rate monitor and GSR, provide similar measures to those used in previous efforts such as Teo, et al. [13]. However, the feasibility of fitness trackers for triggering adaptive communication is unclear, as the manufacturers did not design them with this purpose in mind and may not have the required sensitivity and saliency for accurate physiological state classifications.

1.4 Adaptive Multimodal Interfaces

Accomplishing the vision of adaptive multimodal communication interfaces for soldiers and robots requires a systematic investigation into the performance costs and benefits of single versus multiple modalities and physiological response within different mission contexts and environmental demands. However, a review of the literature shows a limited number of studies to date investigating multimodal communication within mixed-initiative infantry operations [18], with the majority focused on teleoperation [19, 20], humanoid robot assistants [7, 21], and vehicle driving scenarios [22–24]. Few meta-analyses have surveyed the performance costs and benefits of redundant versus single-modality presentation for an interrupting and ongoing task [22, 23]. Moreover, conflicting results across studies demonstrate unclear effects of modality switching on vision-based signal detection tasks like those of cordon and search operations [25]. The goal for this effort is to address this gap by beginning to understand independent and redundant communication modalities and adaptive strategies in squad-level human robot interactions. Specifically, the aim for this paper is to assess the feasibility of using wearable fitness trackers as a means of state identification for adapting multimodal communication.

2 Method

2.1 Participants

A total of 56 (34 males, 22 females) participants between the ages of 18 and 40 ($M = 19.29$, $SD = 2.29$) participated in the study. All participants received credit for their psychology courses for completing the study. Participants were asked not to consume alcohol or any sedative medication for 24 h or caffeine for two hours prior to the study.

2.2 Equipment and Simulation Environment

As previously mentioned, cordon and search is one of the most common operations a squad may perform in an urban environment. It also contains enough complexity to make it well suited for investigating the challenges of mixed-initiative teaming between humans and robots. For the present effort, a custom 3D simulation using the Unreal 4 Game Engine [26] was created, Fig. 2. Within the simulation, participants took the role of a squad leader performing the outer cordon task. This outer cordon activity replicated a signal detection task, [27], where participants were required to look for insurgents walking in front of and around a building at different event rates. If participants detected an insurgent, they used a mouse to click on the character, which the software logged. A 30" monitor with a resolution of 2560×1600 pixels was used to present the environment.



Fig. 2. Unreal 4 game engine simulation used in experiment. Image represents the 3D field of view participants experienced while executing an outer cordon operation. Characters in the environment were animated and walked on and off screen and variable event rates. At the top center is an overlay of the multimodal interface visual display when present.

In addition to the outer cordon signal detection task, participants received information from two virtual robot teammates performing the inner cordon task (not within the participants' field of view). A modified version of the multimodal interface (MMI)

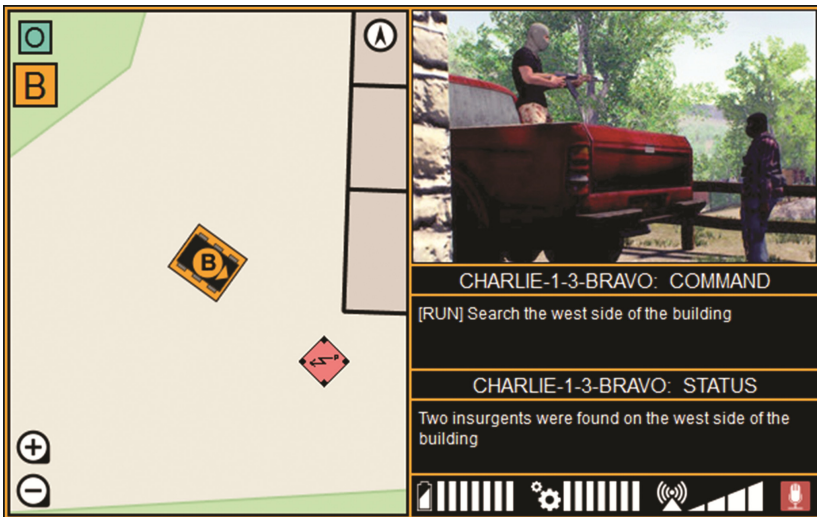


Fig. 3. Multimodal interface (MMI) visual display. Display is comprised of three main areas: semantic map (left) showing robot location and icons of objects found, video/camera (top right), and status (bottom right) illustrating the current command the robot is executing and the most recent report (in text) of what the robot detected.

