Effects of Augmented Feedback and Motor Learning Adaptation on Human–Automation Interaction Factors

Sarah M. O'Meara,* John A. Karasinski,* Casey L. Miller,† Sanjay S. Joshi,‡ and Stephen K. Robinson§
University of California, Davis, Davis, California 95616

https://doi.org/10.2514/1.I010915

Increased interest in body–machine interfaces necessitates understanding how to train users to use nontraditional inputs. In this study, a control task driven by subject-activated surface electromyography was developed as a testbed to observe the effects of automated training methodologies on the development of performance, workload, and trust. Forty-eight subjects learned to use a surface-electromyography-based command system to perform a Fitts's-law-style cursor-to-target task with 120 training trials and 40 evaluation trials. Subjects were divided into four groups: control, concurrent feedback, terminal feedback, and adaptive threshold. The control group trained and learned through repetition using the visual feedback of the cursor position. The concurrent feedback group received additional concurrent visual feedback during command input, and the terminal feedback group had supplementary visual feedback after command input. The adaptive threshold group did not have any additional feedback, but experienced changes in the cursor control designed to induce motor learning adaptation. The results indicate that 1) additional visual feedback improves task performance, workload, and trust during training, and 2) the groups converged in their command proficiency by the end of training.

I. Introduction

HUMAN–MACHINE interfaces (HMIs) traditionally use input devices such as a computer mouse, joystick, and steering wheel. A more recent subset of HMIs leverages signals derived from the human body to communicate intent and has been termed as "body–machine interfaces" (BoMIs) [1]. These BoMIs may be passive, where the body signals are used as a state indicator to monitor or trigger a response to the human, or active, where the human directly controls the input to determine the device output. Furthermore, BoMIs can be characterized by the body signal they use, such as brain signals, gestures, or muscular signals, and whether the measurement is invasive (i.e., penetrates the body) or noninvasive. BoMIs that use brain signals may be termed as "brain–machine interfaces" (BMIs) or "brain–computer interfaces" (BCIs), where the latter is a BMI specifically designed for computer-based applications. The current discussion is focused on noninvasive techniques for aviation and space applications.

An increasing interest in BoMIs for use in aviation has led to the development of several passive and active systems. Passive systems have measured brain signals to predict the occurrence of inattentional deafness to auditory alarms [2], assess cognitive fatigue [3], monitor cognitive load [4], and track vigilance [5]. For example, Di Flumeri et al. [5] demonstrated improved vigilance when a BCI system detected changes in subjects' vigilance and automatically adjusted the automation level to increase/decrease manual engagement for an air traffic control task. Electromyography (EMG), the recording of electrical muscular signals, has also been used for passive BoMIs to warn of gravity-induced loss of consciousness [6,7], monitor muscle fatigue [8,9], and differentiate pilot skill level [10]. Less prevalent has been the inclusion of EMG in active BoMIs for aviation. The work by Jorgensen et al. [11] demonstrated the successful implementation of an EMG BoMI to control the pitch and bank rates for an aircraft landing in simulation. In contrast, there has been an increased ubiquity in the use of body gestures for active BoMIs related to drone control (see Refs. [12–15] for some recent examples). One study observed significantly better performance with torso gestures than a traditional HMI, a joystick [13], lending to the feasibility and potential benefits of these nontraditional interfaces.

In 2004, the European Space Agency's Advanced Concept Team identified noninvasive BMIs as a research area with potential space applications [16]. Broschart et al. [17] proposed incorporation of BMIs for tele-operations, assistive robots, and to increase astronaut productivity, but recognized that the unique physical environment coupled with physiological changes experienced by astronauts would pose implementation challenges. Some initial steps have occurred, such as a BCI mouse for online control of a simplified, simulated spacecraft in yaw and pitch [18]. However, the findings from this study are limited due to low subject number and the simplified simulation but indicate an interesting application. Another preliminary study with limited subjects used a BMI during parabolic flights which simulate microgravity and found that the brain signals measured by electroencephalography (EEG) remained stable enough to achieve an average classification accuracy of 73.1% [19]. In a review of BMIs for space applications, Coffey et al. [20] argued that the implementation of noninvasive active BMIs was unlikely due to gaps in information transfer rates, accuracy, and intuitiveness, whereas passive BMIs were more advantageous because 1) they did not require active engagement or control; 2) they had established reliability for monitoring aspects like attention, workload, and task engagement; and 3) the derived human state information could be fed into adaptable autonomous systems [20]. However, development of BMIs for space applications may prove challenging due to relatively high system complexity and low signal-to-noise ratio (SNR) compared to other BoMIs.

The use of muscular signals in active BoMIs has been understudied for aviation and space applications, but BoMIs that use EMG have multiple advantages. When EMG is measured from the skin's surface, it is noninvasive and referred to as surface EMG (sEMG). An sEMG system can be placed above any muscle location, can be designed to use a single sensor, and has a relatively high SNR. The sEMG signal quality and reproducibility are affected by body pose

Presented as Paper 2020 at the AIAA SciTech Forum, Orlando, FL, January 6–10, 2020; received 13 October 2020; revision received 21 January 2021; accepted for publication 22 January 2021; published online 9 March 2021.

*Graduate Student Researcher, Mechanical and Aerospace Engineering Department. Student Member AIAA.

†Research Assistant, Mechanical and Aerospace Engineering Department.

‡Principal Investigator, Robotics, Autonomous Systems, and Controls Laboratory, Professor, Mechanical and Aerospace Engineering Department. Senior Member AIAA.

§Principal Investigator, Human/Robotics/Vehicle Integration and Performance Laboratory, Professor, Mechanical and Aerospace Engineering Department. Associate Fellow AIAA.

[21], muscle fatigue [22], and sensor displacement [23], and mitigating these issues is an active area of research (see review by Kyranou et al. [24]). Furthermore, EMG control has already been applied to prosthetics (see reviews by Refs. [25,26]) and robotics (e.g., see Refs. [27–30]), which have potential aerospace applications (i.e., tele-operations [31]). In general, BoMIs provide a control modality that can be used in addition to or in replacement of traditional HMIs. Users are not constrained to use an input device with their hands (e.g., a computer mouse), and performance may improve with a BoMI (e.g., see Ref. [13]).

The aforementioned BoMI systems previously studied focused on the development and demonstration of the interface but did not address user training beyond task repetition or test setup familiarization. For a novel input, it is important to efficiently train new users. We decided to apply automation to the skill-based training of learning sEMG control and compare different automated training methodologies. Selection of a training methodology should take into consideration the effects on task and human–automation interaction factors, such as trust and workload, and should also account for context-specific interface elements (e.g., flight displays, communication interfaces).

A. Selected Training Methodologies

Training novice users to effectively use sEMG control methods can be a long and difficult task [32]. Although expert instructors can be effective at increasing learning rates, their time can be costly, and their availability is often limited. For these reasons, we were interested in automated training methodologies that seek to fill the role of an expert instructor. Self-directed practice has been shown to work as well as instructor guided practice for learning complex medical skills [33], though it is unclear how well this paradigm would transfer for nonmedical tasks. The use of augmented feedback strategies, however, has been shown to help reduce training times and they have the potential to reduce cognitive workload, or the amount of mental effort allocated, in especially demanding tasks. Augmented feedback provides information that "cannot be elaborated without an external source; thus, it is provided by a trainer or a display" and has been shown to effectively improve performance in a wide variety of motor tasks [34]. Augmented feedback has been used for training, motor skill acquisition, rehabilitation, and operational assistance for tasks that range from simple, closed-environment laboratory demonstrations to complex, operational cases [34]. Recent approaches to using augmented feedback have focused on multimodal cueing and virtual and augmented reality displays in a variety of medical [35–37], aerospace, and robotics tasks [38–40]. Across these studies, results indicate that providing task-appropriate augmented feedback can improve motor skill acquisition, final performance, and retention, though these results vary depending on the means of the augmented feedback presented. Many of these studies have compared how to provide feedback across different modalities to identify optimal pairings between modality and tasks but investigating when to provide feedback is equally important. Concurrent feedback is presented in real-time, as subjects execute a task, whereas terminal feedback is presented after the task is completed. Some researchers have found that intermittent, concurrent high-frequency feedback is superior to low-frequency feedback [35,41], whereas others have found that terminal feedback outperforms a concurrent approach [42,43], and it is difficult to generalize between different feedback and task types [44,45]. In general, however, concurrent feedback has been shown to be more useful with higher functional task complexity, whereas terminal feedback is often less useful when complexity is high [34]. Concurrent feedback has recently shown great promise in EMG control, though less progress has been made comparing concurrent and terminal feedback strategies or investigating long-term learning effects [46-49]. Therefore, it is important to compare concurrent and terminal feedback approaches within the same task to better understand the appropriate application of training methodology.

Biofeedback, which applies augmented feedback strategies to measured physiological signals such as EMG, has proven to be a useful tool for improving performance and assisting in rehabilitation [50]. Researchers have investigated various augmented biofeedback techniques and found that they help subjects to "become more cognizant of their own EMG signal" [51], allowing for better control of their muscle activity. A recent review of the biofeedback literature suggests that "[b]iofeedback is more effective than usual therapy," though they also note that "[f]urther research is required to determine the long-term effect [biofeedback has] on learning" [52]. In our study, we refer to any additional, visual information given about an sEMG signal input or its processed output as "augmented feedback." The augmented training strategies are analogous to alerting automation and the strategy of complementation in aviation [53], where the augmented training strategies cooperate with and provide additional information to the user.

Another possible training methodology which leverages adaptation comes from motor learning. Users learn and adapt to improve performance throughout practice and training, and induced variability can improve performance. For example, Seow et al. [54] varied a task parameter, where subjects either trained on a set thrust level or a variable one for a spaceship video game. Subjects with variable thrust training performed better than those with a consistent thrust when tasked with a novel thrust level [54]. Braun et al. [55] showed during a planar reaching task that randomly varying the feedback uncertainty (i.e., the difference between the actual and shown position of the hand) led to improved skill generalization. Surface EMG control adaptation may follow Bayesian theory [56], where mapping uncertainty (or the uncertainty of the brain's model of the applicable system) increases adaptation rate. Conversely, increased sensorimotor feedback uncertainty decreases adaptation rate. Lyons and Joshi [57] demonstrated that subjects exposed to a mapping uncertainty during cursor control had higher adaptation rates to sensorimotor feedback when the mapping uncertainty was removed, indicating that artificially added mapping uncertainty during training may increase adaptation rate. Additionally, a noisier sEMG control method (i.e., increased mapping uncertainty) provides more information about the control signal, which can lead to a larger adaptation rate compared with a more filtered classification method [58]. Similarly, introducing noise in a joystick-controlled final approach phase flight task has been shown to increase the rate of motor skill acquisition [38]. Taken together, these studies indicate that varying a task parameter [54,55] and increasing mapping uncertainty [57,58] can potentially improve skill acquisition. However, research in this area is not well-studied for sEMG control or for tasks that do not use the planar reaching paradigm [59]. Therefore, we have included an adaptation-based strategy in our study to compare with more established training methodologies to investigate the feasibility of this method.

B. Workload and Trust in Automation

When evaluating the different automated training methodologies, it was important to select appropriate training feedback that did not cause additional cognitive demands. Karasinski et al. [60] investigated the effects of concurrent feedback in a simulated, four-degree-of-freedom, manually controlled spacecraft inspection task. Subjects in the feedback group performed the task much faster and more accurately than those in a control group and reported a significantly lower workload. Workload can be defined as "the cost incurred by a human operator to achieve a particular level of performance" [61,62], and is commonly assessed using quantitative, subjective techniques such as the NASA-TLX [61,62] and Modified Bedford Workload [63] scales. Although there are limits to these subjective techniques, such as the possibility of large intersubject variability, they provide a rapid and easily administered tool for workload assessment. The Modified Bedford Workload scale was administered frequently throughout this study as a rapid means of estimating spare cognitive load. In addition to subjective workload measures, there are a variety of techniques that aim to objectively estimate workload. Physiological measurements such as heart rate variability have been used to assess workload for many types of actual and simulated aerospace flight tasks [64], and heart rate and heart rate variability can also provide objective measurements that have been previously correlated with subjective scales such as the Bedford and NASA-TLX ones [65].

