Toward Intuitive Teleoperation in Surgery: Human-Centric Evaluation of Teleoperation Algorithms for Robotic Needle Steering

Conference Paper · May 2018

DOI: 10.1109/ICRA.2018.8460729

3 authors, including:

Ziheng Wang
University of Texas at Dallas
5 PUBLICATIONS **1** CITATION

SEE PROFILE

2018 IEEE International Conference on Robotics and Automation (ICRA)
May 21-25, 2018, Brisbane, Australia

Toward Intuitive Teleoperation in Surgery: Human-centric Evaluation of Teleoperation Algorithms for Robotic Needle Steering

Ziheng Wang[1], *Student Member, IEEE* , Isabella Reed[2], Ann Majewicz Fey[1], *Member, IEEE*

***Abstract*— The effectiveness of control algorithms for teleoperated systems is typically evaluated through experimental performance measures, post-experimental user surveys, and theoretical analysis. However, none of these methods provide an objective assessment of teleoperation algorithms with respect to the real-time changes of human users during teleoperated tasks in terms of physiological, kinematic, or cognitive metrics. In this study, we recruited subjects to control robotically steered needles in a randomized experiment, using four different teleoperation mappings (joint space control, steering control, and Cartesian space control with and without force feedback). We investigated how the choice of these algorithms affect both performance and user response. Our novel steering control mapping, which mimics hub-centered steering, is significantly correlated with decreased cognitive stress and improved teleoperation performance when compared to joint space control. Overall, user experience and teleoperation performance were significantly improved with Cartesian space control, resulting in faster needle insertion, higher targeting accuracy, lower cognitive load, and smoother movements. Furthermore, while additional haptic feedback in Cartesian space provided an improved performance, it may increase user cognitive workload and muscle fatigue. These results highlight the importance of considering human-centric metrics when designing novel teleoperation strategies for complex systems.**

I. INTRODUCTION

Many algorithms have been developed to enable human-in-the-loop control for robotic systems. In scenarios where the master manipulator and slave robot are kinematically similar, position-based teleoperation is often used. More complex systems, such as single-master-multiple-slave systems, multiple-master-single-slave systems, or kinematically dissimilar systems, require more complex mappings between user inputs and system outputs [1]–[3]. Robotic needle steering is an example of a kinematically dissimilar teleoperated system. Steerable needles behave with nonholonomic kinematics [4], yet the human operator does not have these constraints. As a result, manual control of steerable needles can be challenging and non-intuitive. Several teleoperation algorithms have developed for ease of control, such as joint space control [5] and Cartesian space control [6]. However, there is a lack of clear guidelines on how to choose one algorithm over the others.

Current methods to assess the effectiveness of teleoperation systems are largely limited to (1) measure of performance metrics, (2) post-experiment user surveys, and (3) control theoretical analyses. For example, studies involving endoscopic catheterization and needle steering typically evaluate outcome-based features such as targeting accuracy, position and orientation errors, and path length [5]–[12]. Furthermore, several common metrics for general task-oriented human-robot interaction are widely used to assess performance across tasks and systems [13], including task completion time and mission completion percentage. Self-reporting questionnaires, such as the NASA Task Load Index (NASA-TLX) [14], can also assess task workload in the human-machine teleoperation [15]; however, surveys are subject-dependent, inconsistent, and often biased [16]. Finally, theoretical analysis of system stability and transparency have been well-studied, both analytically and experimentally, to evaluate the effectiveness of bilateral teleoperation systems [17]–[19]. While these methods have been successful in evaluating teleoperation systems from a control perspective, they do not take into consideration the physical and cognitive response of human users. These may be important for evaluation purposes as several cases have been reported of a mismatch between user acceptance of some feedback method, and whether or not performance was improved [6], [11], [20]–[22].

Human-centric metrics, derived from physiological sensors and user motion, could prove useful in evaluating teleoperation algorithms, from a human perspective. The analysis of physiological signals from human central and peripheral nervous systems has enabled objective measurements of user response in a variety of studies. For example, electroencephalogram (EEG) can provide an objective measure of human perception, cognition, and technical skills [23], [24]; Electromyography (EMG) signals, generated by active muscles, are useful in revealing underlying motion patterns, physical effort, and stress [25], [26]. Heart rate response (HR) and galvanic skin response (GSR), fluctuations of skin electrical conductance, have been shown to be reliable indicators of human dynamic mental workload, emotion, and cognitive stress [27]–[29]. Additionally, user kinematic profiles, derived from position and acceleration sensors, prove to be useful in revealing human capabilities in perform complex motor tasks [30]–[32].

While the studies above have shown the usefulness of physiological and kinematic metrics on evaluating some type of user response, few, if any, have considered the relationship between multiple human-centric metrics and more traditional task-based metrics. In this study, we propose a framework for objectively evaluating the effectiveness of various teleoperation algorithms by exploring meaningful metrics derived

This work is supported by National Science Foundation (NSF#1464432)

[1]Ziheng Wang, and Ann Majewicz Fey, are with Department of Mechanical Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA. zxw151130@utdallas.edu

[2] Isabella Reed is with the Department of Bioengineering, The University of Texas at Dallas, Richardson, TX 75080, USA.