developed by Barber et al. [5] was used to deliver auditory and visual reports from robot teammates. The visual display of the interface, illustrated in in Fig. 3, appeared on top of the 3D simulation at the top center area of the screen at a resolution of 602×377 pixels, Fig. 2. The resolution and size of the visual display was scaled to match 1:1 to the physical size of the Toughpad FZ-M1 tablet used in [5] when shown on the 30" monitor used in the study. For each visual report, the display was present for 10 s before being hidden off screen. For the auditory modality, text reports were converted to speech using Microsoft Speech Platform SDK version 11 text-to-speech (TTS) and the default male voice of the Windows 10 operating system, [28].

For physiological data capture, the Microsoft Band 2 was used [17]. The Microsoft Band 2 was selected because it provided sensor data corresponding with previous research efforts for classifying physiological state (e.g. interbeat interval (IBI)), and Microsoft providing a Windows software development kit (SDK) enabling real-time capture of data during the simulation. A custom software application using the provided SDK captured and recorded data from the device.

2.3 Design

A 3 (Adaptive Strategy: Constant, MMI, User) \times 3 (Modality: Auditory, Visual, Auditory and Visual) \times 2 (Environmental Demand: High, Low) repeated measures design was employed. For adaptive strategy, constant (C) meant that no adaptation to communication modality occurred, MMI indicated that the multimodal interface triggered changes to modality, and for user (U), the participants triggered which modality was

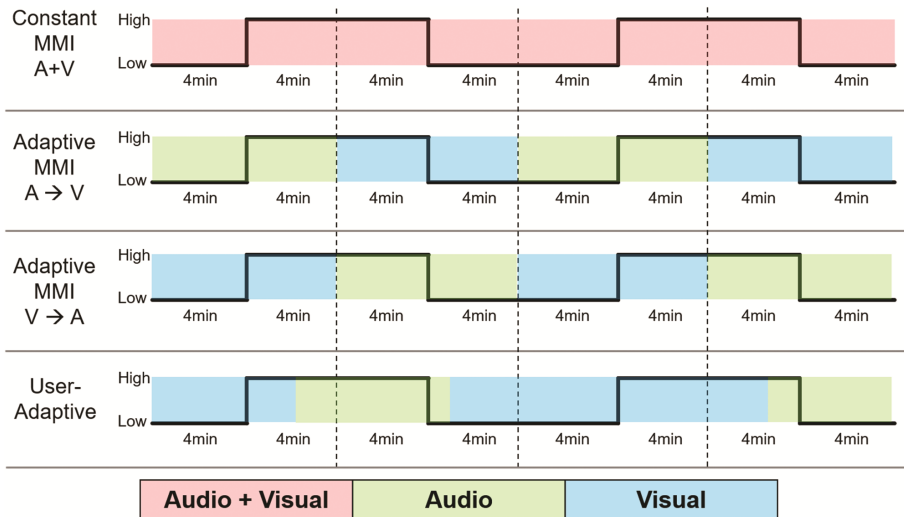


Fig. 4. Scenario design for the experiment. Four scenarios were created for each participant, one constant adaptive strategy (both visual and audio reports), two MMI adaptive (audio to visual and visual to audio), and one user-adaptive. Event rate (high/low) for the signal detection task is indicated by the square wave function, with report periods divided into 4 min blocks.

used with the spacebar key on the keyboard. Two modalities were used auditory (A) and visual (V) and depending on the adaptive strategy were presented standalone or redundantly. Environmental demand during scenarios was either high (H) or low (L). To manipulate environmental demand, event rates of 15 events/minute and 60 events/minute on the signal detection task were used corresponding to low and high task load respectively. Selection of these event rates was taken from Abich et al. [29], which established event rates for a similar signal detection task that elicited distinct levels of low and high workload as reported by the NASA-TLX.