Researchers have had mixed success in being able to generally relate physiological measures to subjective workload scales, however, and physiological workload estimates were not used here. Among the most common objective measurement techniques is the secondary task, which requires subjects to complete the primary task then use any spare cognitive margin to respond to an additional task [66]. This secondary task approach does not work well for simpler tasks as completing it begins to compete for attention with the primary task [67]. No objective workload measurements were included in this study.

Trust is another important factor when considering human–automation interaction, and inappropriate trust can lead to the disuse or misuse of automated systems. Trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [68]. The reliability of a system, in particular, has been shown to be an important aspect of an operator's trust in a system [69]. Although there have been many proposed models for trust, Hoff and Bashir's three-layer model deserves particular attention [70]. After performing a systematic review of the literature, they developed a three-layer model of trust that is split between dispositional trust, situational trust, and learned trust. Though researchers have little control over dispositional trust, they can affect situational trust by varying an experimental interface or environment and learned trust can be evaluated using repeated measures. Our study aimed to alter situational trust through different automated training methodologies, where the agent or the automated feedback in place of an instructor helped the subjects learn a skill. We expected subjects in each group to perceive whether the agent will help them achieve their goals differently based on the assigned training methodology. We also assessed learned trust with measurements taken periodically during the task. The subjects did not have an option to turn off or alter the agent but may have felt that the agent was not aiding in their goals. If subjects became reliant on the agent, we expected to see decreased performance when the agent was removed.

C. Study Objective

Although it is not the objective of this work to develop or optimize an sEMG system for one particular application, the sEMG system can act as a testbed in which the effects of different automated training methodologies are evaluated, and the novelty in using these systems enables the observation of early learning effects. The sEMG controlled Fitts's law task naturally lends itself as an excellent testbed for evaluating augmented feedback strategies. Most participants cannot easily predict or quickly understand the signal they generate and pass into the sEMG controller, and augmented feedback can provide insight into what would otherwise be an inherently noisy and elusive process. Modern systems increasingly incorporate humans with automation, necessitating an enhanced understanding of human–automation interaction. The purpose of this study was to investigate human–automation interaction with an emphasis on performance, workload, and trust during early learning across different automated training methodologies.

To the authors' knowledge, there is no study investigating differences in performance, workload, and trust across different automated training methodologies using augmented feedback and motor learning adaptation. In this study, we address the effects of automated training methodology on performance, workload, and trust during a computer-based Fitts's law [71] style cursor-to-target task. The training methodologies include repetition, concurrent feedback, terminal feedback, and an adaptive method. The treatments are removed for the evaluation phase to assess if subjects in the augmented feedback conditions (concurrent feedback and terminal feedback) succumbed to the guidance hypothesis [72], which would indicate that they have become reliant on the feedback in order to perform the task. The experiment is designed to provide subjects with a sufficiently challenging task so that we may observe early learning effects and the evolution of performance, workload, and trust.

II. Materials and Methods

A. Subjects and Experimental Setup

The UC Davis Institutional Review Board approved the study protocol, and subjects were recruited from the university student population. Exclusion criteria for subjects included a history of neuromuscular disorders, physical limitations of dominant arm, and prior sEMG control experience. Subjects provided written consent before participation. A total of 55 subjects volunteered, and subjects were released from the study due to equipment issues ($N = 3$), withdrawal request ($N = 2$), and significant motivation issues (e.g., not attempting the task; $N = 2$). The remaining 48 subjects completed the protocol and had an average age of 20.1 ± 1.4 years ($\mu \pm \sigma$), included 2 left-hand-dominant subjects, and had an equal participation of biological sexes.

Subjects were trained to control a cursor with sEMG to perform a Fitts's-law-style cursor-to-target task with a center-out paradigm [73,74]. During the experiment, subjects sat in front of a desk and computer screen (see Fig. 1a) with electrodes on their forearms. Two electrodes (ConMed 1620 Ag/AgCl center snap) approximately 2.5 cm apart were placed on the dominant hand side near the extensor digitorum proximal attachment. A reference electrode was located near the lateral epicondyle of the humerus (see Fig. 1b). The electrodes' signal was acquired as described in Ref. [75], the signal processing followed Ref. [74], and the experimental software used the Python AxoPy library [76]. The root mean square (rms) value for each time-domain sample window was calculated, normalized by a manually set calibration constant (in mV), and incorporated into a 0.5 s moving average window to yield an updated and processed sEMG signal, $\bar{x}$ (mV/mV), at 16 Hz [used in Eq. (1)]. The subjects learned to manipulate their sEMG signals to produce serial patterns that translated to cursor motion.

After the electrode placement, the subjects viewed their sEMG signal on an oscilloscope shown on the computer screen and were instructed to flex their hand to induce signal changes. This oscilloscope activity confirmed proper electrode adhesion and illustrated the subjects' abilities to intentionally change their sEMG signal. For the

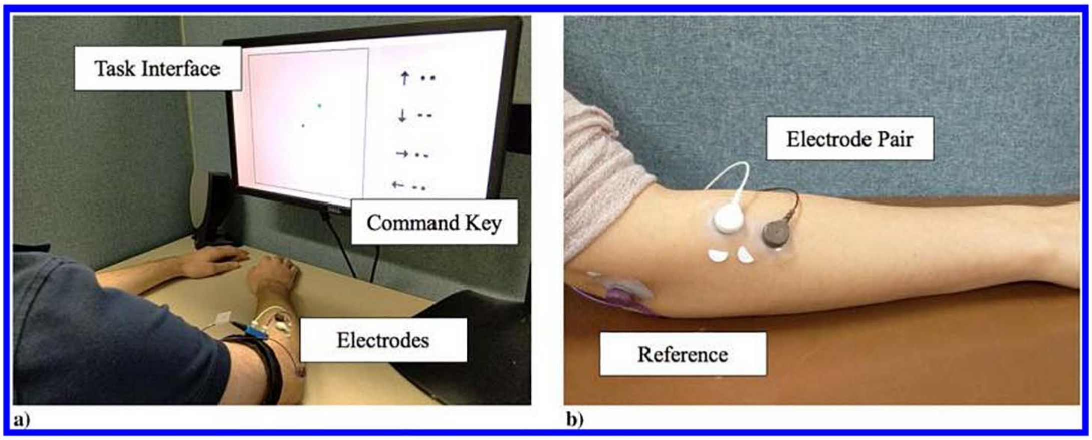

Fig. 1 Setup with interface displayed during training and testing of cursor-to-target task (a) and electrode placement (b).

http://arc.aiaa.org | DOI: 10.2514/1.I010915

Table 1 Summary of muscle activation strategies by group

	Group				
Strategy, N	Control	Concurrent feedback	Terminal feedback	Adaptive threshold	All
Flex wrist	7	10	9	8	34
Make fist	5	3	2	3	13
Raise finger	1	0	2	1	4
Raise multiple fingers	3	6	5	6	20
One strategy	9	8	8	8	33
Multiple strategies	3	4	4	4	15

remainder of the study, the subjects individually determined how to produce sufficient muscle activation and were prompted to try different, self-selected movements and contractions during the manual calibration. Subjects self-reported the muscle activation strategies they used at the end of the study. Flexing the wrist was the most common strategy ($N = 34$), possibly due to its introduction during the oscilloscope activity. Other strategies included making a fist, raising a single finger, or raising multiple fingers. The majority of subjects ($N = 33$) only used one strategy. Fifteen subjects elected to use multiple strategies, and five of those subjects used specific strategies for different actions (e.g., raising fingers for shorter inputs). A summary of the strategies used by experimental group is shown in Table 1. A more detailed discussion of the experimental groups is provided in a later section.

B. Cursor-to-Target Task

In our study, the Fitts's-law-style task served to provide an interactive, engaging environment in which the subjects developed their proficiency of sEMG commands by moving a cursor to hit a target. The Fitts's law task used a center-out paradigm and a similar interface as in Ref. [74]. The square cursor interface had normalized horizontal and vertical bounds of $[-1, 1]$ and a length of 2 units. Each trial began with the cursor at the center of the interface and a target in a pseudorandom position. There were 40 unique target positions that covered a range of index of difficulties (IDs) from 1.00 to 4.09 bits (calculated by the Shannon formulation [77]). The subject used sEMG to convey commands and moved the cursor to the target, and the cursor had to dwell on the target for 1 s to successfully complete the trial. The maximum trial time was 60 s, which was determined by reviewing previous data [74] and preliminary testing.

For sEMG control, it is necessary to designate a threshold below which the signal input is considered at rest because sEMG signals effectively always have nonzero values. The threshold l_1 (in mV/mV) defined the crossover value for a signal at "rest" versus active (see Fig. 3a). From our previous observations, cursor control improved when the initial motion had a slow, constant velocity before switching to a velocity proportional to the signal input, as it allowed subjects to maintain a slow velocity when desired (e.g., in close vicinity to a target). Therefore, we designated a second threshold, l_2 (in mV/mV), that delineated the constant from the proportional velocity control. The cursor was either stationary ($\bar{x} < l_1$); moving at a small, constant velocity ($l_1 < \bar{x} \leq l_2$); or moving at a velocity proportional to the input ($\bar{x} > l_2$). When the input exceeded l_2, the cursor velocity was calculated by Eq. (1), where v_c is the minimum velocity (0.10 units/s), v_m is the maximum velocity (0.50 units/s), l_2 is 0.30 mV/mV, and $\bar{x}$ is the filtered, averaged sEMG signal. The l_1 value was nominally 0.20 mV/mV, except for the adaptive threshold group in which the value was randomly selected for each trial ($l_1 = 0.10, 0.15, 0.20, 0.25, 0.30$ mV/mV).

$$v = v_c + (v_m - v_c)\left[(\bar{x} - l_2)/(1 - l_2)\right] \tag{1}$$

Because the subjects could only select one command at a time, the cursor could either move up, down, left, or right. It was not possible to combine commands, which restricted the cursor motion to a rectilinear trajectory. To illustrate the resulting cursor trajectories, Fig. 2 displays selected trials that are either within one standard deviation of the median (see Fig. 2a) or outside of one standard deviation (see Fig. 2b) for a path efficiency metric. Example trajectories are superimposed on a cursor position heat map of successful trials for the applicable target (ID = 4.09). The trajectories and heat maps are from across groups for the selected target position.

C. Command Design and Group Treatments

The premise of the sEMG control methodology was to use serial patterns of muscle activation ("inputs") to convey commands, similar to Morse code [74]. In our scheme, the first two inputs selected the command and a third input allowed for continuous control of forward movement in the selected direction (see Fig. 3a). A timeout between the first two inputs allowed a reset in the case of errors during command selection. The command inputs were defined by the duration that the sEMG signal exceeded the threshold l_1 and each input was identified as "short" ($\leq$0.50 s) or "long" ($>$0.50 s). For example, the combination of two short inputs selected the "up" command. Subjects learned the two-input code for four commands: up, down, left, and right. A portion of the user interface contained a command

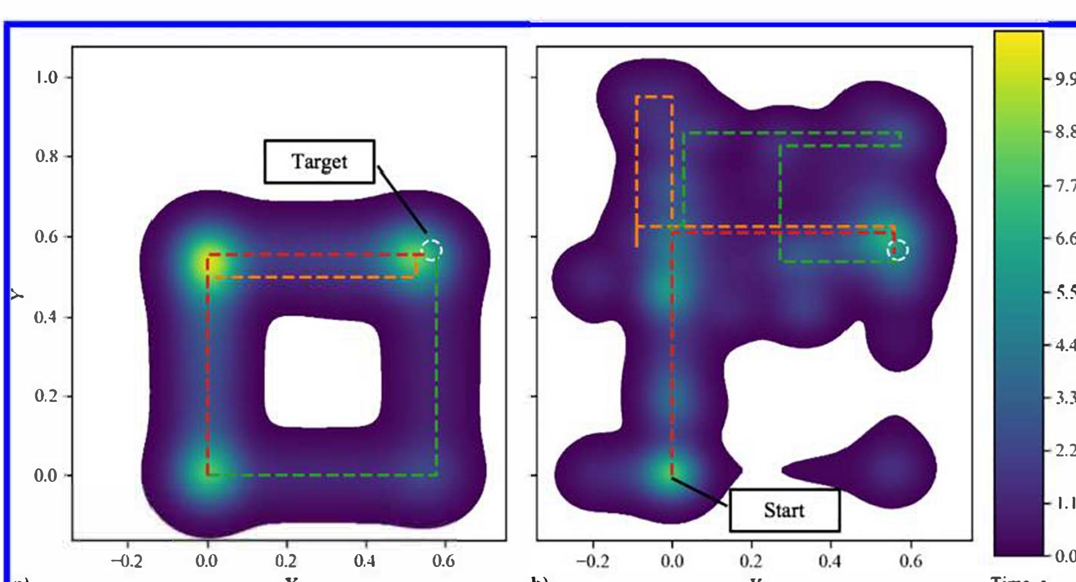

Fig. 2 Sample cursor trajectories for a target (ID = 4.09) either a) within or b) outside one standard deviation of the median for path efficiency.