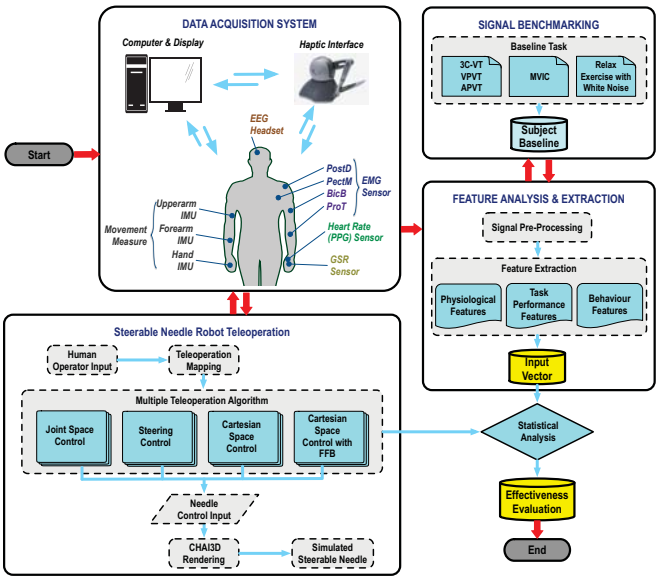


Fig. 1: Our teleoperation evaluation framework scheme consists of three major parts: data acquisition and signal benchmarking, feature extraction, and statistical analysis and evaluation, applied to various simulated teleoperation algorithms for robotic needle steering.

from both human-response and task performance (Fig. 1). We apply this method to evaluate four possible algorithms for teleoperation of steerable needles, using a haptic device. Four different teleoperation algorithms include (1) joint space control (JC), (2) a novel steering control (SC), (3) Cartesian space control (CC), and (4) Cartesian space control with position-error-based force feedback (CFB).

II. NEEDLE STEERING TELEOPERATION

A. Simulated Needle Kinematics

Using the homogeneous matrix representation, the rigid transformation of needle in body frame n, attached to the needle tip, with respect to a fixed starting base frame s, is described as $g_{sn}(t) = \begin{bmatrix} R_{sn}(t) & p_{sn}(t) \\ 0 & 1 \end{bmatrix} \in SE(3)$, where $p_{sn}(t) \in \mathbb{R}^3$ denotes needle tip position, $R_{sn}(t) \in SO(3)$ denotes the rotation between the frame n and s. Forward kinematics of needle are defined as $g_{sn}(t) = g_{sn}(0) g_{rot}(t) \exp(t\widehat{V}^n_{sn})$, where $g_{rot}(t) = \begin{bmatrix} R_{\vec{e}_2}(\gamma(t)) & \mathbf{0} \\ 0 & 1 \end{bmatrix}$ is the relative rotation about the needle axis $\vec{e}_2$ by angle γ (constrained to 0° or 180°), a 3×3 matrix $\widehat{V}^n_{sn}$ is the instantaneous needle tip velocity, and $g_{sn}(0)$ is the previous needle configuration. Detailed definitions of needle kinematics can be found in the prior work [6].

For simulation, the position and orientation of needle are updated, at each time step t, given user control inputs of needle orientation and velocity for each mapping (details are provided in the following section).

B. Teleoperation Mapping

The four teleoperation mapping strategies designed for the needle steering task are shown in Fig. 2.

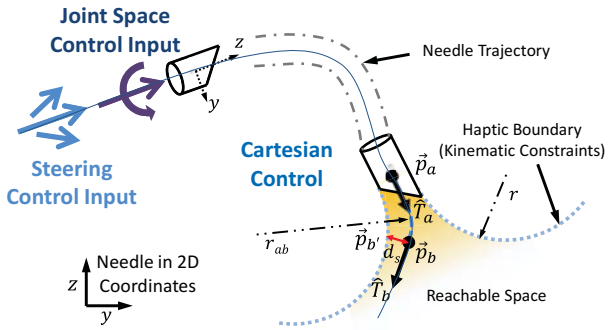


Fig. 2: Teleoperation algorithms for 2D needle steering. In both joint space and steering control, needle is controlled via insertion and rotation cues. In Cartesian control, steering is achieved by updating tip positions in reachable space.

1) Joint Space Mapping: Joint space teleoperation (JC) was implemented in the insertion and rotation of steerable needle [6]. For needle insertion, the insertion velocity was defined as $v(t) = k_1 p_i$ where k_1 is the scalar gain $k_1 = 0.00075$, p_i is insertion distance along z axis of the device stylus; for rotation, changes in gimbal angle γ around the z axis of the device stylus are mapped to changes in rotation of the needle tip. In 2D steering, the needle tip orientation was constrained to 0° and 180° (i.e. curve left and curve right).

2) Steering Mapping: A novel steering control mapping (SC) was developed, similar to hub-centered steering (used in motorcycles and bicycles). Our hypothesis is that steering control might be a more effective and intuitive form of joint-based needle control. In this mapping, needle insertion is defined as for joint space control; however, the control of rotation degree of freedom (DOF) is different: the SC teleoperation was implemented by direction-pointing (left or right-pointing) using haptic device stylus to define a steering angle (i.e., needle rotation about x axis). A rate control law, $v(t) = k_2 d_i$, was applied to calculate insertion velocity, where the scalar gain is $k_2 = 0.00075$, and d_i is the insertion distance along the pointing direction.

3) Cartesian Space Mapping: Needle steering using Cartesian-space Control (CC) was achieved by updating a constant-curvature path segment in a reachable space at each time step. The path segment with a curvature, r_{ab}, was obtained for the needle from the current needle tip position, $\vec{p}_a$, and the user-controlled stylus position (user position), $\vec{p}_b$ [6]. Additionally, to simulate realistic needle steering, the kinematic constraints of the needle are respected. This is done by rendering a haptic virtual boundary, determined by the minimum achievable radius of curvature of the needle, r, and the current needle heading, $\widehat{T}_a$. For user positions outside the reachable space, a virtual repulsive haptic force with a negative stiffness gain $k_s = -100$ is applied to constrain the user to only reachable positions. The magnitude of the force is proportional to the distance, d_s, where $d_s = ||\vec{p}_b - \vec{p}_{b'}||$, the distance between the current position $\vec{p}_b$ and the orthogonal-projected position on the closest side of the kinematic boundary, $\vec{p}_{b'}$.

Furthermore, a secondary form of Cartesian teleoperation

with an additional haptic guidance feedback (CFB), was also evaluated. In addition to kinematic constraints of needle, a position-error-based force feedback $\vec{F}_{FB}$ was presented to the subjects. This virtual force is proportional, with a gain $k_{FB} = 20$, to the distance between the current needle tip and user position, $||\vec{p}_a - \vec{p}_b||$. Apart from the additional force feedback $\vec{F}_{FB}$, Cartesian space control with force feedback (CFB) was the same to the design of Cartesian space control (CC).