A total of four scenarios were created to capture the experimental design and collect an equal amount of data across adaptive strategy, communication modality, and environmental demand (event rate), illustrated in Fig. 4.

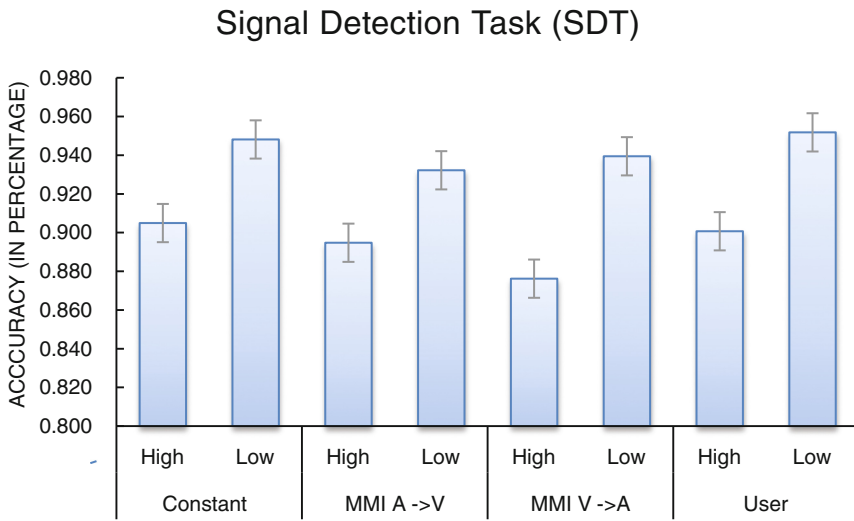


Fig. 5. Comparison of IBI between high and low environmental demands across adaptive strategy scenarios. Error bars represent standard error.

Each scenario was sub-divided into eight 4-minute blocks where participants received nine reports each, for a total of 72 reports. After three reports, participants were asked two questions regarding the information received to measure their situational awareness (SA). These “SA probes” were delivered via pre-recorded audio, and participants responded verbally. Six SA probes were given for each 4-minute block for a total of 48 per scenario. Through the manipulation of event rate within each scenario, and the break down of modality transitions, performance and physiological response was captured across adaptive strategies and during low and high environmental demands. For each scenario, a different building location was used which was counterbalanced across manipulations and participants. Furthermore, presentation order for each scenario was randomized and counterbalanced across participants as well.

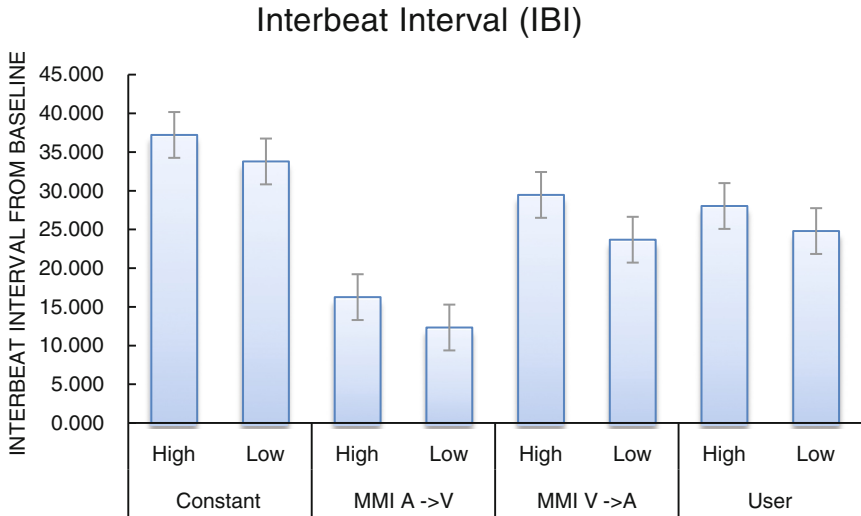


Fig. 6. Comparison of HRV between high and low environmental demands across adaptive strategy scenarios. Error bars represent standard error.