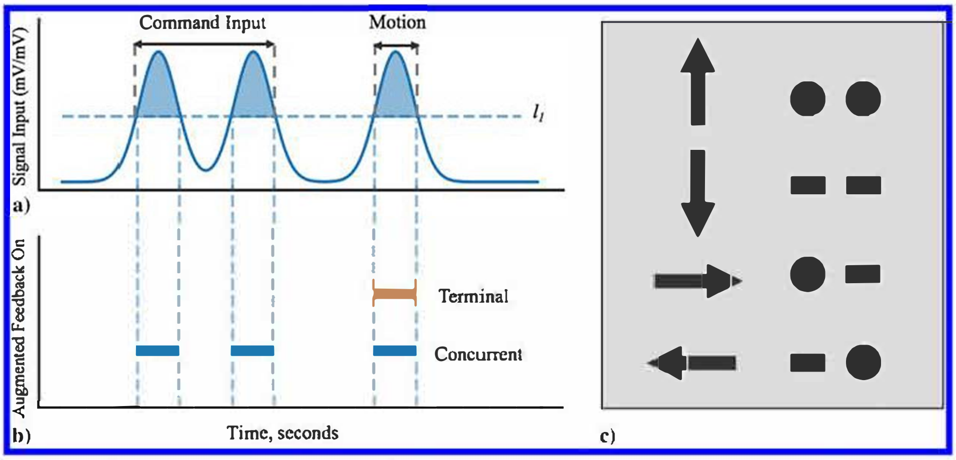

Fig. 3 Illustrative signal input (a) with the feedback timing (b) for applicable groups. The command key (c) is always present (dot = short, dash = long) [78]. Revised with permission from O'Meara et al. [78].

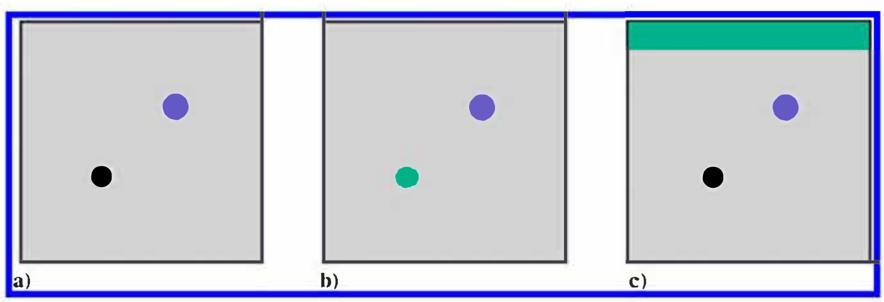

Fig. 4 Cursor interfaces for a) the control and adaptive threshold, b) concurrent feedback, and c) terminal feedback groups (not to scale) [78]. Reprinted with permission from O'Meara et al. [78].

key so that the subjects did not need to memorize the serial, sEMG patterns to produce commands (see Fig. 3c).

The four groups were each exposed to one training methodology:

1) The control group trained solely through task repetition and was able to view the motion of the cursor.

2) The concurrent feedback group received additional visual feedback during the task that indicated when the processed sEMG signal $\bar{x}$ exceeded the threshold $l_1 = 0.20$ mV/mV, which was important for selecting commands (see Fig. 3b). The subjects received feedback to confirm reception of the signal.

3) The terminal feedback group received visual feedback after entering a command (see Fig. 3b). The subjects received feedback regarding the interpretation of the signal to a command, and signal reception during motion.

4) The adaptive threshold group was the same visually as the control group, but the threshold l_1 was randomly selected at each trial among values between 0.10 and 0.30. In this group, the processed sEMG signal $\bar{x}$ had to exceed the threshold that changed each trial in order to input commands. These subjects trained with variable parameters/increased mapping uncertainty (i.e., whether their signal will cross the threshold) with the aim of improving motor skill acquisition.

As shown in Fig. 4, all of the groups viewed a very similar interface. Our early results in providing augmented feedback for manual control tasks found that visual augmented feedback worked best when placed on an element of the display that operators were already visually engaged with, and that visual feedback placed elsewhere would both lower performance and increase workload [79]. This takes advantage of operators' normal patterns of visual attention, which is "drawn to display items that are ..., colorful, and changing (e.g., blinking)" [80]. For this study, we sought to identify differences between providing feedback concurrently, every time subjects crossed l_1, and terminally, only after they had successfully input a command. Additionally, we selected the feedback colors to be colorblind agnostic. The additional visual feedback in the concurrent feedback group was provided by changing the color of the cursor (Fig. 4b). The terminal feedback group had additional visual elements at the edges of the cursor interface that changed color to reflect the command (Fig. 4c). The visual element locations were selected based on pilot testing. Both of the two augmented feedback techniques provided users with feedback when the system had received a valid input (see the colored regions for the timing shown in Fig. 3b), and the terminal feedback group received the additional feedback of the system's interpretation of the direction to move the cursor. A recent review of augmented feedback techniques (see Ref. [34]) found that terminal feedback is more effective for tasks with low functional complexity, whereas concurrent feedback is more effective for tasks with high functional complexity. A summary of the training type, displays, and visual feedback is provided in Table 2 (refer to Supplemental Storyboard S1 and Supplemental Video S1 for walkthrough of each interface). It should be noted that the subject did not have a choice to disuse or turn off the automated training methodology. For example, this is similar to the predetermined automation design of a space exploration autonomous behavioral health tool reported in [81].

Table 2 Summary of group treatments and displays

Group	Type	Display	Visual feedback timing
Control	Repetition	Fig. 4a	N/A
Concurrent feedback	Augmented feedback	Fig. 4b	Fig. 3b
Terminal feedback	Augmented feedback	Fig. 4c	Fig. 3b
Adaptive threshold	Motor adaptation	Fig. 4a	N/A

Subjects were assigned to one of four groups (12 subjects per group) and proceeded through the same protocol. The protocol consisted of the following general steps: consent, entrance survey, experiment overview and setup, pretraining assessments, training phase, post training assessments, evaluation phase, post evaluation assessments, and an exit survey. A maximum voluntary muscle contraction measurement was collected pretraining and post evaluation to confirm that subjects were not significantly fatigued during the experiment, and which has been shown to affect sEMG signals [22].

We developed a "command accuracy test" to assess each subject's ability to produce commands. The command accuracy test was conducted pretraining, post training, and post evaluation to provide an assessment similar to offline classification accuracy (see Ref. [82] for an EMG example). A Fitts's-law-style task was implemented for the training and evaluation phases. There were a total of 160 trials, with the first 120 trials used for training and the last 40 trials for evaluation. Trials were grouped into sets of 10 to form a block; there was a 30 s minimum rest time after each block and during which subjects completed surveys for workload and trust.

III. Analysis and Hypotheses

The analysis focused primarily on command and interaction (i.e., workload and trust) metrics to observe the effects of the automated training methodology. Task-based metrics were also analyzed. A brief description of the analysis is provided in the following subsections.

A. Command Metrics

The command accuracy test provided an opportunity to evaluate subjects' proficiency in producing desired commands outside of the Fitts's law task: pretraining, post training, and post evaluation. During the command accuracy test, subjects responded to a prompt to produce a specified command (up, down, left, or right) and performed each command 5 times for a total of 20 commands in a pseudorandomized order. The command was scored as successful when the input matched the prompt; inputs classified as any other command were not considered successful. The percent of successful commands was calculated for each subject and averaged within the group for a command accuracy metric. For all the successful commands, command time and average command amplitude were also calculated. The command time is the duration between the start and end of a command input (see Fig. 3a). The average command amplitude is the average value of the input above the threshold during the command input (i.e., the shaded regions in first two inputs of Fig. 3a); an average command amplitude close to zero represents an efficient input. Chhabra and Jacobs [83] used a similar signal magnitude metric to evaluate input proficiency. These metrics assessed the accuracy and efficiency of command input but were not calculated during the Fitts's law task because subject intention (i.e., desired command) would need to be inferred for most time points.

B. Interaction Metrics: Workload and Trust

Workload was measured using the Modified Bedford Workload scale [63]; the subject rated their perceived cognitive workload using a flow diagram. Trust was evaluated using Jian et al.'s 12-statement questionnaire, which measures trust between people and automated systems [84]. Each of the 12 statements was evaluated on a 7-point Likert-type scale. Subjects completed these surveys after each block in the training and evaluation phases. A block consisted of 10 trials, where a trial is a single Fitts's-law-style cursor-to-target task.

C. Cursor-to-Target Task Metrics

The task metrics from the training and evaluation phases included percent of successful trials, completion time, throughput (TP), and normalized path length (nPL). The percent of successful trials consisted of the number of successful trials out of the 10 trials within a block, and only successful trials contributed to the completion time, TP, and nPL metrics. Completion time was defined as the time from the start of the trial to the trial completion but did not include the 1 s dwell time. TP is a common metric for evaluating Fitts's law tasks and is defined as $TP = ID/MT$, where ID is the index of difficulty and MT is the movement time. Given the control scheme and the target positions, there was an optimal, rectilinear path that minimized the distanced traveled by the cursor to select the target. The nPL metric assessed the efficiency of the path traveled by using the optimal path distance to normalize each value. There were 10 unique optimal paths due to the radial symmetry of the target positions.

The percent of successful trials was evaluated every block. The 40 unique target positions repeated every four blocks (a.k.a. a session), so the completion time and nPL metrics were averaged over every four blocks. We found completion time and nPL to be sensitive to target location or ID, and therefore averaging over four blocks (i.e., the 40 unique target positions) was more representative of overall performance. TP is intended to evaluate an input device's efficacy under the assumption that the user is proficient. Therefore, TP was only calculated during the evaluation phase when subjects had reached a performance plateau.

D. Hypotheses

Based on our prior experience with augmented feedback and sEMG cursor control, we formed the following hypotheses:

1) The adaptive threshold group will have the highest command proficiency by the end, followed by the concurrent feedback and terminal feedback groups, and then the control. Varying parameters (i.e., thresholds) would improve sensitivity to the sEMG dynamics. Augmented feedback would provide better sensitivity to the sEMG dynamics than the control.

2) The concurrent feedback and terminal feedback groups will have a high level of trust during training with some slight decrease during evaluation. The control group's trust will continually increase. The adaptive threshold group will have lower trust during training, which will increase in the evaluation phase.

3) The workload will continually decrease during the training phase for all groups with the largest decreases for the concurrent feedback and terminal feedback groups. There will be no significant difference in workload in the evaluation phase for all groups because the training phase will be sufficiently long enough for all groups to proficiently learn to execute the Fitts's-law-style task.

4) During the training phase, the concurrent feedback group will have the highest performance followed by terminal feedback, then control, and finally the adaptive threshold groups. All groups will perform similarly in the evaluation phase because the training phase will be sufficiently long enough for all groups to proficiently execute the Fitts's-law-style task.