III. METHODOLOGY

A systematic framework for evaluating teleoperation algorithms for steerable needles was developed, as shown in Fig. 1. The various algorithms are evaluated through human-centric data acquisition and signal benchmarking, feature extraction, and statistical analysis and evaluation.

A. Simulation of Needle Steering

A simulated needle steering teleoperation environment was developed using C++ code and the CHAI3D library. The objective of the teleoperation task was to steer the simulated needle to reach one of four predefined targets (5 mm in diameter and mirrored vertically), as defined by one of the four teleoperation algorithms presented, by moving a haptic stylus (Geomagic Touch, 3D Systems, SC, USA). Figure 3 shows a human subject interacting with the simulated haptic and graphical environment. The necessary input commands to the haptic device for each algorithm are shown in Fig. 4. The target layout is shown in Fig. 5(a).

The user experiment consisted of unrecorded training trials and recorded movement trials. During training, subjects manipulated the stylus to become acquainted with device control and experiment procedures. After training, subjects were presented with a random teleoperation algorithm and conducted insertions to random targets, five times for each target. It results in a total of 160 needle insertions for every participant. To reduce unwanted effects of user boredom or fatigue due to repetitions, subjects took 45-second breaks between each new teleoperation algorithm block.

Six healthy adults in total (5 males, 1 female, all right-hand dominant) participated in this study. All subjects provided informed consent in accordance with University of Texas at Dallas Institutional Review Board (UTD IRB #14-57).

B. Data Acquisition and Signal Benchmarking

Real-time sensor signals (human physiological response, user kinematics, and performance data) were acquired simultaneously, using custom Robotic Operating System (ROS) and C++ codes, at sampling rates defined by the hardware.

To measure muscle activity, surface EMG signals were acquired (1024 Hz sampling rate) using non-invasive Al/AgCl electrodes via Bluetooth modules (Shimmer Sensing, Inc., Dublin, Ireland). Four muscles, primarily active during the required inputs for the haptic stylus, were selected for EMG monitoring, including: Pectoralis major (PectM), Deltoid Posterior (PostD), Biceps Brachii (BicB) and Pronator Teres (PronT). Heart response was measured using photoplethysmography (PPG) via an optical pulse sensor on the subject's

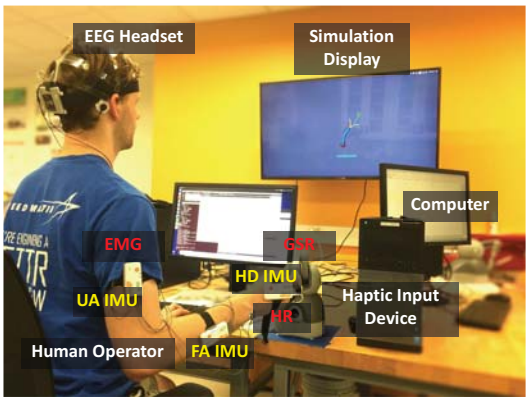
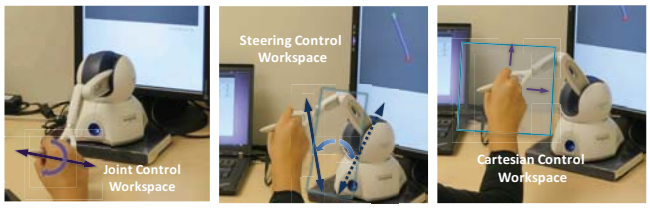


Fig. 3: A participant interacted with the teleoperation simulation by manipulating the end-effector of a haptic device, while real-time physiological response and movement kinematics were recorded by data acquisition system.



(a) JC Setup (b) SC Setup (c) CC Setup

Fig. 4: User input mappings for each teleoperation method (a) joint space control; (b) steering control; and (c) Cartesian space control.

fingertip (512 Hz sampling rate). GSR signals were acquired (512 Hz sampling rate) via two electrodes placed on two neighboring fingers. EEG signals were acquired (1024 Hz sampling rate) using a B-Alert X10 wireless headset (Advanced Brain Monitoring, CA, USA). Nine Al/AgCl electrodes measuring electric potentials were placed on the scalp, according to the sensor specifications [33]. User kinematic data was acquired using inertial measurement units (IMU, Shimmer Sensing Inc., Ireland) placed on the hand (HD), forearm (FA), and upper arm (UA) (512 Hz sampling rate). The kinematics of the haptic master controller were obtained at the sample rate of 512 Hz.

To acquire subject-independent features of physiological response and reduce unwanted noise, signals were normalized through pre-experimental benchmarking for each subject. The benchmarking of EEG cognition states consisted of three interactive sessions, included with the BIOPAC Cognitive State Analysis Software: a 3-choice vigilance task (3C-VT), a visual psychomotor vigilance task (VPVT), and an auditory psychomotor vigilance task (APVT) [34]. To normalize EMG signals across subjects, the maximum voluntary isometric contraction (MVIC) was performed for each muscle. Each subject was asked to contract each muscle with a maximum force and hold the contraction for 20 seconds, three times in a row, with two-minute rest intervals. The maximum muscle response in the MVIC trials was recorded and used as an EMG baseline for each subject. To normalize GSR and PPG heart response, a baseline trial was

conducted wherein subjects were asked to close eyes while listening to white noise for three minutes.

Finally, NASA-TLX questionnaires were acquired to obtain subjective ratings on the user experience and task load after each teleoperation algorithm block. The survey includes six multidimensional evaluation questions (mental demand, physical demand, temporal demand, perceived performance, effort and frustration). Subjects were asked to rate scores on each question, ranging from 0 to 20.

C. Feature Extraction

A variety of processing techniques were used to obtain features from the time series data. The features are categorized into three groups: physiological features, motion kinematics, and task-specific performance metics. All features used are task-level metrics, computed for each needle steering trial.