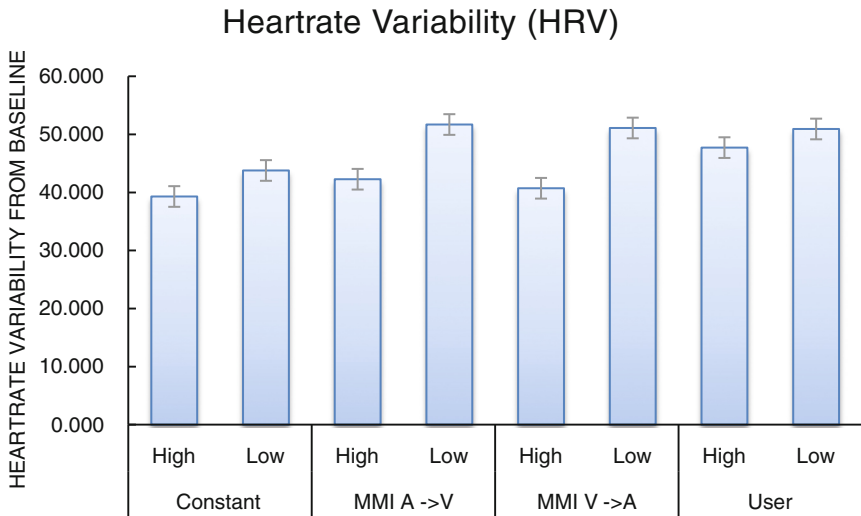


Fig. 7. Comparison of GSR between high and low environmental demands across adaptive strategy scenarios. Error bars represent standard error.

Dependent Variables

Signal Detection Task (SDT). The accuracy of participants in identifying enemy insurgents [27]. For each time period analyzed, the total number of correctly identified insurgents was divided by the total number of insurgents presented to obtain an accuracy percentage.

Interbeat interval (IBI). Interbeat interval as reported from the Microsoft Band 2, measured as the time in milliseconds between heartbeat RR peaks of the QRS complex [30]. For each time period analyzed, the mean IBI was calculated and normalized across participants by subtracting the mean resting baseline value.

Heartrate Variability (HRV). The variance of the interbeat interval reported from the Microsoft Band 2. For each time period analyzed, the mean HRV was calculated and normalized across participants by subtracting the mean resting baseline value.

Galvanic Skin Response (GSR). Mean skin resistance converted to Siemens reported from the Microsoft Band 2 [30]. For each time period analyzed, the mean GSR was calculated and normalized across participants by subtracting the mean resting baseline value.

3 Procedure

Upon arrival, participants first completed an informed consent document, and then were equipped with the Microsoft Band 2 on the wrist of their non-dominant hand after the area was cleaned with an alcohol pad. Participants then completed a demographics questionnaire, followed by measurement of a five-minute wakeful resting baseline with the Microsoft Band 2. Next, they were trained on each of the tasks they would perform individually, then in combination. To do this, participants were first trained on the character models used in the signal detection task and which models were considered enemies to detect and which were not. They then performed the signal detection task with a low to high event rate transition as practice. Following the signal detection task training, example visual and audio reports from the robot were demonstrated, with focus on what information they would need to recall during situation awareness (SA) probes. Participants then performed practice scenarios with SA probes for each of the modalities (A, V, A + V). After these practice scenarios, participants completed four additional practice scenarios with the combined signal detection task and robot reports that covered the four types of experimental scenarios they would encounter. After completing the practice scenarios, they performed each of the experimental scenarios. Performance during practice scenarios was not used to screen participants from performing experimental scenarios. After completing all four experimental scenarios, participants were debriefed and dismissed.