IV. Results

Subjects were evenly divided into four groups: control, concurrent feedback (visual feedback when $\bar{x} > l_1$), terminal feedback (visual feedback after command selected), and adaptive threshold (l_1 varied on a trial-by-trial basis). Subjects completed a total of 16 blocks, each comprised 10 cursor-to-target trials, for a total of 160 trials over the course of the study. When applicable, sets of four blocks were grouped into a session. Each session was identical, such that they contained the same sequence of pseudorandomized target positions. Muscle fatigue did not appear to alter the results, as the maximum voluntary contraction was not significantly different between the beginning and end of the study ($F(1, 46) = 2.554$, $p = 0.117$). One subject was removed from this analysis as they changed their arm position/contraction method during testing, which affects the signal.

We ran two-factor linear mixed effect models to investigate changes in command and task performance, workload, and trust with one between-subjects factor, group, and one within-subjects repeated measure, block/session/test. When significant effects were observed, post hoc comparisons using the Tukey honest significance difference (HSD) test were performed and considered significant at the $p < 0.05$ level, and the Satterthwaite method was used to calculate the adjusted degrees of freedom using the lmerTest package in R [85].

A. Command Metrics

The results in this section (command accuracy, command time, average command amplitude) were assessed from the command accuracy test. The command accuracy test occurred three times (before training, after training, and after evaluation), and the average was calculated within a group for each test. We did not find evidence that analyzing results by command type (i.e., up, down, left, or right) altered the overall findings.

Subjects responded to command prompts in each command accuracy test. The command accuracy indicates the percentage of the 20 prompts in each test that subjects correctly performed (see Fig. 5). There was a significant main factor of test ($F(2, 88) = 108.485$, $p < 0.001$), but group was not significant ($F(3, 44) = 2.631$, $p = 0.062$). The interaction effect between group and test was not significant ($F(6, 88) = 0.826, p = 0.553$). Investigation into the test variable showed that subjects performed significantly better between tests 1 and 2, and between tests 1 and 3, but not between tests 2 and 3. These results demonstrated that there was a significant improvement in the percent of accurate commands by the end of the training phase.

The command time and average command amplitude metrics were calculated only for successful commands. For command time, there was a significant factor of test ($F(2, 2055.358) = 9.417$, $p < 0.001$), but group ($F(3, 44.080) = 0.662, p = 0.580$) and the interaction between group and test ($F(6, 2054.727) = 1.005, p = 0.420$) were not significant. Examining the test variable, the command times significantly improved from test 1 to test 2 and test 1 to test 3, but not from test 2 to test 3 (see Fig. 6). The average command times across groups were 1.81, 1.69, and 1.63 s in order of test.

Average command amplitude had a significant interaction factor of group and test ($F(6, 2044.772) = 17.143, p < 0.001$), and main factor of test ($F(2, 2044.882) = 13.886$, $p < 0.001$), but not group ($F(3, 44.211) = 1.062, p = 0.375$). The only significant group and test interactions occurred in test 1 between the control and adaptive threshold groups and the concurrent feedback and adaptive threshold groups. Notably, the adaptive threshold group's average command amplitude increased from test 1, whereas all other groups generally declined (see Fig. 7). An average command amplitude close to zero represents an efficient input.

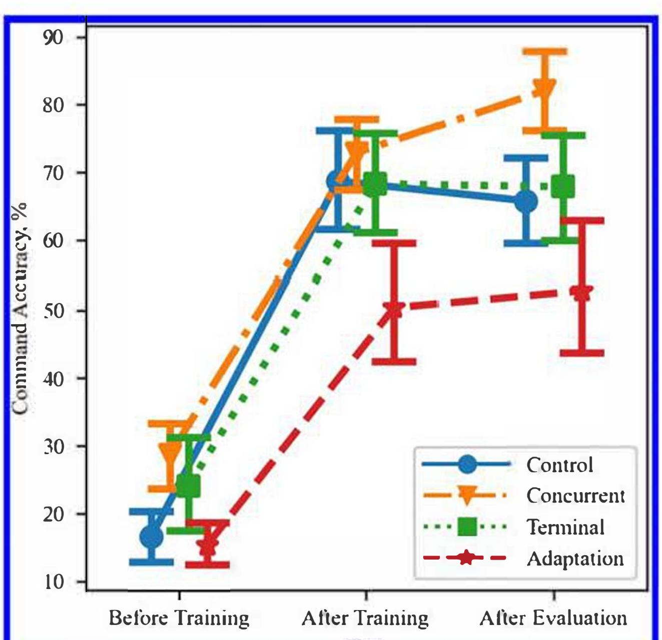

Fig. 5 **Command accuracy results. Tests 1, 2, and 3 occurred before the training phase, after the training phase, and after the evaluation phase, respectively [78]. Revised with permission from O'Meara et al. [78].**

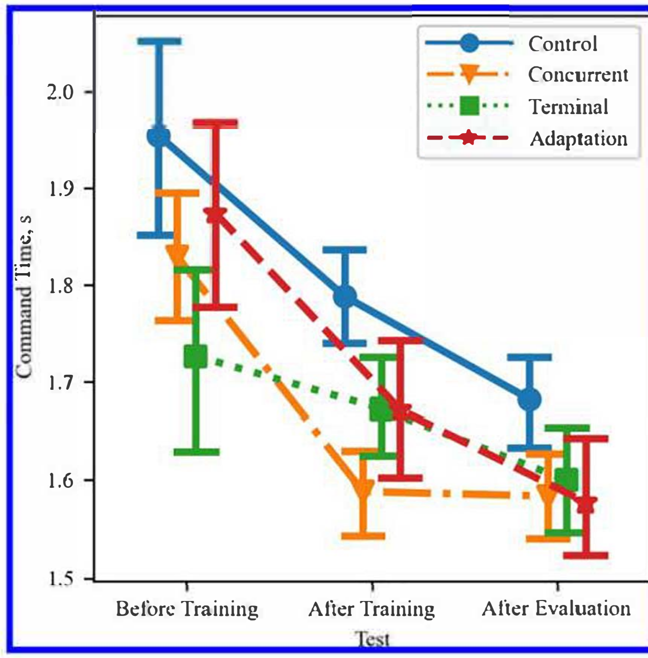

Fig. 6 **Command time results. Tests 1, 2, and 3 occurred before the training phase, after the training phase, and after the evaluation phase, respectively.**

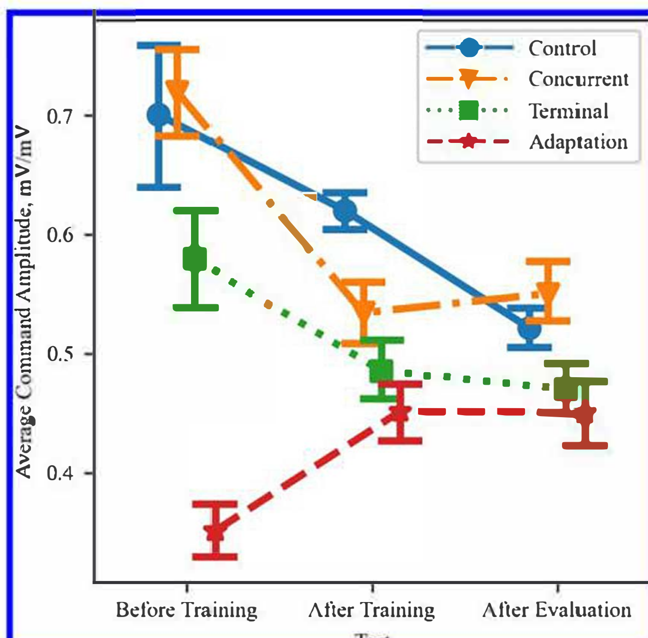

Fig. 7 **Average command amplitude. Tests 1, 2, and 3 occurred before the training phase, after the training phase, and after the evaluation phase, respectively.**

B. Interaction Metrics: Workload and Trust

The study assessed the changes in workload: an important consideration for human–automation interaction. The Modified Bedford Workload metric is a subjective measurement of workload that ranges from 1 to 10, where 1 indicates low workload and 10 indicates high workload. There was a significant main factor of block ($F(15, 660) = 18.284, p < 0.001$), but group was not found to be significant ($F(3, 44) = 2.164, p = 0.106$). There was also a significant interaction effect between group and block ($F(45, 660) = 1.818, p = 0.001$) (see Fig. 8). The interaction effect resulted from subjects reporting lower workload as they learned the task at different rates, indicated by the block factor. In further investigation of the interaction, we observed that the adaptive threshold group reported a

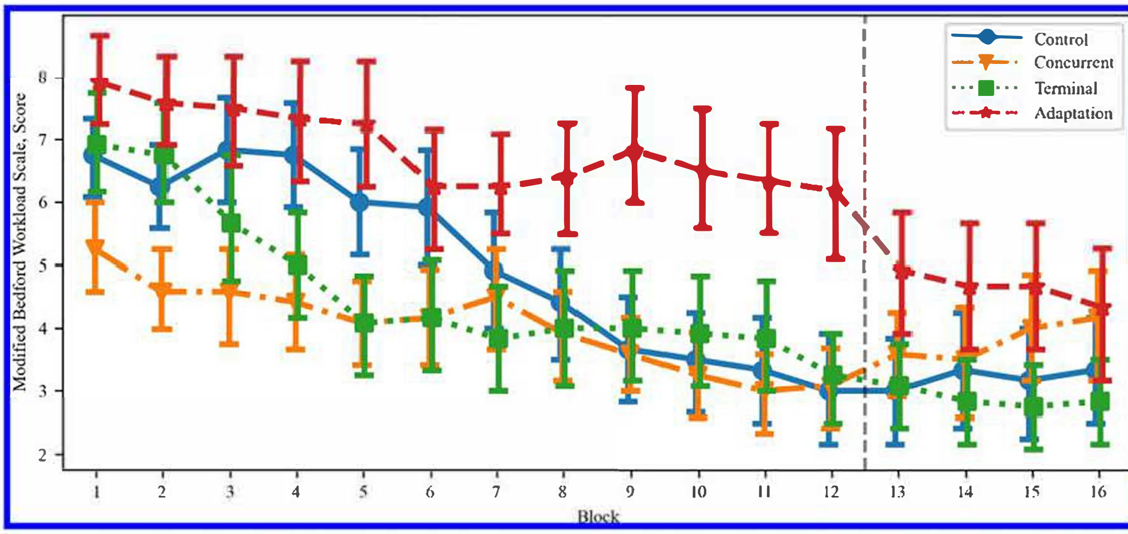

Fig. 8 Modified Bedford Workload Score. The vertical dashed line represents the transition from training to evaluation. Error bars shown are the standard error of the mean [78]. Revised with permission from O'Meara et al. [78].

significantly higher workload than the concurrent feedback group for blocks 9, 10, and 11. This perception of high workload may have resulted from the challenges and uncertainties of the changing threshold. None of the groups showed a significant difference in workload compared with the control group, and all four groups reported statistically similar workloads in the evaluation phase.

Intragroup changes in workload were also of interest. The concurrent feedback group showed no statistically significant changes in performance between blocks, though they did demonstrate a nonsignificant, increasing workload trend in the evaluation phase of the experiment. The terminal feedback group had statistically higher initial workload for blocks 1, 2, and 3, but the remainder of the blocks had a statistically similar level of workload. The control group's workload was significantly higher for blocks 1–6 but leveled off for the remainder of the trials. Finally, the adaptive threshold group reported the highest workload in blocks 1–5, but also saw the largest improvement transitioning into the evaluation phase where their threshold l_1 stabilized to the same fixed value as the other groups. The second human–automation interaction factor evaluated in the study was trust. Trust was measured using Jian et al.'s 12-question trust survey [84]. Each question was rated on a seven-point Likert scale, the five reverse-coded questions were reversed, and the results were averaged to create a single trust score (see Fig. 9). There was a significant main factor of block ($F(15, 660) = 13.051$, $p < 0.001$), but group was not found to be significant ($F(3, 44) = 2.588$, $p = 0.065$). There was also significant interaction effect between group and block ($F(45, 660) = 1.996$, $p < 0.001$). The significant main effect of block shows a gradual increase in trust throughout the duration of the study. After investigating the interaction effect, we saw that no group reported a significantly different trust level than the control group on any block, but that the adaptive threshold group recorded a significantly lower trust than the concurrent feedback group on blocks 3–6 and 9. Similar to workload, the primary interaction effects appeared driven by intragroup differences. The concurrent feedback group showed no statistically significant changes through the study, the terminal feedback and control groups displayed significant increases in trust in blocks 1–6, and the adaptive threshold group reported significantly higher trust in the evaluation session than during early blocks. We investigated the ratings for each individual question in the trust survey but did not observe any trends that contradicted the overall trust score patterns shown in Fig. 9.