1) Physiological Features: To process EMG data, a Fourier transform of the autocorrelation function was employed on the EMG time series data to obtain the power spectral density (PSD), which is commonly used for EMG analysis [35]. One frequency-domain feature, mean power (*MNP*), defined as Eq. 1, was extracted from the PSD in the range of 20 - 450 Hz, where P_i is the power value at the i-th frequency bin and N is the length of frequency bins.

$$MNP = \frac{1}{N}\sum_{i=1}^{N} P_i \tag{1}$$

For analysis of EEG time series, two probabilistic metrics, *Engagement* and *Workload*, are obtained from the BIOPAC Cognitive State Analysis Software. There metrics are classified from real-time raw EEG signals and computed as probability measures, normalized to each subject through a pre-experimental benchmarking session [24], [36]. Details regarding the EEG processing technique and the validation of computed metrics can be found in the literature [34].

To process GSR signals, a second-order lowpass FIR filter (with cutoff frequency 5 Hz) was applied to remove signal artifacts. Average skin conductance levels, SC_{avg}, and global variance of skin conductance, SC_{vr}, (changes between the global maximum and minimum [37]), were extracted to assess the individual galvanic response in teleoperation, and normalized to a baseline GSR signal for each subject.

Finally, to assess user cognitive stress and workload, average heart rate, *HR*, defined as $HR = 60/PP$, was derived from finger-tip PPG signals, where the *PP* is the average normal-to-normal interval of PPG pulse waveforms [28], [38]. The calculated *HR* was normalized for each subject by the heart rate value measured in the baseline task.

2) User Kinematics: Kinematic features were computed from human arm motions captured with the inertial motion units (IMU). Average angular velocity (*AngVel*) and average linear acceleration (*LinAcc*) were captured as measures of movement magnitudes for the duration of each trial. Non-dimensional jerk (*Jerk*), a feature for assessing the movement smoothness [39], was calculated from kinematic data in 3D space. The calculation of non-dimensional jerk is defined as Eq. 2, where x, y, z are the arm position components, L is length of arm trajectory, and T is manipulation time in total.

$$J = \left(\frac{1}{2}\int_0^T (x'''(t)^2 + y'''(t)^2 + z'''(t)^2)dt\right)\frac{T^5}{L^2} \tag{2}$$

3) Steering Task Performance: Task-specific performance metrics reflect how well the tasks were executed, and are widely adopted in current teleoperation assessment studies. Four performance metrics were chosen to evaluate needle steering teleoperation. Task movement time (*MT*) is defined as the total time for simulated needle to reach targets. The path straight deviation (*PathStrDev*), shown as Eq. 3, is defined as the average magnitude of the orthogonal projection of the current needle tip position P_i, onto the vector $\vec{a}$, where $\vec{a}$ is the vector between the starting point P_0 and final needle position P_n. The *PathStrDev* measures the straightness deviation of needle trajectory, where the value of zero indicates a purely straight trajectory.

$$PathStrDev = \frac{1}{n}\sum_{i=1}^{n} d(P_i,(\vec{a})) = \frac{1}{n}\sum_{i=1}^{n} \frac{\left\| \vec{P_0P_i} \times \vec{a} \right\|}{\|\vec{a}\|} \tag{3}$$

The standard deviation of needle speed ($NeedleVel_{SD}$) is a measure of users ability to continuously steer the needle [6]. Finally, the position targeting error (*TargetErr*) is the normalized distance between the target and needle tip at the end of each trial, and is a performance measure of targeting accuracy.

D. Statistical Analysis

A two-way analysis of variance (ANOVA) was conducted to identify the effects of teleoperation algorithms, as well as target location, on the human-centric and task performance metrics. A post-hoc Tukey's honest significant difference (HSD) test was performed to identify significantly different groups among different teleoperation algorithms and target sites. The level of significance was determined by $p-$values less than 0.05 for all statistics.

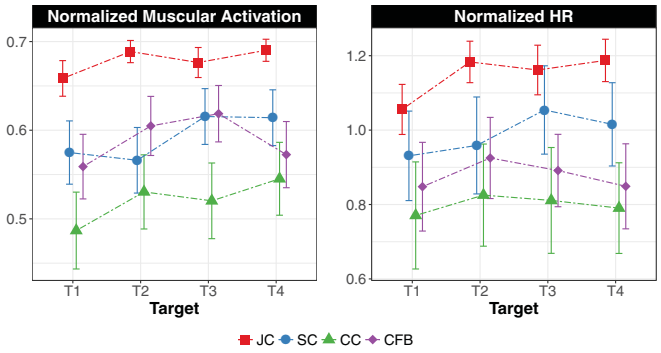
IV. RESULTS AND DISCUSSION

A. Human-centric Assessment

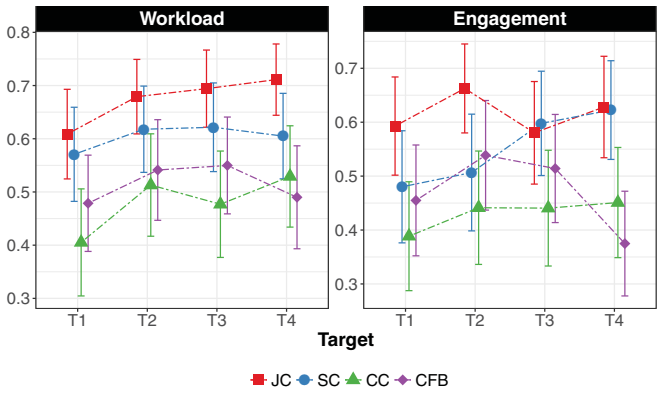
Target geometric configuration and sequential trajectories of needle tips for each teleoperation mapping algorithms (selected from three random subjects for illustration) are shown in Fig. 5, colored by different targets. Fig. 6, Fig. 7 and Fig. 8 show physiological response, movement kinematics, and task-specific performance metrics across all subjects, respectively. Fig. 7 shows the motion kinematics averaged from user hand, forearm, and upper arm, as these metrics all shared similar trends. Error bars illustrate data variances in a 95% confidence interval. We summarize results of ANOVA and post-hoc analysis for all the computed metrics in Tab. I.

