Human-as-a-Robot Performance in Mixed Reality Teleultrasound

David Black[1*] and Septimiu Salcudean[1]

[1*]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada.

*Corresponding author(s). E-mail(s): dgblack@ece.ubc.ca; Contributing authors: tims@ece.ubc.ca;

Abstract

Purpose: In "human teleoperation", mixed reality (MR) and haptics are used to tightly couple an expert leader to a human follower [1]. To determine the feasibility of human teleoperation for teleultrasound, we quantify the ability of humans to track a position and/or force trajectory via MR cues. The human response time, precision, frequency response, and step response were characterized, and several rendering methods were compared. **Methods:** Volunteers (n=11) performed a series of tasks as the follower in our human teleoperation system. The tasks involved tracking pre-recorded series of motions and forces, each time with a different rendering method. The order of tasks and rendering methods was randomized to avoid learning effects and bias. The volunteers then performed a series of frequency response tests and filled out a questionnaire. **Results:** Rendering the full ultrasound probe as a position target with an error bar displaying force led to the best position and force tracking. Following force and pose simultaneously was more difficult but did not lead to significant performance degradation versus following one at a time. On average, subjects tracked positions, orientations, and forces with RMS tracking errors of $\mathbf{6.2 \pm 1.9}$ mm, $\mathbf{5.9 \pm 1.9°}$, $\mathbf{1.0 \pm 0.3}$ N, steady-state errors of $\mathbf{2.8 \pm 2.1}$ mm, $\mathbf{0.26 \pm 0.2}$ N, and lags of $\mathbf{345.5 \pm 87.6}$ ms respectively. Performance decreased with input frequency, until the person could no longer follow, depending on the input amplitude. **Conclusion:** This paper characterizes human tracking ability in MR human teleoperation for teleultrasound, which is important for designing future tightly-coupled guidance and training systems using MR.

Keywords: Teleoperation, Human Computer Interaction, Ultrasound, Robotics, AR/VR

1 Introduction

Teleguidance is becoming more important in a wide range of fields, from remote maintenance and monitoring [2] to telemedicine [3]. The latter has grown in importance particularly in the last years due to the COVID-19 pandemic. One important and heavily studied form of telemedicine is teleultrasound (TUS). In the last two years alone, this has found applications in rural or under-resourced environments [4, 5], for COVID-19 safety [6], and for training [7, 8]. In [1], a novel teleultrasound system dubbed "human teleoperation" was introduced based on mixed reality (MR). In brief, a novice person carries out an US procedure by following the motion and force of a virtual US probe projected onto the patient via a mixed reality headset. The virtual probe is controlled in real time by a remote expert who sees the US images and a video stream of the patient. This improves performance over video conference-based teleguidance by creating a tighter coupling between the leader and follower, yet is more flexible and accessible than robotic TUS.

MR has long been used to augment a person's sensory capabilities by overlaying medical images or volumetric models in situ for ultrasound-guided needle biopsies [9, 10], robot-assisted surgery [11], laparoscopic surgery [12, 13], and transcranial magnetic stimulation [14]. Some implementations also include static guides illustrating where to align an instrument [15, 16]. Recent work on "holoportation" has attempted to render a realistic avatar of a person in another person's headset to provide more meaningful remote social interactions [17]. However, to our knowledge no other system has used dynamic hand-over-hand guidance to enable an expert to effectively teleoperate a novice person. This has traditionally been the purpose of robots, whose performance characteristics are precisely known before use. However, given the novelty of the concept, the performance characteristics of the human follower in human teleoperation have never been studied.

Using a prototype system fully implemented in [18], it was possible to characterize human capabilities and performance limitations in human teleoperation. Similar research was carried out for the development of haptic devices. Stocco and Salcudean surveyed human hand capabilities to inform haptic interface design [19] while others have tested maximum hand torque [20] or just-noticeable-differences (JND) in force [21]. Of particular interest are the visual-motor response time (RT), frequency response of the hand, and accuracy in positioning and force exertion. Tan compiled several measures including an approximate JND for force sensing of 7%, a JND for wrist angle of 2°, and a resolution for applied forces of 0.36 N [22]. The spatial resolution of hand movements is approximately 0.5-2.5 mm [23]. It was found that the somatosensory system can perceive vibrotactile stimuli up to 1 kHz, though the ability to apply forces is far more limited, with a force control bandwidth of approximately 20-30 Hz. In practice, however, the bandwidth is closer to 7 Hz and differs between unexpected signals (1-2 Hz), periodic signals (2-5Hz), internally-generated or learned trajectories (5 Hz), and reflexive actions (10 Hz) [24].