C. Task Metrics

Although the command metrics were the primary proficiency indicator, we did evaluate aspects of task performance, including percent of successful trials, throughput (TP), completion time, and normalized path efficiency (nPL). The percent of successful trials metric measured the percentage of successfully completed trials within a block. For percent of successful trials, there were significant

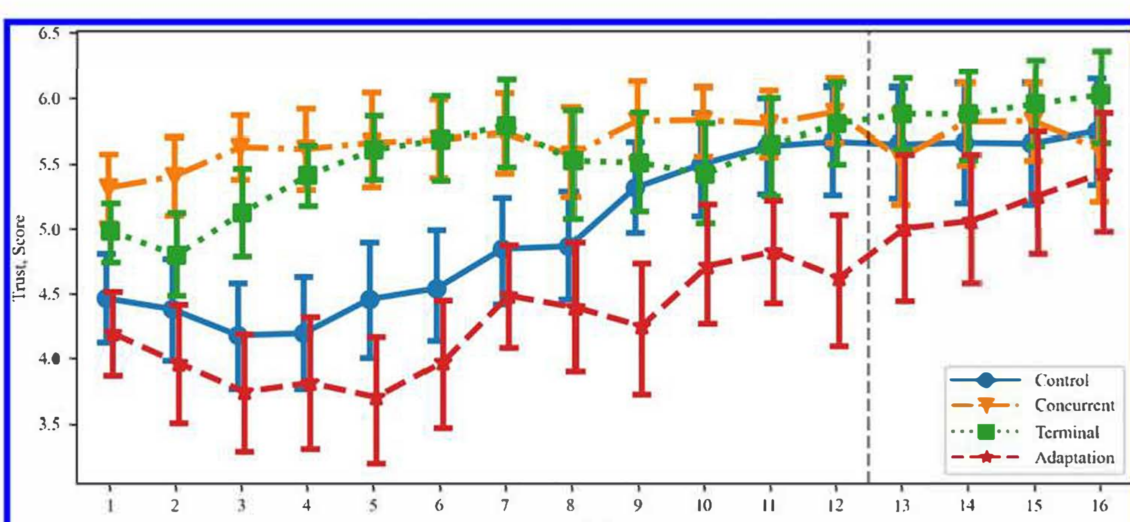

Fig. 9 Trust score. The vertical dashed line represents the transition from training to evaluation. Error bars shown are the standard error of the mean [78]. Revised with permission from O'Meara et al. [78].

main factors of group ($F(3, 44) = 8.183$, $p < 0.001$) and block ($F(15, 660) = 31.805$, $p < 0.001$). There was also a significant interaction effect between group and block ($F(45, 660) = 3.903$, $p < 0.001$). Despite the presence of an interaction effect that resulted from subjects learning the task (as indicated by the block factor), the main effect of group could still be interpreted. A Tukey test showed that the subjects in the groups differed significantly, with subjects in the concurrent feedback group performing significantly better than those in the control group ($p = 0.020$). The Tukey test also showed that subjects in the adaptive threshold group performed significantly worse than those in the terminal feedback and concurrent feedback groups ($p = 0.009, < 0.001$, respectively), but not worse than the control ($p = 0.394$).

The interaction effect resulted from different learning rates between the groups (see Fig. 10), where subjects learned in the following order (fastest to slowest): concurrent feedback, terminal feedback, control, and adaptive threshold. Compared with the control group, the concurrent feedback group significantly outperformed them for the first six blocks. Unlike the concurrent feedback group, the terminal feedback and adaptive threshold groups started with the same initial performance as the control group. The terminal feedback group learned more quickly than the control group, however, and significantly outperformed the control group for blocks 4 and 5. Compared with the control group, all groups performed at statistically similar level after block 6. Investigating the immediate evaluation effects when the group-specific treatments were removed in block 13, the mixed model showed no change in performance for any of the groups ($p > 0.989$ for all groups). As such, the percentage of successfully completed trials did not show any effect from the guidance hypothesis (i.e., the subjects did not rely on the feedback to complete the task and removing the feedback did not result in decreased performance).

The TP was calculated for the evaluation phase and averaged across blocks 13–16 (i.e., fourth session). Throughput is generally used to assess an input device, which should be measured when the subjects can complete the task. Because there were no significant differences in the evaluation phase for percent complete, we only calculated TP at this time, since initial blocks may have been biased toward low IDs. There was no significant difference in throughput between the groups ($F(3, 44) = 1.625$, $p = 0.197$). The mean throughput for all subjects was found to be 0.564 ± 0.023 bits/s ($\mu \pm \sigma$).

The pseudorandomization of the 40 unique target positions occurred over four blocks; thus it seemed appropriate to average completion time over a session (i.e., set of four blocks). Completion time was only defined for successfully completed trials; the results are displayed in Fig. 11. There were significant main factors of group ($F(3, 43.968) = 4.390$, $p = 0.009$) and session ($F(3, 131.074) = 24.906$, $p < 0.001$). The interaction effect between group and session was not significant ($F(9, 131.069) = 0.782$, $p = 0.633$). A Tukey test showed that the concurrent feedback group performed significantly better than those in the adaptive threshold group ($p = 0.004$), which was the only significant difference between groups, and no groups significantly outperformed the control group. Analysis of the session factor showed increased performance ($p < 0.05$) until the last two sessions, which were not statistically different ($p = 0.646$). These results further support that the guidance hypothesis did not occur.

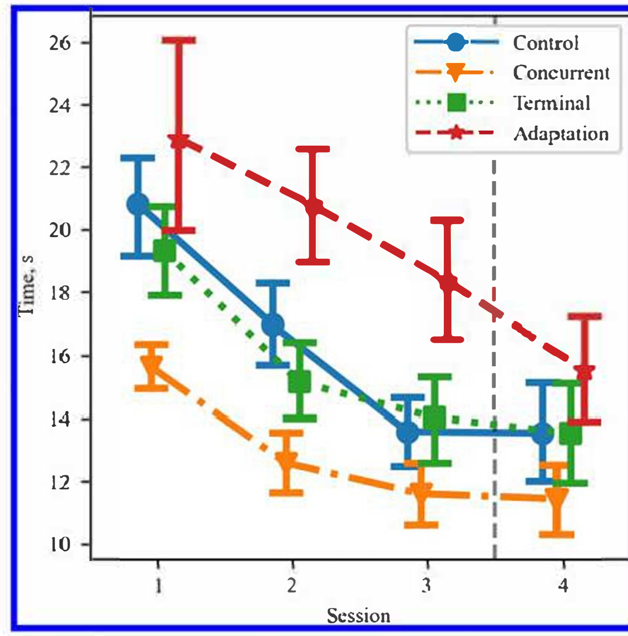

Fig. 11 Completion time by session (set of 4 blocks) across groups. The vertical dashed line represents the transition between the training and evaluation phases [78]. Revised with permission from O'Meara et al. [78].

The significant difference in completion time between the concurrent feedback and adaptive threshold groups may have been explained by increased normalized path length, but there were no significant group differences ($F(3, 44.099) = 0.908$, $p = 0.445$). Increased normalized path length correlated with lower ID targets (see Table 3).

V. Discussion

The study investigated the effects of automated training methodologies to efficiently and accurately use an sEMG command system, and elucidate the relationship between performance, workload, and

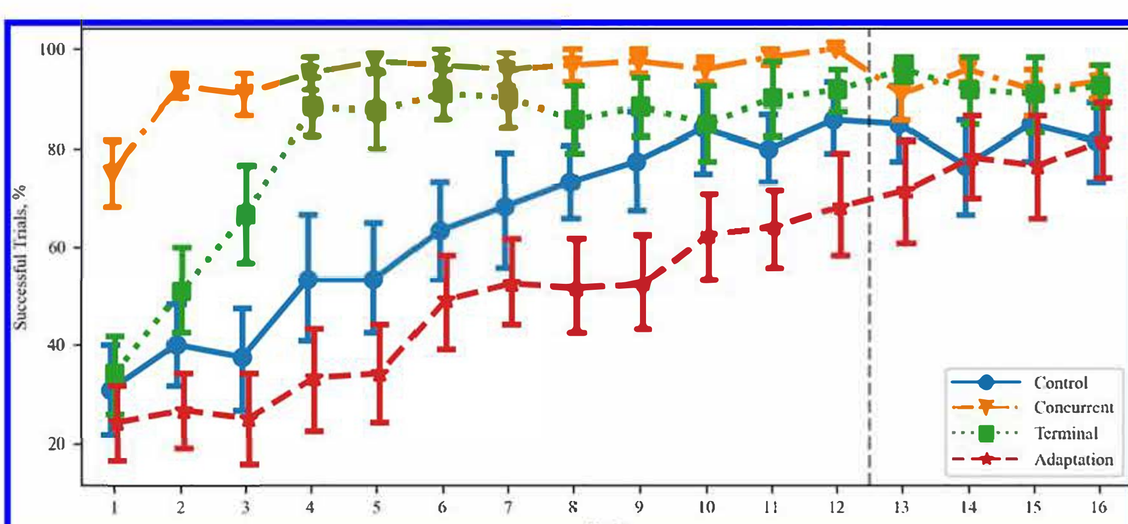

Fig. 10 Percent of successful trials. The vertical dashed line represents the transition from training to evaluation. Error bars shown are the standard error of the mean [78]. Revised with permission from O'Meara et al. [78].

Table 3 Normalized path efficiency for different targets

Index of difficulty	1.00	1.58	2.32	3.17	4.09
Average	32.11	3.80	1.80	1.38	1.20
Standard error of the mean	2.90	0.29	0.08	0.04	0.02

trust during training. The command metrics from the command accuracy test largely did not indicate group differences. We expected the groups to perform similarly during test 2, which was administered at the conclusion of the training phase and when the subjects should have reached a performance plateau at the end of training. However, in test 3 we expected that the adaptive threshold group would optimize their inputs to the stabilized threshold, but we did not observe significant differences in test 3. We did not expect to observe differences between groups in test 1, but there was a significant difference in average command amplitude between the control and adaptive threshold and concurrent feedback and adaptive threshold. Before test 1, the subjects across groups had received a generic experiment introduction, group-specific training tailored to their different interfaces, and a manual calibration of their signals. After investigating the data from test 1, we concluded that the differences may be a result of the group-specific training and the manual calibration. The manual calibration for the adaptive threshold group had to accommodate the range of thresholds ($0.1 \leq l_1 \leq 0.3$), which increased the normalization value by approximately 24% higher than the other groups. As a result, the normalized signal input $\bar{x}$ decreased by the same proportion for the same rms input value. Combined with the standard threshold value ($l_1 = 0.2$) during all code accuracy tests, the adaptive group would have needed to input larger signals to compensate. There are no significant differences after test 1, during the evaluation phase, or in the MVC (which would indicate fatigue); therefore the effect seems to be limited to pretraining. Their average command amplitude values did increase for test 2 and test 3 compared with test 1. Tests 2 and 3 occurred after the Fitts's law task when the adaptive threshold group would have seen the effect of their input on the cursor motion. The adaptive threshold group did not pursue a strategy of maximally contracting their muscles evident by their convergent average command amplitude values with the other groups, particularly the concurrent feedback group, which had feedback on threshold crossings. Our subjects' adaptation follows other studies that observed adaptation to input noise characteristics [83] and the refinement of muscle synergies [30]. The subjects in this study improved across the command metrics after the training phase, demonstrating adaptation to the sEMG command system. In terms of these command metrics, the specific automated training methodology was less important.