(a) Target configuration (b) JC (c) SC (d) CC (e) CFB

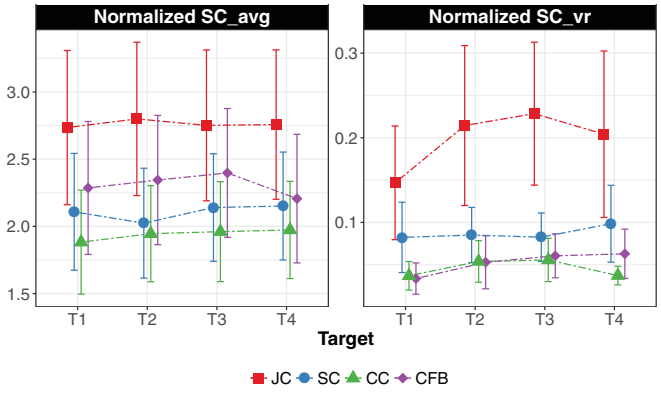
Fig. 5: Configuration of targets for needle steering experiments (a). Representative needle trajectories from three randomly-chosen subjects using (b) JC, (c) SC, (d) CC and (e) CFB. Needle paths are colored by the different targets.



(a) Muscular Activation & Heart Rate Response



(b) EEG Cognitive Metrics



(c) Skin Conductance

Fig. 6: Physiological response across all subjects for teleoperation (JC, SC, CC and CFB) on different target sites.

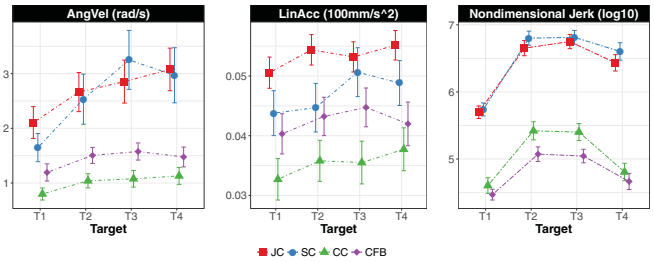


Fig. 7: Kinematic metrics across all subjects for teleoperation (JC, SC, CC and CFB) on different target sites.

1) Teleoperation Mappings: Joint space teleoperation is significantly ($p < 0.001$) associated with the higher levels for all measured physiological response than all other teleoperation algorithms, with the exception of cognitive workload and engagement, for which both joint space and steering control were significantly higher than both Cartesian control algorithms.

For user kinematic metrics, compared to joint space and steering control, both Cartesian control CC and CFB have significantly smaller motion amplitudes, angular velocity and linear acceleration, on forearm and hand ($p < 0.001$). Moreover, significantly smaller values of non-dimensional jerk indicate that Cartesian control can achieve smoother movements during steering tasks ($p < 0.001$). This could be explained by the fact that human operators have the direct control of needle tip in Cartesian space teleoperation; in joint space and steering control, users are required to not only have a strong cognitive model of nonholonomic needle kinematics, but also actively engage in trajectory planning, observation, and control, which could lead to the increased complexity of the task and more erratic movements.

2) Additional Haptic Feedback: Within Cartesian control, position-error-based force feedback (CFB) is associated with decreased non-dimensional jerk, higher needle velocity variances, and significantly lower movement time. These results suggest that the additional haptic feedback, $\vec{F}_{FB}$, could assist operators to better perform steering tasks.

However, the presence of position-error-based force feedback could potentially result in muscle fatigue, increased cognitive workload and stress. As shown in Fig. 6a, CFB is correlated with increase levels of muscular activation ($p < 0.001$), when compared to CC. This result could be explained

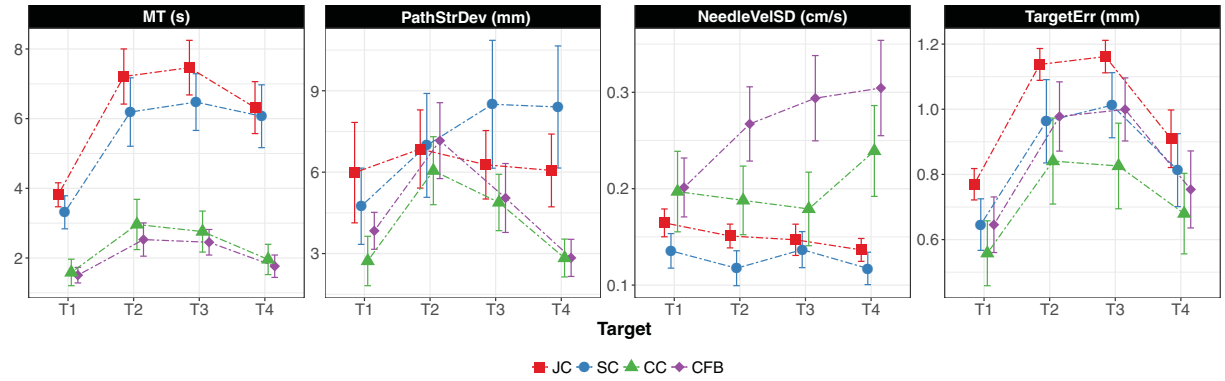


Fig. 8: Task-specific performance across all subjects for teleoperation (JC, SC, CC and CFB) on different target sites.

as the user is moving against a resistive force. Though not significant, CFB is associated with higher EEG-based workload and engagement, heart rate, as well as average skin conductance levels and variances. The results indicate the error-based force feedback does not always guarantee a natural and intuitive user experience in needle steering.

3) Environmental Geometry: Another attention of our analysis is focused on the environmental geometry of targets during needle steering. First, varied distances and relative layouts between the initial position and insertion targets can potentially affect the control easiness in needle steering. As shown in Fig. 5a, target T1 is the closest target but requires the highest curvatures to reach and T4 is the farthest target, but can be reached with a straighter path. Both targets T2 and T3 require some changes in path curvatures and longer trajectories than T1. Though not significant, as distances to targets increases (i.e, T1, T2, T3), cognitive load, muscle activation, and magnitudes of limb motions also increase. One explanation is that in order to reach far-away targets, participants would have to increasingly engage in path planning, and allocate greater efforts mentally and physically while maintaining performance accuracy. However, the computed metrics for T4 decrease, which could reflect the fact that relatively straight-line paths require less modification in control efforts to attain.