4 Results

4.1 SDT

A 4 (Adaptive Strategy: Constant, MMI Audio to Visual, MMI Visual to Audio, User) \times 2 (Environmental Demand: High, Low) repeated measures ANOVA was performed for performance on the signal detection task revealing a significant main effect for adaptive strategy ($F(2.50, 97.33) = 3.30, p = .03, \eta^2 = .08$), Fig. 8. A pairwise

comparison with a Bonferroni correction indicated that participants in the constant adaptive strategy ($M = 0.94, SD = 0.04$) identified insurgents more accurately than participants in the audio to visual adaptive strategy ($M = 0.92, SD = 0.05, p = .002$). A significant main effect for environmental demand ($F(1, 42) = 88.13, p < .001, \eta^2 = .68$) Fig. 8, was also found, such that performance was higher in a low event rate ($M = 0.96, SD = 0.04$) than in a high event rate ($M = 0.91, SD = 0.05$).

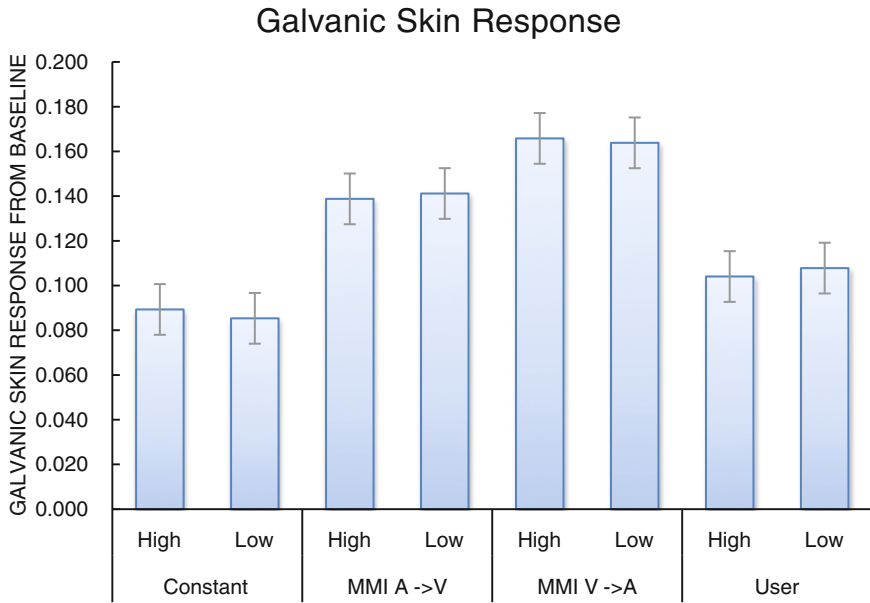


Fig. 8. Comparison of SDT between high and low environmental demands across adaptive strategy scenarios. Error bars represent standard error

4.2 IBI

A 4 (Adaptive Strategy: Constant, MMI Audio to Visual, MMI Visual to Audio, User) \times 2 (Environmental Demand: High, Low) repeated measures ANOVA was performed and showed no significant main effect for mean IBI between adaptive strategies or environmental demand, Fig. 5.

4.3 HRV

A 4 (Adaptive Strategy: Constant, MMI Audio to Visual, MMI Visual to Audio, User) \times 2 (Environmental Demand: High, Low) repeated measures ANOVA was performed and showed no significant main effect HRV for adaptive strategy. There was however a significant main effect for environmental demand, ($F(1, 36) = 27.66, p < .001, \eta^2 = .44$), such that HRV during high demand ($M = 42.49, SD = 37.56$) was lower

than during low ($M = 49.367$, $SD = 36.06$), Fig. 6. No significant interaction between adaptive strategy and modality was shown.