The adaptive threshold group had increased mapping uncertainty during training, whereas the concurrent feedback and terminal feedback groups received augmented feedback about the sEMG control system. Augmented feedback was more effective in assisting subjects in achieving a performance plateau, requiring only a third of the prescribed training phase and following the findings of Basmajian [51]. The control and adaptive threshold groups needed the entire training phase to achieve a similar performance level, indicating that the augmented feedback enhanced the early acquisition of motor learning skills. The early motor learning we observed, where the concurrent group led the terminal group in initial performance gains, was expected as this sEMG testbed was designed to investigate a challenging task. Functionally complex motor control tasks are generally expected to respond better to concurrent feedback [34], though it remains difficult to estimate a task's functional task complexity, or instances where concurrent versus terminal feedback will be more effective. Both feedback levels eventually resulted in the same performance, but the high levels of performance initially observed in the concurrent group, both likely resulted from the subject's increased internal understanding of the task dynamics. The augmented feedback groups appear to have the strongest performance advantage.

However, the adaptive threshold group's results are interesting for two reasons. First, varying thresholds/increased mapping uncertainty did not appear to cause adverse effects compared with the control group. Unreliable automation behavior can lead to poor human–automation interaction [86], but was not the case in this study. Secondly, the adaptive threshold group only trained at the standard threshold ($l_1 = 0.2$) for 20% of the trials compared with the other groups. The changing threshold did affect task metrics during the training phase; the varying threshold levels were not inconsequential as evident by the significant differences between the adaptive threshold and concurrent feedback groups for workload, trust, percent of successful trials, and completion time. However, the adaptive threshold group's performance was indistinguishable during the evaluation. The training challenge of varying thresholds/increased mapping uncertainty did not cause poorer performance during evaluation, but also did not see significantly improved performance. It is possible that the adaptive training methodology could be redesigned to better elicit the benefits. Alternatively, we could have evaluated all groups at a novel threshold. To observe potential task generalization benefits, this methodology may be better tested with a different task that still uses the sEMG commands (a.k.a. a transfer task) or a similar underlying task structure, instead of returning to a stable condition. For example, Braun et al. [55] showed that subjects trained with varying visuomotor tasks were able to quickly generalize to novel tasks that still retained similar features as the original training task. In one sense, the adaptive threshold group does have better command proficiency due to similar performance with less training at the standard threshold, which is encouraging although not strong support for the first hypothesis. We maintain that the subsequent experiments should continue investigating the use of adaptive strategies for training.

The task results suggest that subjects reached a similar level of task proficiency by the end of the training phase, as we observed no differences between groups during the evaluation phase. However, a more complex interaction between task performance, workload, and trust occurred during the training phase and varied by training methodology. The concurrent feedback group started with the highest task performance, lowest workload, and highest level of trust. By the fifth block, the terminal feedback group overlapped with the concurrent feedback group in terms of percent of successful trials, workload, and trust. These two groups then tracked each other for the remainder of the study. In both augmented feedback groups, our feedback displays act as an expert "instructor" model, providing information to subjects in real-time about either $\bar{x}$ or their command state. Correspondingly, the results we observed closely align with prior literature when subjects are provided with expert feedback; subjects in both augmented feedback groups displayed reduced "cognitive load during initial practice, helping [them] integrate declarative knowledge with physical skills" [37]. The concurrent feedback and terminal feedback groups did not have significantly different average completion times throughout the study. In their initial learning, the concurrent feedback group demonstrated high initial performance with smaller, incremental gains, whereas the terminal feedback group had large block-to-block improvements. The adaptive threshold group did not appear to reach a performance plateau for the percent of successful trials and steadily, but not significantly, improved in subsequent evaluation blocks. Interestingly, the trust score also continually increased for both the control and adaptive threshold groups. Although the changing cursor dynamics in the adaptive threshold group does not seem to have adversely affected trust compared with the control group, the adaptive threshold group did report a significantly higher workload for blocks 9–11 than the concurrent feedback group. Once the threshold stabilized, the workload and trust in the adaptive threshold group was not statistically different from the other groups. These effects that may be explained by Hoff and Bashir's three-layer model [70], where the training methodology altered situational trust and the continued interaction affected learned trust. Jian et al.'s trust survey specifically asks subjects about their familiarity with and confidence in the system, and asks subjects to rate the reliability of the system and its outputs [84]. Changes in trust scores may reflect differences in components like confidence and reliability, but also include a question for trust. Analysis of the individual questions did not contradict the averaged results of the trust scale, however, and it likely that this reflects that learned trust, not simply understanding of the system, is increasing with time. Overall, these results suggested that augmented feedback led to earlier task performance gains with improvements in trust and workload. Surprisingly, the varying thresholds/increased mapping uncertainty did not adversely affect trust, but the mental cost was reflected in the task performance and workload. In general, the second, third, and fourth hypotheses were largely supported with increased trust and decreased workload by the evaluation phase, with no differences between groups.

The workload scores indicated that the task of learning a novel sEMG cursor control system to hit targets on a screen was sufficiently challenging. The average workload scores by block across groups ranged from 6.71 to 3.65 with the highest average score occurring in the first block. The average workload score decreased with subsequent blocks. According to the Modified Bedford Workload Scale [63], a score of 7 indicated that the task was possible, but that there was "... *minimal* spare time for additional tasks." A score of 6 meant that the task was possible and workload tolerable and "there was *some but not enough* spare time available for additional tasks" [63]. By the end of training, the average workload score was rated approximately at a 4, as the task was considered possible with tolerable workload and "... *ample time* to attend to additional tasks" [63]. In a simulated, four-degree-of-freedom manually controlled spacecraft inspection task with a secondary task and verbal callouts, the average Modified Bedford Workload scores ranged from 7.17 to 3.94 [60]. The similarity in the workload scores indicates that the subjects found the task to learn sEMG control to be sufficiently difficult until the end of training. We intended for the subjects to become proficient in sEMG cursor control by the end of training. Therefore, for an initial evaluation of automated training methodologies, we believe the task complexity to be sufficient. In future development of sEMG control for aerospace applications, it would be beneficial to include additional tasks to simulate a more realistic environment.

There were not many active BoMIs for aerospace applications available for comparison and with similar metrics. The code accuracy reported in this study, 67.2% in test 3, appeared similar to the classification accuracy of a BCI mouse at 73.1% [19]. The percent of successful trials during the evaluation phase (86.18%) exceeded that of a BCI mouse for a simplified, simulated spacecraft (66.7%) [18], and was similar to the success rate for crossing gates in gesture-based drone control (87.67%) [13]. However, the differences in the tasks and requirements for success limit the extent and interpretability of the comparisons. As another point of comparison, the throughput values during the evaluation phase for all groups fell within previously published results for sEMG cursor control systems. Our previous single-site sEMG cursor control system with 2 degrees of freedom (counterclockwise rotation and forward) reported 2.24 and 0.23 bits/s for control methodologies that used different levels of automation [74]. Multisite systems have achieved 0.4 [87], 0.84 [88], and 1.3 bits/s [89]. The sEMG cursor control system used in this study had a throughput of 0.56 bits/s and may be of additional interest to the BCI community. However, the purpose of the sEMG cursor control system in this study was primarily to provide a testbed for automated training that lent itself to motor learning adaptation and was sufficiently challenging to probe the relationship between command and task performance, workload, and trust.

VI. Conclusions

The study results largely supported our hypotheses. The percent of successful trials performance during the training phase followed the order of concurrent feedback, terminal feedback, control, and adaptive threshold, but not for all times during that phase. All groups performed similarly in percent of successful trials during the evaluation phase. The subjects' trust followed our expectations, with concurrent feedback and terminal feedback having the highest levels, whereas the control group's trust continually increased at a slow rate. The adaptive threshold group had lower trust during training, and the trust increased to the level of the other groups during evaluation. The workload results also supported our hypotheses that concurrent feedback and terminal feedback groups would have the largest decrease in workload, and that all groups would have similar workload during the evaluation phase. Interestingly, the concurrent feedback and terminal feedback groups converged across performance, workload, and trust by block 5. Our hypothesis that the adaptive threshold group would achieve the highest command proficiency was not entirely supported, although we observed encouraging indications. In future studies it may be of interest to include a transfer task or conduct the command accuracy test throughout the training phase to improve understanding of the adaptive methodology. This study provided insights on the relationship between performance, workload, and trust for various automated training methodologies, and highlighted the advantage of certain methodologies during the training phase. By directly comparing methodologies and identifying these human–automation interaction effects, this research provides a quantitative basis for exploring combined training approaches in future research. It would be interesting to investigate a mixed training methodology approach that combines concurrent feedback with the motor adaptation approach, where subjects potentially benefit from both increased feedback, adaptation rate, and skill generalization. This research also contributes to training users of BoMIs, but it is not known if the study results generalize to BoMIs with other signal types. The muscle activation patterns were mapped to cursor commands, which were relevant to the task, but could be translated to other commands as needed for an application, and sEMG control is not limited to an on-screen application. The availability of multiple communication modalities provides flexibility and the opportunity to create new interaction paradigms.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1934792. This work was supported by a Space Technology Research Institutes grant from NASA's Space Technology Research Grant Program. Additional support was provided by the Link Foundation (Link Fellowship in Link Foundation Modeling, Simulation, and Training) and Zonta International Foundation (Amelia Earhart Fellowship). The authors would like to thank the subjects who participated in the experiment.

References

[1] Casadio, M., Ranganathan, R., and Mussa-Ivaldi, F. A., "The Body-Machine Interface: A New Perspective on an Old Theme," *Journal of Motor Behavior*, Vol. 44, No. 6, 2012, pp. 419–433. https://doi.org/10.1080/00222895.2012.700968

[2] Duprès, A., Roy, R. N., Scannella, S., and Dehais, F., "Pre-Stimulus EEG Engagement Ratio Predicts Inattentional Deafness to Auditory Alarms in Realistic Flight Simulator," *3rd International Mobile Brain/Body Imaging Conference (MOBI 2018)*, Berlin, Germany, 2018, pp. 114–115.

[3] Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., and Roy, R., "Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Inst. of Electrical and Electronics Engineers, New York, 2018, pp. 544–549. https://doi.org/10.1109/SMC.2018.00102

[4] Gateau, T., Ayaz, H., and Dehais, F., "In Silico Versus over the Clouds: On-the-Fly Mental State Estimation of Aircraft Pilots, Using a Functional near Infrared Spectroscopy Based Passive-BCI," *Frontiers in Human Neuroscience*, Vol. 12, May 2018, p. 187. https://doi.org/10.3389/fnhum.2018.00187

[5] Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohneiser, O., Kramer, J., Aricò, P., Borghini, G., Babiloni, F., Bagassi, S., and Piastra, S., "Brain–Computer Interface-Based Adaptive Automation to Prevent Out-of-the-Loop Phenomenon in Air Traffic Controllers Dealing with Highly Automated Systems," *Frontiers in Human Neuroscience*, Vol. 13, Sept. 2019, p. 296. https://doi.org/10.3389/fnhum.2019.00296

[6] Choi, B., Kim, D., Kim, M., and Hwang, S., "Effectiveness of EMG in Development of G-Induced Loss of Consciousness (G-LOC) Warning System," *Proceedings of the International Conference on Biomedical Engineering and Systems*, Avestia, Orleans, ON, Canada, 2014,

pp. 129-1–129-4, https://avestia.com/proceeding/proceedings-of-the-international-conference-on-biomedical-engineering-and-systems-icbes14/.

[7] Kim, S., Cho, T., Lee, Y., Koo, H., Choi, B., and Kim, D., "G-LOC Warning Algorithms Based on EMG Features of the Gastrocnemius Muscle," *Aerospace Medicine and Human Performance*, Vol. 88, No. 8, 2017, pp. 737–742.
https://doi.org/10.3357/AMHP.4781.2017

[8] Harrison, M. F., Neary, J. P., Albert, W. J., Veillette, D. W., McKenzie, N. P., and Croll, J. C., "Helicopter Cockpit Seat Side and Trapezius Muscle Metabolism with Night Vision Goggles," *Aviation, Space, and Environmental Medicine*, Vol. 78, No. 10, 2007, pp. 995–998.
https://doi.org/10.3357/ASEM.2067.2007

[9] de Oliveira, C. G., and Nadal, J., "Back Muscle EMG of Helicopter Pilots in Flight: Effects of Fatigue, Vibration, and Posture," *Aviation, Space, and Environmental Medicine*, Vol. 75, No. 4, 2004, pp. 317–322.