Second, the effects of the target distance and relative layout on the measures of user response and performance are different. Of all metrics computed, limb motion kinematics and task performance were most significantly affected by target geometry. One possible explanation is that the geometry has a direct and distinct impact on human motion kinematics, whereas the varied geometric configurations may not be enough to elicit a significant effect on physiological response.

Finally, task performance using different control algorithms was affected by the geometric configuration of targets. In general, as shown in Fig. 8, Cartesian space control was the best control strategy for the four targets. It was associated with faster steering speed, higher needle manipulability, straighter paths, and smaller targeting errors. On the other hand, joint space control and steering control show a different trend: the needle velocity variation decreases as insertion distances increase. This indicates users in joint space and steering control modes have smaller and limited capability of manipulation in far-reaching tasks.

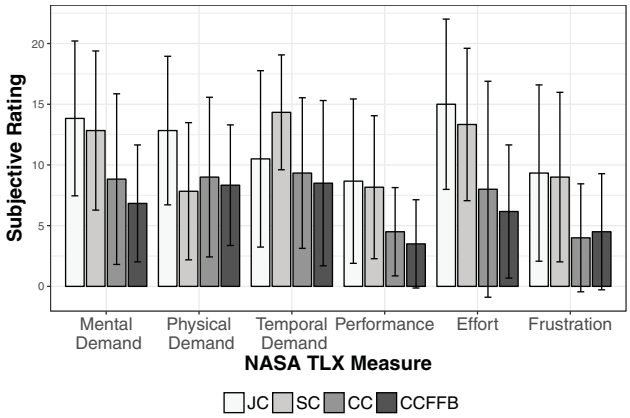


Fig. 9: Subjective multi-dimensional reports of teleoperation algorithms acquired by NASA-TLX user questionnaires. Smaller rating value indicates less perceived task load. Error bars denote the 95% confidence intervals (CI).

B. Comparison of Subjective User Reports

Fig. 9 shows NASA-TLX user reports on teleoperation mapping algorithms in terms of six multi-dimension 20-point subscales: *mental*, *physical*, *and temporal task demands*, *perceived performance*, *effort*, and *frustration*. To identify the significant effects of teleoperation algorithms and subjects on these reports, we conducted ANOVA analysis on each dimension and total aggregated ratings, as shown in Tab. II.

Generally, NASA-TLX results mirror our human-centric evaluation results, though show less statistical significance. As shown from the overall aggregated score, teleoperation using joint space and steering control mappings resulted in the highest average task workload (average total ratings: 70.17 ± 35.17 and 65.50 ± 31.02 respectively), compared to both Cartesian teleoperation algorithms (43.67 ± 34.47 for CC and 37.83 ± 24.40 for CFB). Specifically, cognitive demand of joint space control was significantly higher than for Cartesian space control with force feedback ($p = 0.020$). Both joint space and steering control required significantly higher control effort than Cartesian space control with force

TABLE I: Statistical ANOVA and post-hoc analysis showing the effects of algorithm and target layout on features extracted from real-time physiological response, kinematics of hand, forearm and upper-arm motions, and task-specific performance.

	CATEGORY	Features	Teleoperation Algorithms		Targets	
			p	Significant Groups	p	Significant Groups
Physiological Response	Muscular Activation Level	$PostD_{MNP}$	<0.001	JC>(CFB, SC, CC)	0.352	N/A
		$PectM_{MNP}$	<0.001	JC>(SC, CFB)>CC	0.294	N/A
		$BicB_{MNP}$	<0.001	JC>(CFB, SC)>CC	0.214	N/A
		$PronT_{MNP}$	<0.001	JC>(SC, CFB)>CC	0.408	N/A
	EEG BIOPAC Metrics	*Engagement*	<0.001	(JC, SC)>(CFB, CC)	0.314	N/A
		Workload	<0.001	(JC, SC)>(CFB, CC)	0.047	N/A
	Skin Conductance Level	SC_{avg}	<0.001	JC>(CFB, SC, CC)	0.986	N/A
		SC_{vr}	<0.001	JC>(SC, CFB, CC)	0.257	N/A
	Heart Rate Response	*HR*	<0.001	JC>SC>(CFB, CC)	0.148	N/A
Kinematic Metrics	Hand Motion	$AngVel_{HD}$	<0.001	SC>JC>(CFB, CC)	<0.001	3>2>1, 4>1
		$LinAcc_{HD}$	<0.001	JC>SC>CFB>CC	0.003	(3,4)>1, 3>1
		$Jerk_{HD}$	<0.001	SC>JC>(CC, CFB)	<0.001	(3,2)>4>1
	Forearm Motion	$AngVel_{FA}$	<0.001	JC>SC>CFB>CC	<0.001	(4,3,2)>1
		$LinAcc_{FA}$	<0.001	JC>SC>CFB>CC	0.098	N/A
		$Jerk_{FA}$	<0.001	(SC, JC)>CC>CFB	0.002	(3,2)>4>1
	Upper-arm Motion	$AngVel_{UA}$	<0.001	JC>(CFB, SC)>CC	<0.001	(2,3,4)>1
		$LinAcc_{UA}$	<0.001	JC>(SC, CFB)>CC	0.538	N/A
		$Jerk_{UA}$	<0.001	(JC, SC)>(CC, CFB)	<0.001	(3,2,4)>1
Task Performance	Movement Time	*MT*	<0.001	JC>SC>(CC, CFB)	<0.001	(3,2)>4>1
	Path Straight Deviation	*PathStrDev*	<0.001	(SC, JC)>(CFB, CC)	<0.001	(3,2)>1, 2>4
	SD of Needle Velocity	$NeedleVel_{SD}$	<0.001	CFB>CC>(JC, SC)	0.132	N/A
	Targeting Error	*TargetErr*	<0.001	JC>(SC, CFB)>CC	<0.001	(3,2)>4>1

TABLE II: ANOVA analysis of NASA-TLX subjective ratings on teleoperation mapping algorithms and subjects.