4.4 GSR

A 4 (Adaptive Strategy: Constant, MMI Audio to Visual, MMI Visual to Audio, User) \times 2 (Environmental Demand: High, Low) repeated measures ANOVA was performed and revealed no significant effects for mean GSR between adaptive strategies or high and low task demands, Fig. 7.

5 Conclusion

The present study described a starting point in the advancement of dynamic multimodal interfaces capable of changing presentation format to ensure robust communication. An experimental design demonstrating different adaptive strategies and levels of environmental demands was used to measure impacts on task performance and the sensitivity of a commercial-off-the-shelf wearable (Microsoft Band 2) to detect these changes. An analyses of task performance on an SDT supported previous findings related to manipulation of event rate reported by Abich et al. [29], where participants' detection accuracy decreased during higher event rates. Furthermore, a performance difference was also revealed for adaptive strategy type, with highest SDT performance in the constant (dual modality) condition. Following these findings, analyses of the Microsoft Band 2 data showed that heart rate information was most sensitive to changes in environmental demands, with GSR showing no effects. Specifically, HRV results showed significant differences between low and high task demand within adaptive strategies, such that HRV was higher during low, and lower for high environmental demand. This finding supports previous research correlating HRV and workload, [31]. Although promising, further work is still needed to determine if these findings are consistent across different task types within this domain before attempting to dynamically change modalities. Furthermore, although differences in performance on the SDT were shown between adaptive strategies, more analyses are still required to understand impacts of these strategies on working memory and situational awareness on longer duration exercises, and whether when you adapt the modality during changes to environmental demand matters.

Acknowledgments. This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0016. The views and conclusions contained in this document are those of the author's and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Phillips, E., Ososky, S., Grove, J., Jentsch, F.: From tools to teammates: toward the development of mental models for intelligent robots. In: Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting (2011)
2. U.S. Congress, National Defense Authorization Act for Fiscal Year 2001, Washington, D.C. (2001)
3. U.S. Army Research, Laboratory, Robotics, 1 June 2011. <http://www.arl.army.mil/www/default.cfm?page=392>. Accessed 10 Feb 2017
4. Hearst, M., Allen, J., Guinn, C., Horvitz, E.: Mixed-initiative interaction: trends & controversies. In: IEEE Intelligent Systems, pp. 14–23 (1999)
5. Barber, D.J., Abich, J., Phillips, E., Talone, A., Jentsch, F., Hill, S.G.: Field assessment of multimodal communication for dismounted human-robot teams. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2015)
6. Lackey, S.J., Barber, D.J., Reinerman-Jones, L., Badler, N., Hudson, I.: Defining next-generation multi-modal communication in human-robot interaction. In: Human Factors and Ergonomics Society Conference, Las Vegas (2011)
7. Bischoff, R., Graefe, V.: Dependable multimodal communication and interaction with robotic assistants. In: 11th IEEE International Workshop on Robot and Human Interactive Communication (2002)
8. Partan, S., Marler, P.: Communication goes multimodal. *Science* **283**(5406), 1272–1273 (1999)
9. Sutherland, J., Baillergeon, R., McKane, T.: Cordon and search operations: a deadly game of hide and seek. *Air Land and Sea Bulletin: Cordon and Search*, pp. 4–10 (2010)
10. Cosenzo, K., Chen, J., Reinerman-Jones, L., Barnes, M., Nicholson, D.: Adaptive automation effects on operator performance during a reconnaissance mission with an unmanned ground vehicle. In: Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, Los Angeles, CA (2010)
11. Parasuraman, R., Hancock, P.: Adaptive workload and control. In: *Workload and Fatigue*, Mahwah, NJ, USA, pp. 305–320. Lawrence Erlbaum Associates (2001)
12. Taylor, G.S., Reinerman-Jones, L., Szalma, J.L., Mouloua, M., Hancock, P.A.: What to automate: addressing the multidimensionality of cognitive resources through system design. *J. Cogn. Eng. Deci. Making* **7**(4), 311–329 (2013)
13. Teo, G., Reinerman-Jones, L., Matthews, G., Barber, D., Harris, J., Hudson, I.: Augmenting robot behaviors using physiological measures of workload state. In: *Human Computer Interaction International*, Toronto, CA (2016)
14. Barber, D., Reinerman-Jones, L., Lackey, S., Hudson, I.: Augmenting robot behaviors using physiological measures. In: Schmorow, D.D., Fidopiastis, C.M. (eds.) *FAC 2011*. LNCS, vol. 6780, pp. 567–572. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-21852-1_65](https://doi.org/10.1007/978-3-642-21852-1_65)
15. Abich, J., Reinerman-Jones, L.R., Taylor, G.S.: Investigating workload measures for adaptive training systems (2013)
16. Mercado, J., Reinerman-Jones, L., Barber, D., Leis, R.: Investigating workload measures in the nuclear domain. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2014)
17. Microsoft, Microsoft Band Features, Microsoft. <https://www.microsoft.com/microsoft-band/en-us/features>. Accessed 10 Feb 2017
18. Redden, E., Elliot, L.: Robotic control systems for dismounted soldiers. In: Jentsch, F.G., Barnes, M.J. (eds.) *Human-Robot Interactions in Future Military Operations*, pp. 335–352. Ashgate (2010)