[10] Hewson, D. J., McNair, P. J., and Marshall, R. N., "Aircraft Control Forces and EMG Activity: Comparison of Novice and Experienced Pilots During Simulated Take-Off and Landing," *Aviation, Space, and Environmental Medicine*, Vol. 70, No. 8, 1999, pp. 745–751.

[11] Jorgensen, C., Wheeler, K., Stepniewski, S., and Norvig, P., "Bioelectric Control of a 757 Class High Fidelity Aircraft Simulation," *Proceedings of the World Automation Congress*, 2000, pp. 11–16, https://ntrs.nasa.gov/citations/20010083844.

[12] Macchini, M., Havy, T., Weber, A., Schiano, F., and Floreano, D., "Hand-Worn Haptic INTERFACE for Drone Teleoperation," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, New York, 2020, pp. 10,212–10,218.

[13] Miehlbradt, J., Cherpillod, A., Mintchev, S., Coscia, M., Artoni, F., Floreano, D., and Micera, S., "Data-Driven Body–Machine Interface for the Accurate Control of Drones," *Proceedings of the National Academy of Sciences*, Vol. 115, No. 31, 2018, pp. 7913–7918.
https://doi.org/10.1073/pnas.1718648115

[14] Macchini, M., Schiano, F., and Floreano, D., "Personalized Telerobotics by Fast Machine Learning of Body-Machine Interfaces," *IEEE Robotics and Automation Letters*, Vol. 5, No. 1, 2019, pp. 179–186.
https://doi.org/10.1109/LRA.2019.2950816

[15] Menshchikov, A., Lopatkin, D., Tsykunov, E., Tsetserukou, D., and Somov, A., "Realizing Body-Machine Interface for Quadrotor Control Through Kalman Filters and Recurrent Neural Network," *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vol. 1, Inst. of Electrical and Electronics Engineers, New York, 2020, pp. 595–602.
https://doi.org/10.1109/ETFA46521.2020.9212138

[16] Summerer, L., "Thinking Tomorrows' Space–Research Trends of the ESA Advanced Concepts Team 2002–2012," *Acta Astronautica*, Vol. 95, Feb. 2014, pp. 242–259.
https://doi.org/10.1016/j.actaastro.2013.11.002

[17] Broschart, M., de Negueruela, C., Millán, J. D. R., and Menon, C., "Augmenting Astronaut's Capabilities Through Brain-Machine Interfaces," *Proceedings of the 20th International Joint Conference on Artificial Intelligence, Workshop on Artificial Intelligence for Space Applications*, Morgan Kaufmann Publ., San Francisco, CA, 2007, https://dl.acm.org/doi/proceedings/10.5555/1625275#sec Authors.

[18] Poli, R., Cinel, C., Matran-Fernandez, A., Sepulveda, F., and Stoica, A., "Towards Cooperative Brain-Computer Interfaces for Space Navigation," *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, Assoc. for Computing Machinery, New York, 2013, pp. 149–160.
https://doi.org/10.1145/2449396.2449417

[19] Millàn, J. D. R., Ferrez, P. W., and Seidl, T., "Validation of Brain–Machine Interfaces During Parabolic Flight," *International Review of Neurobiology*, Vol. 86, Jan. 2009, pp. 189–197.
https://doi.org/10.1016/S0074-7742(09)86014-5

[20] Coffey, E. B., Brouwer, A.-M., Wilschut, E. S., and van Erp, J. B., "Brain–Machine Interfaces in Space: Using Spontaneous Rather than Intentionally Generated Brain Signals," *Acta Astronautica*, Vol. 67, Nos. 1–2, 2010, pp. 1–11.
https://doi.org/10.1016/j.actaastro.2009.12.016

[21] Scheme, E., Fougner, A., Stavdahl, Ø., Chan, A. D., and Englehart, K., "Examining the Adverse Effects of Limb Position on Pattern Recognition Based Myoelectric Control," *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Inst. of Electrical and Electronics Engineers, New York, 2010, pp. 6337–6340.
https://doi.org/10.1109/IEMBS.2010.5627638

[22] Barszap, A. G., Skavhaug, I. M., and Joshi, S. S., "Effects of Muscle Fatigue on the Usability of a Myoelectric Human-Computer Interface," *Human Movement Science*, Vol. 49, Oct. 2016, pp. 225–238.
https://doi.org/10.1016/j.humov.2016.06.009

[23] Hargrove, L., Englehart, K., and Hudgins, B., "A Training Strategy to Reduce Classification Degradation due to Electrode Displacements in Pattern Recognition Based Myoelectric Control," *Biomedical Signal Processing and Control*, Vol. 3, No. 2, 2008, pp. 175–180.
https://doi.org/10.1016/j.bspc.2007.11.005

[24] Kyranou, I., Vijayakumar, S., and Erden, M. S., "Causes of Performance Degradation in Electromyographic Pattern Recognition in Upper Limb Prostheses," *Frontiers in Neurorobotics*, Vol. 12, Sept. 2018, p. 58.
https://doi.org/10.3389/fnbot.2018.00058

[25] Roche, A. D., Rehbaum, H., Farina, D., and Aszmann, O. C., "Prosthetic Myoelectric Control Strategies: A Clinical Perspective," *Current Surgery Reports*, Vol. 2, No. 3, 2014, p. 44.
https://doi.org/10.1007/s40137-013-0044-8

[26] Iqbal, N. V., and Subramaniam, K., "A Review on Upper-Limb Myoelectric Prosthetic Control," *IETE Journal of Research*, Vol. 64, No. 6, 2018, pp. 740–752.
https://doi.org/10.1080/03772063.2017.1381047

[27] Artemiadis, P. K., and Kyriakopoulos, K. J., "An EMG-Based Robot Control Scheme Robust to Time-Varying EMG Signal Features," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 14, No. 3, 2010, pp. 582–588.
https://doi.org/10.1109/TITB.2010.2040832

[28] Hussain, I., Salvietti, G., Spagnoletti, G., and Prattichizzo, D., "The Soft-SixthFinger: A Wearable EMG Controlled Robotic Extra-Finger for Grasp Compensation in Chronic Stroke Patients," *IEEE Robotics and Automation Letters*, Vol. 1, No. 2, 2016, pp. 1000–1006.
https://doi.org/10.1109/Lra.2016.2530793

[29] Hussain, I., Spagnoletti, G., Salvietti, G., and Prattichizzo, D., "An EMG Interface for the Control of Motion and Compliance of a Supernumerary Robotic Finger," *Frontiers in Neurorobotics*, Vol. 10, Nov. 2016, p. 18.
https://doi.org/10.3389/fnbot.2016.00018

[30] Ison, M., and Artemiadis, P., "Proportional Myoelectric Control of Robots: Muscle Synergy Development Drives Performance Enhancement, Retainment, and Generalization," *IEEE Transactions on Robotics*, Vol. 31, No. 2, 2015, pp. 259–268.
https://doi.org/10.1109/TRO.2015.2395731

[31] Chen, J., Glover, M., Li, C., and Yang, C., "Development of a User Experience Enhanced Teleoperation Approach," *2016 International Conference on Advanced Robotics and Mechatronics (ICARM)*, Inst. of Electrical and Electronics Engineers, New York, 2016, pp. 171–177.
https://doi.org/10.1109/ICARM.2016.7606914

[32] Skavhaug, I. M., Lyons, K. R., Nemchuk, A., Muroff, S. D., and Joshi, S. S., "Learning to Modulate the Partial Powers of a Single sEMG Power Spectrum Through a Novel Human-Computer Interface," *Human Movement Science*, Vol. 47, June 2016, pp. 60–69.
https://doi.org/10.1016/j.humov.2015.12.003

[33] Safir, O., Dubrowski, A., Hui, Y., Backstein, D., and Carnahan, H., "Self-Directed Practice Scheduling is Equivalent to Instructor Guided Practice When Learning a Complex Surgical Skill," *Procedia-Social and Behavioral Sciences*, Vol. 2, No. 2, 2010, pp. 792–796.
https://doi.org/10.1016/j.sbspro.2010.03.105

[34] Sigrist, R., Rauter, G., Riener, R., and Wolf, P., "Augmented Visual, Auditory, Haptic, and Multimodal Feedback in Motor Learning: A Review," *Psychonomic Bulletin & Review*, Vol. 20, No. 1, 2013, pp. 21–53.
https://doi.org/10.3758/s13423-012-0333-8

[35] Bosse, H. M., Mohr, J., Buss, B., Krautter, M., Weyrich, P., Herzog, W., Jünger, J., and Nikendei, C., "The Benefit of Repetitive Skills Training and Frequency of Expert Feedback in the Early Acquisition of Procedural Skills," *BMC Medical Education*, Vol. 15, No. 1, 2015, p. 22.
https://doi.org/10.1186/s12909-015-0286-5

[36] Al-Saud, L. M., Mushtaq, F., Allsop, M. J., Culmer, P. C., Mirghani, I., Yates, E., Keeling, A., Mon–Williams, M., and Manogue, M., "Feedback and Motor Skill Acquisition Using a Haptic Dental Simulator," *European Journal of Dental Education*, Vol. 21, No. 4, 2017, pp. 240–247.
https://doi.org/10.1111/eje.12214

[37] Cecilio-Fernandes, D., Cnossen, F., Coster, J., Jaarsma, A. C., and Tio, R. A., "The Effects of Expert and Augmented Feedback on Learning a Complex Medical Skill," *Perceptual and Motor Skills*, Vol. 127, No. 4, 2020, pp. 766–784.
https://doi.org/10.1177/0031512520914680

[38] Huet, M., Jacobs, D. M., Camachon, C., Missenard, O., Gray, R., and Montagne, G., "The Education of Attention as Explanation of Variability of Practice Effects: Learning the Final Approach Phase in a Flight Simulator," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 37, No. 6, 2011, p. 1841.
https://doi.org/10.1037/a0024386

[39] Eisma, Y. B., Borst, C., Paassen, R. V., and Winter, J. D., "Augmented Visual Feedback: Cure or Distraction?" *Human Factors*, 2020,

Paper 0018720820924602.
https://doi.org/10.1177/0018720820924602

[40] Triantafyllidis, E., Mcgreavy, C., Gu, J., and Li, Z., "Study of Multimodal Interfaces and the Improvements on Teleoperation," *IEEE Access*, Vol. 8, April 2020, pp. 78,213–78,227.
https://doi.org/10.1109/ACCESS.2020.2990080

[41] Marschall, F., Bund, A., and Wiemeyer, J., "Does Frequent Augmented Feedback Really Degrade Learning? A Meta-Analysis," *Bewegung und Training*, Vol. 1, 2007, pp. 75–86.

[42] Liebermann, D. G., Katz, L., Hughes, M. D., Bartlett, R. M., McClements, J., and Franks, I. M., "Advances in the Application of Information Technology to Sport Performance," *Journal of Sports Sciences*, Vol. 20, No. 10, 2002, pp. 755–769.
https://doi.org/10.1080/026404102320675611

[43] Billings, D., "Efficacy of Adaptive Feedback Strategies in Simulation-Based Training," *Military Psychology*, Vol. 24, No. 2, 2012, pp. 114–133.
https://doi.org/10.1080/08995605.2012.672905

[44] Hattie, J., and Timperley, H., "The Power of Feedback," *Review of Educational Research*, Vol. 77, No. 1, 2007, pp. 81–112.
https://doi.org/10.3102/003465430298487

[45] Wijewickrema, S., Ma, X., Bailey, J., Kennedy, G., and OLeary, S., "Feedback Techniques in Computer-Based Simulation Training: A Survey," arXiv preprint arXiv:1705.04683, 2017.