NASA-TLX	Teleoperation Algorithms		Subjects	
	p	Significant Groups	p	Significant Groups
Mental Demand	0.020	JC>CFB	0.003	(1,2,4,6)>3
Physical Demand	0.101	N/A	0.002	4>6>(3,5)
Temporal Demand	0.067	N/A	0.002	(2,4)>3
Performance	0.056	N/A	0.012	5>(3,6)
Effort	0.011	(JC, SC)>CFB	0.004	2>(3,5)
Frustration	0.095	N/A	0.026	4>5
Aggregated Score	0.021	JC>CFB	0.004	(1,2,4,5,6)>3

feedback ($p = 0.011$). There were no statistically significant differences between Cartesian space control with and without feedback; however, it is interesting to note that average values of subjective ratings with the haptic feedback were lower than without feedback for all metrics, except *frustration*.

Also, significant differences were found between subjects on ratings of six-dimension NASA-TLX measures in Tab. II. Specifically, the third subject gave significantly lower scores for all measures, while the second subject gave higher (but not significantly so) ratings on the *mental*, *temporal demand*, and *effort*. This result highlights the limitations of user response surveys, as factors such as uncertainty of word choices, varied levels of user expectations, and inconsistent perceptions of self-performance, could explain these differences.

V. CONCLUSION

In this paper, we presented an novel method for objectively evaluating robotic teleoperation systems through human-centric analysis of physiological, kinematic, and task-specific metrics. This method was applied to evaluate four different types of control algorithms in the case of needle steering. From the experimental user studies, the following is concluded.

1) Cartesian teleoperation results in better task performance with fast insertion, higher accuracy, and increased straightness of needle paths as well as lower cognitive load and smoother movements for users. 2) Position-error-based haptic guidance, widely considered as a valuable navigation tool during teleoperation procedures, might not guarantee an intuitive user experience in needle steering. It was shown to improve steering performance in terms of movement time, consistent with prior work showing enhanced performance with the force feedback in various ways [6], [40], while also reducing the movement smoothness. However, this type of feedback led to increased muscle fatigue, higher cognitive engagement and workload for users. 3) Environmental geometry, which could be a factor of interest in teleoperation, would potentially affect the intuitiveness of control, and result in the variations of needle steering experience. Significant differences between performance and kinematic metrics on the target locations within the user workspace highlighted the impacts of target geometry on motion characteristics.

One limitation of our study is that a simulated needle, as opposed to real needle, was used for experimental evaluation. The presence of a real, physical system may affect user engagement and response. This will be an important consideration for future work.

By applying methods from cognitive analysis, affective computing, and human motor research to the field of teleoperation, it may be possible to design more natural and intuitive interfaces for a variety of human-in-the-loop systems such as surgical robots, rehabilitation robots, and

collaborative manufacturing robots. In future work, we will extend our framework to assess overall states of human users in real-time. This could lay the groundwork for designing adaptive, human-centric controllers for teleoperated robotic systems.

REFERENCES

[1] P. F. Hokayem and M. W. Spong, "Bilateral teleoperation: An historical survey," *Automatica*, vol. 42, no. 12, pp. 2035–2057, 2006.

[2] H. Rafii-Tari, C. J. Payne, and G.-Z. Yang, "Current and emerging robot-assisted endovascular catheterization technologies: a review," *Ann. Biomed. Eng.*, vol. 42, no. 4, pp. 697–715, 2014.

[3] H. I. Son, A. Franchi, L. L. Chuang, J. Kim *et al.*, "Human-centered design and evaluation of haptic cueing for teleoperation of multiple mobile robots," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 597–609, 2013.

[4] R. J. Webster III, J. S. Kim, N. J. Cowan, G. S. Chirikjian, and A. M. Okamura, "Nonholonomic modeling of needle steering," *Intl. J. Robot. Res.*, vol. 25, no. 5-6, pp. 509–525, 2006.

[5] J. M. Romano, R. J. Webster, and A. M. Okamura, "Teleoperation of steerable needles," in *2007 IEEE Int. Conf. on Robot. Autom. (ICRA)*. IEEE, 2007, pp. 934–939.

[6] A. Majewicz and A. M. Okamura, "Cartesian and joint space teleoperation for nonholonomic steerable needles," in *2013 World Haptics Conf. (WHC)*. IEEE, Apr. 2013, pp. 395–400.

[7] C. Pacchierotti, M. Abayazid, S. Misra, and D. Prattichizzo, "Teleoperation of steerable flexible needles by combining kinesthetic and vibratory feedback," *IEEE Trans. Haptics*, vol. 7, no. 4, pp. 551–556, 2014.

[8] M. Li, G. Li, B. Gonenc, X. Duan, and I. Iordachita, "Towards human-controlled, real-time shape sensing based flexible needle steering for mri-guided percutaneous therapies," *Intl. J. Med. Robot. Comput. Assisted Surgery*, vol. 13, no. 2, 2017.

[9] M. Abayazid, C. Pacchierotti, P. Moreira, R. Alterovitz *et al.*, "Experimental evaluation of co-manipulated ultrasound-guided flexible needle steering," *Intl. J. Med. Robot. Comput. Assisted Surgery*, vol. 2, no. 12, pp. 219–230, 2016.

[10] A. Devreker, P. T. Tran, B. Rosa, H. De Praetere *et al.*, "Intuitive control strategies for teleoperation of active catheters in endovascular surgery," *J. Med. Robot. Res.*, vol. 1, no. 03, p. 1640012, 2016.

[11] C. Rossa, J. Fong, N. Usmani, R. Sloboda, and M. Tavakoli, "Multi-actuator haptic feedback on the wrist for needle steering guidance in brachytherapy," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 852–859, 2016.