19. Pettitt, R., Redden, E.S., Carsten, C.: Scalability of robotic controllers: speech-based robotic controller evaluation (ARL-TR-4858). Aberdeen Proving Ground, MD: US Army Research Laboratory, pp. 1–46 (2009)
20. Oron-Gilad, T., Redden, E.S., Minkov, Y.: Robotic displays for dismounted warfighters. *J. Cogn. Eng. Decis. Making* **5**(1), 29–54 (2011)
21. Bennewitz, M., Faber, F., Joho, D., Behnke, S.: Fritz - a humanoid communication robot. In: *The 16th IEEE International Symposium on Robot and Human interactive Communication* (2007)
22. Wickens, C.D., Prinnet, J., Hutchins, S., Sarter, N.B., Sebok, A.: Auditory-visual redundancy in vehicle control interruptions: two meta-analyses. In: *Human Factors and Ergonomics Society 55th Annual Meeting*, Santa Monica, CA, USA (2011)
23. Wickens, C.D., Gosney, J.: Redundancy, modality, and priority in dual task interference. In: *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, Santa Monica, CA, USA (2003)
24. Lu, S.A., Wickens, C.D., Prinnet, J.C., Hutchins, S.D., Sarter, N., Sebok, A.: Supporting interruption management and multimodal interface design: three meta-analyses of task performance as a function of interrupting task modality. *Hum. Factors* **55**(4), 697–724 (2013)
25. Reinerman-Jones, L.R., Taylor, G., Sprouse, K., Barber, D., Hudson, I.: Adaptive automation as a task switching congruence challenge. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2011)
26. Epic Games, Unreal Engine, Epic Games. <https://www.unrealengine.com/what-is-unreal-engine-4>. Accessed 10 Feb 2017
27. Green, D.M., Swets, J.A.: *Signal Detection Theory and Psychophysics*. Wiley, New York (1966)
28. Microsoft: Microsoft Speech Platform - Software Development Kit (SDK) (Version 11), Microsoft. <https://www.microsoft.com/en-us/download/details.aspx?id=27226>. Accessed 10 Feb 2017
29. Abich, J., Taylor, G., Reinerman-Jones, L.: Establishing workload manipulations utilizing a simulated environment. In: *Human Computer Interaction International*, Las Vegas, NV, USA (2013)
30. Microsoft: Microsoft Band SDK. <https://developer.microsoftband.com/Content/docs/Microsoft%20Band%20SDK.pdf>. Accessed 10 Feb 2017
31. Matthews, G., Reinerman-Jones, L., Barber, D., Abich, J.: The psychometrics of mental workload: multiple measures are sensitive but divergent. *Hum. Factors* **57**(1), 125–143 (2015)