[46] Dosen, S., Markovic, M., Strbac, M., Belić, M., Kojić, V., Bijelić, G., Keller, T., and Farina, D., "Multichannel Electrotactile Feedback with Spatial and Mixed Coding for Closed-Loop Control of Grasping Force in Hand Prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 25, No. 3, 2016, pp. 183–195.
https://doi.org/10.1109/TNSRE.2016.2550864

[47] Markovic, M., Karnal, H., Graimann, B., Farina, D., and Dosen, S., "GLIMPSE: Google Glass Interface for Sensory Feedback in Myoelectric Hand Prostheses," *Journal of Neural Engineering*, Vol. 14, No. 3, 2017, Paper 036007.
https://doi.org/10.1088/1741-2552/aa620a

[48] Schweisfurth, M. A., Markovic, M., Dosen, S., Teich, F., Graimann, B., and Farina, D., "Electrotactile EMG Feedback Improves the Control of Prosthesis Grasping Force," *Journal of Neural Engineering*, Vol. 13, No. 5, 2016, Paper 056010.
https://doi.org/10.1088/1741-2560/13/5/056010

[49] Shehata, A. W., Engels, L. F., Controzzi, M., Cipriani, C., Scheme, E. J., and Sensinger, J. W., "Improving Internal Model Strength and Performance of Prosthetic Hands Using Augmented Feedback," *Journal of Neuroengineering and Rehabilitation*, Vol. 15, No. 1, 2018, p. 70.
https://doi.org/10.1186/s12984-018-0417-4

[50] Huang, H., Wolf, S. L., and He, J., "Recent Developments in Biofeedback for Neuromotor Rehabilitation," *Journal of Neuroengineering and Rehabilitation*, Vol. 3, No. 1, 2006, p. 11.
https://doi.org/10.1186/1743-0003-3-11

[51] Basmajian, J. V., "Control and Training of Individual Motor Units," *Science*, Vol. 141, No. 3579, 1963, pp. 440–441.
https://doi.org/10.1126/science.141.3579.440

[52] Stanton, R., Ada, L., Dean, C. M., and Preston, E., "Biofeedback Improves Performance in Lower Limb Activities More Than Usual Therapy in People Following Stroke: A Systematic Review," *Journal of Physiotherapy*, Vol. 63, No. 1, 2017, pp. 11–16.
https://doi.org/10.1016/j.jphys.2016.11.006

[53] Pritchett, A. R., "Aviation Automation: General Perspectives and Specific Guidance for the Design of Modes and Alerts," *Reviews of Human Factors and Ergonomics*, Vol. 5, No. 1, 2009, pp. 82–113.
https://doi.org/10.1518/155723409X448026

[54] Seow, R. Y. T., Betts, S., and Anderson, J. R., "Transfer Effects of Varied Practice and Adaptation to Changes in Complex Skill Acquisition," *Proceedings of the 17th International Conference on Cognitive Modelling*, Univ. of Waterloo, Waterloo, Canada, 2019, pp. 222–227, https://iccm-conference.neocities.org/2019/proceedings/ICCM2019Proceedings.pdf.

[55] Braun, D. A., Aertsen, A., Wolpert, D. M., and Mehring, C., "Motor Task Variation Induces Structural Learning," *Current Biology*, Vol. 19, No. 4, 2009, pp. 352–357.
https://doi.org/10.1016/j.cub.2009.01.036

[56] Johnson, R. E., Kording, K. P., Hargrove, L. J., and Sensinger, J. W., "Does EMG Control Lead to Distinct Motor Adaptation?" *Frontiers in Neuroscience*, Vol. 8, Sept. 2014, p. 302.
https://doi.org/10.3389/fnins.2014.00302

[57] Lyons, R. K., and Joshi, S. S., "Effects of Mapping Uncertainty on Visuomotor Adaptation to Trial-by-Trial Perturbations with Proportional Myoelectric Control," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Inst. of Electrical and Electronics Engineers, New York, 2018, pp. 5178–5181.
https://doi.org/10.1109/EMBC.2018.8513412

[58] Shehata, A. W., Scheme, E. J., and Sensinger, J. W., "The Effect of Myoelectric Prosthesis Control Strategies and Feedback Level on Adaptation Rate for a Target Acquisition Task," *2017 International Conference on Rehabilitation Robotics (ICORR)*, Inst. of Electrical and Electronics Engineers, New York, 2017, pp. 200–204.
https://doi.org/10.1109/ICORR.2017.8009246

[59] Sternad, D., "It's Not (Only) the Mean that Matters: Variability, Noise and Exploration in Skill Learning," *Current Opinion in Behavioral Sciences*, Vol. 20, April 2018, pp. 183–195.
https://doi.org/10.1016/j.cobeha.2018.01.004

[60] Karasinski, J., Robinson, S., Handley, P., and Duda, K., "Real-Time Performance Feedback in a Manually-Controlled Spacecraft Inspection Task," *AIAA Modeling and Simulation Technologies Conference*, AIAA Paper 2017-1314, 2017.
https://doi.org/10.2514/6.2017-1314

[61] Hart, S. G., "NASA-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50, Sage Publ., Los Angeles, 2006, pp. 904–908.
https://doi.org/10.1177/154193120605000909

[62] Hart, S. G., and Staveland, L. E., "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, Vol. 52, Jan. 1988, pp. 139–183.
https://doi.org/10.1016/S0166-4115(08)62386-9

[63] Roscoe, A. H., and Ellis, G. A., "A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use," Royal Aerospace Establishment TR ADA227864, Farnborough, England, U.K., 1990.

[64] Jorna, P., "Heart Rate and Workload Variations in Actual and Simulated Flight," *Ergonomics*, Vol. 36, No. 9, 1993, pp. 1043–1054.
https://doi.org/10.1080/00140139308967976

[65] Casner, S. M., and Gore, B. F., "Measuring and Evaluating Workload: A Primer," NASA TM 216395, 2010.

[66] Gawron, V. J., *Human Performance, Workload, and Situational Awareness Measures Handbook*, Vol. 2, CRC Press, Boca Raton, FL, 2019, p. 16.

[67] Williges, R. C., and Wierwille, W. W., "Behavioral Measures of Aircrew Mental Workload," *Human Factors*, Vol. 21, No. 5, 1979, pp. 549–574.
https://doi.org/10.1177/001872087902100503

[68] Lee, J. D., and See, K. A., "Trust in Automation: Designing for Appropriate Reliance," *Human Factors*, Vol. 46, No. 1, 2004, pp. 50–80.
https://doi.org/10.1518/hfes.46.1.50_30392

[69] Wickens, C. D., Gempler, K., and Morphew, M. E., "Workload and Reliability of Predictor Displays in Aircraft Traffic Avoidance," *Transportation Human Factors*, Vol. 2, No. 2, 2000, pp. 99–126.
https://doi.org/10.1207/STHF0202_01

[70] Hoff, K. A., and Bashir, M., "Trust in Automation: Integrating Empirical Evidence on Factors that Influence Trust," *Human Factors*, Vol. 57, No. 3, 2015, pp. 407–434.
https://doi.org/10.1177/0018720814547570

[71] Fitts, P. M., "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement," *Journal of Experimental Psychology*, Vol. 47, No. 6, 1954, p. 381.
https://doi.org/10.1037/h0055392

[72] Salmoni, A. W., Schmidt, R. A., and Walter, C. B., "Knowledge of Results and Motor Learning: A Review and Critical Reappraisal," *Psychological Bulletin*, Vol. 95, No. 3, 1984, p. 355.
https://doi.org/10.1037/0033-2909.95.3.355

[73] Choi, C., Na, Y., Rim, B., Kim, Y., Kang, S., and Kim, J., "An SEMG Computer Interface Using Three Myoelectric Sites for Proportional Two-Dimensional Cursor Motion Control and Clicking for Individuals with Spinal Cord Injuries," *Medical Engineering & Physics*, Vol. 35, No. 6, 2013, pp. 777–783.
https://doi.org/10.1016/j.medengphy.2012.08.006

[74] O'Meara, S. M., Shyr, M. C., Lyons, K. R., and Joshi, S. S., "Comparing Two Different Cursor Control Methods Which use Single-Site Surface Electromyography," *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, Inst. of Electrical and Electronics Engineers, New York, 2019, pp. 1163–1166.
https://doi.org/10.1109/NER.2019.8716903

[75] Lyons, K. R., and Joshi, S. S., "Real-Time Evaluation of a Myoelectric Control Method for High-Level Upper Limb Amputees Based on Homologous Leg Movements," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Inst. of Electrical and Electronics Engineers, New York, 2016, pp. 6365–6368.

[76] Lyons, K., and Margolis, B., "AxoPy: A Python Library for Implementing Human-Computer Interface Experiments," *Journal of Open Source Software*, Vol. 4, No. 34, 2019, p. 1191.
https://doi.org/10.21105/joss.01191

[77] Soukoreff, R. W., and MacKenzie, I. S., "Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI," *International Journal of Human-Computer Studies*, Vol. 61, No. 6, 2004, pp. 751–789.
https://doi.org/10.1016/j.ijhcs.2004.09.001

[78] O'Meara, S. M., Karasinski, J., Miller, C. L., Joshi, S., and Robinson, S. K., "The Effects of Training Methodology on Performance, Workload, and Trust During Human Learning of a Computer-Based Task," *AIAA SciTech 2020 Forum*, AIAA Paper 2020-1110, 2020.
https://doi.org/10.2514/6.2020-1110

[79] Karasinski, J. A., Robinson, S. K., Duda, K. R., and Prasov, Z., "Development of Real-time Performance Metrics for Manually-Guided Spacecraft Operations," *2016 IEEE Aerospace Conference*, Inst. of Electrical and Electronics Engineers, New York, 2016, pp. 1–9.
https://doi.org/10.1109/AERO.2016.7500734

[80] Wickens, C., and Hollands, J., *Engineering Psychology and Human Performance*, Prentice Hall, Upper Saddle River, NJ, 2000, p. 75.

[81] Lyons, K. D., Slaughenhaupt, R. M., Mupparaju, S. H., Lim, J. S., Anderson, A. A., Stankovic, A. S., Cowan, D. R., Fellows, A. M., Binsted, K. A., and Buckey, J. C., "Autonomous Psychological Support for Isolation and Confinement," *Aerospace Medicine and Human Performance*, Vol. 91, No. 11, 2020, pp. 876-885.
https://doi.org/10.3357/AMHP.5705.2020

[82] Lyons, K. R., and Joshi, S. S., "Upper Limb Prosthesis Control for High-Level Amputees via Myoelectric Recognition of Leg Gestures," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 26, No. 5, 2018, pp. 1056-1066.
https://doi.org/10.1109/TNSRE.2018.2807360

[83] Chhabra, M., and Jacobs, R. A., "Near-Optimal Human Adaptive Control Across Different Noise Environments," *Journal of Neuroscience*, Vol. 26, No. 42, 2006, pp. 10,883–10,887.
https://doi.org/10.1523/JNEUROSCI.2238-06.2006

[84] Jian, J.-Y., Bisantz, A. M., and Drury, C. G., "Foundations for an Empirically Determined Scale of Trust in Automated Systems," *International Journal of Cognitive Ergonomics*, Vol. 4, No. 1, 2000, pp. 53–71.
https://doi.org/10.1207/S15327566IJCE0401_04

[85] Bates, D., Mächler, M., Bolker, B., and Walker, S., "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, Vol. 67, No. 1, 2015, p. 48.
https://doi.org/10.18637/jss.v067.i01

[86] Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P., "The Role of Trust in Automation Reliance," *International Journal of Human-Computer Studies*, Vol. 58, No. 6, 2003, pp. 697–718.
https://doi.org/10.1016/S1071-5819(03)00038-7

[87] Cler, M. J., and Stepp, C. E., "Discrete Versus Continuous Mapping of Facial Electromyography for Human-Machine Interface Control: Performance and Training Effects," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 23, No. 4, 2015, pp. 572–580.
https://doi.org/10.1109/TNSRE.2015.2391054

[88] Williams, M. R., and Kirsch, R. F., "Evaluation of Head Orientation and Neck Muscle EMG Signals as Command Inputs to a Human-Computer Interface for Individuals with High Tetraplegia," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 16, No. 5, 2008, pp. 485-496.
https://doi.org/10.1109/Tnsre.2008.2006216

[89] Williams, M. R., and Kirsch, R. F., "Case Study: Head Orientation and Neck Electromyography for Cursor Control in Persons with High Cervical Tetraplegia," *Journal of Rehabilitation Research & Development*, Vol. 53, No. 4, 2016, pp. 519–530.
https://doi.org/10.1682/JRRD.2014.10.0244

E. Atkins
Associate Editor