[12] R. J. Kuiper, D. J. Heck, I. A. Kuling, and D. A. Abbink, "Evaluation of haptic and visual cues for repulsive or attractive guidance in non-holonomic steering tasks," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 672–683, 2016.

[13] A. Steinfeld, T. Fong, D. Kaber, M. Lewis *et al.*, "Common metrics for human-robot interaction," in *Proc. the 1st ACM SIGCHI/SIGART Conf. on Human-robot Interaction*. ACM, 2006, pp. 33–40.

[14] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. in Psycho.*, vol. 52, pp. 139–183, 1988.

[15] D. B. Kaber, J. M. Riley, R. Zhou, and J. Draper, "Effects of visual interface design, and control mode and latency on performance, telepresence and workload in a teleoperation task," in *Proc. Human Factors and Ergonomics Soc. Annual Meeting*, vol. 44, no. 5, 2000, pp. 503–506.

[16] E. M. Hufnagel and C. Conca, "User response data: The potential for errors and biases," *Inf. Syst. Res.*, vol. 5, no. 1, pp. 48–73, 1994.

[17] B. Willaert, D. Reynaerts, H. Van Brussel, and E. B. Vander Poorten, "Bilateral teleoperation: quantifying the requirements for and restrictions of ideal transparency," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 1, pp. 387–395, 2014.

[18] D. A. Lawrence, "Stability and transparency in bilateral teleoperation," *IEEE Trans. Robot. Autom.*, vol. 9, no. 5, pp. 624–637, 1993.

[19] N. Diolaiti, G. Niemeyer, F. Barbagli, and J. K. Salisbury, "Stability of haptic rendering: Discretization, quantization, time delay, and coulomb effects," *IEEE Trans. Robot.*, vol. 22, no. 2, pp. 256–268, 2006.

[20] J. C. Gwilliam, M. Mahvash, B. Vagvolgyi, A. Vacharat *et al.*, "Effects of haptic and graphical force feedback on teleoperated palpation," in *2009 IEEE Int. Conf. on Robot. Autom. (ICRA)*. IEEE, 2009, pp. 677–682.

[21] W. McMahan, J. Gewirtz, D. Standish, P. Martin *et al.*, "Tool contact acceleration feedback for telerobotic surgery," *IEEE Trans. Haptics*, vol. 4, no. 3, pp. 210–220, 2011.

[22] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady, "Evaluating display fidelity and interaction fidelity in a virtual reality game," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 4, pp. 626–633, 2012.

[23] T. O. Zander, K. Shetty, R. Lorenz, D. R. Leff *et al.*, "Automated task load detection with electroencephalography: Towards passive brain–computer interfacing in robotic surgery," *J. Med. Robot. Res.*, vol. 2, no. 01, p. 1750003, 2017.

[24] C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau *et al.*, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviation, Space, and Environ. Med.*, vol. 78, pp. B231–B244, 2007.

[25] C. A. Vernooij, G. Rao, D. Perdikis, R. Huys *et al.*, "Functional coordination of muscles underlying changes in behavioural dynamics," *Sci. Rep.*, vol. 6, 2016.

[26] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Trans. Syst.*, vol. 6, no. 2, pp. 156–166, 2005.

[27] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen, "Galvanic skin response (GSR) as an index of cognitive load," in *2007 Extended Abstracts on Hum. Factors in Comput. Syst. (CHI)*. ACM, 2007, pp. 2651–2656.

[28] R. Castaldo, P. Melillo, U. Bracale, M. Caserta *et al.*, "Acute mental stress assessment via short term hrv analysis in healthy adults: A systematic review with meta-analysis," *Biomed. Signal Proc. Control*, vol. 18, pp. 370–377, 2015.

[29] M. Swangnetr and D. B. Kaber, "Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 1, pp. 63–75, 2013.

[30] N. Nordin, S. Q. Xie, and B. Wünsche, "Assessment of movement quality in robot-assisted upper limb rehabilitation after stroke: a review," *J. Neuroeng. Rehabil.*, vol. 11, no. 1, p. 137, 2014.

[31] E. B. Mazomenos, P.-L. Chang, R. A. Rippel, A. Rolls *et al.*, "Catheter manipulation analysis for objective performance and technical skills assessment in transcatheter aortic valve implantation," *Intl. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 6, pp. 1121–1131, 2016.

[32] I. Nisky, M. H. Hsieh, and A. M. Okamura, "Uncontrolled manifold analysis of arm joint angle variability during robotic teleoperation and freehand movement of surgeons and novices," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2869–2881, 2014.

[33] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, 2001.

[34] R. R. Johnson, D. P. Popovic, R. E. Olmstead, M. Stikic *et al.*, "Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model," *Biol. Psychol.*, vol. 87, no. 2, pp. 241–250, 2011.

[35] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere *et al.*, "Emg feature evaluation for improving myoelectric pattern recognition robustness," *Expert Systems with Applications*, vol. 40, no. 12, pp. 4832–4840, 2013.

[36] M. Stikic, C. Berka, D. J. Levendowski, R. F. Rubio *et al.*, "Modeling temporal sequences of cognitive state changes based on a combination of EEG-engagement, EEG-workload, and heart rate metrics," *Front. Neurosci.*, vol. 8, 2014.

[37] A. C. Granero, F. Fuentes-Hurtado, V. N. Ornedo, J. G. Provinciale *et al.*, "A comparison of physiological signal analysis techniques and classifiers for automatic emotional evaluation of audiovisual contents," *Front. Comput. Neurosci.*, vol. 10, 2016.

[38] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand, "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography," *J. Med. Eng. Tech.*, vol. 32, no. 6, pp. 479–484, 2008.

[39] S. Estrada, C. Duran, D. Schulz, J. Bismuth *et al.*, "Smoothness of surgical tool tip motion correlates to skill in endovascular tasks," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 647 – 659, 2016.

[40] M. K. OMalley, A. Gupta, M. Gen, and Y. Li, "Shared control in haptic systems for performance enhancement and training," *J. Dynamic Syst., Meas., Control*, vol. 128, no. 1, pp. 75–85, 2006.