The human response time has been studied carefully since the late 1800s. Wargo described the delay of every step in the processing chain, from receptor delays (15-38 ms), to afferent neural transmission (5-10 ms), central processing (90-300 ms), efferent transmission (10-20 ms), and muscle activation (1 ms) [25]. Keele and Posner measured 190-260 ms visual processing time [26], while Beggs and Howarth found it to be $<$ 300 ms [27]. More recently, Liu et al. were able to decode object categories from intracranial field potentials in the visual cortex in as little as 100 ms after onset of stimulus [28].

While these values constitute approximate limits in expected human performance, they do not describe the ability of a human to track an input signal, whether visual, haptic, or auditory. This ability to track has great implications for AR/VR interfaces and teleguidance systems, enabling technologies such as our human teleoperation system. This paper therefore characterizes human tracking performance, using mixed reality input to follow positions, orientations, and forces, both individually and simultaneously. We determine frequency responses, tracking bandwidth, root-mean-square (RMS) error, steady state error, reaction time, and rise time, among other important factors. Step response tests with the system are described in [18].

These results are important not only to determine the feasibility and expected effectiveness of our specific TUS system, but also more broadly to inform future mixed or virtual reality human computer interfaces. The expected performance measures give designers a baseline in building new applications, including how fast a person will react or follow virtual cues, how precisely a person will move or apply forces, and what type of rendering the human brain reacts to best. The speed and precision values also allow such a system to be modeled, which will be essential in designing teleoperation controllers.

2 Methods

The human teleoperation prototype used in this paper was introduced in [1]. Improvements were presented in [29]. Human teleoperation for ultrasound (US) involves an expert sonographer at a medical centre (the “expert”), and a novice “follower” in a remote location who carries out the US examination. The follower wears a mixed reality headset (Microsoft HoloLens 2) which projects a three-dimensional virtual US probe into the leader’s visual field. This virtual probe is controlled in real time using a haptic device (3DSystems Touch X) by the expert, who sees the US images and video feed from the HoloLens in real time, and receives force feedback via the haptic device. The follower matches

Fig. 1: System Overview: a) electromagnetic tracker (on device and separate, with thumb tack for scale); b) dummy US probe; c) force sensor; d) electromagnetic sensing transmitter with ArUco markers for registration; e) follower with HoloLens 2; f) follower side user interface; g) force rendering schemes (color and error-bar) with full and partial probe.

the position and orientation (pose) and force of the virtual probe as it moves according to the expert's input, thus achieving teleoperation.

The system logs all data during the tests and can play back prerecorded or generated motion and force sequences on the follower side. An overview is shown in Fig. 1. The follower was given a 3D printed object shaped exactly like a Clarius C3HD3 probe, but with a built-in 6-axis force/torque sensor at the tip (ATI Nano25), and an electromagnetic pose sensor (driveBAY, Northern Digital) embedded in the hand grip, as far from the metal force sensor as possible. In this way, the pose and force of the follower were measured throughout the tests. Optical pose tracking was also evaluated due to its potentially better accuracy, but it was found to be impractical due to repeated occlusion of the reflective markers during motions.

The sensors were connected to a Windows PC running a .NET program which read in the data and sent it to the HoloLens via WiFi, over a WebRTC connection, which adds up to 5 ms delay to the measurements [18]. This delay is measured in real time, and the reported latencies in the results are adjusted accordingly. The electromagnetic force sensor readings were transformed to the HoloLens frame by placing two ArUco markers [30] in known positions on the magnetic transmitter, which defines the sensor frame. The HoloLens detects the pose of both markers (Mitchell Doughty GitHub[1]) and finds the optimal transform between the frames. According to the data sheets, the electromagnetic tracking is accurate to 1.4 mm and 0.5°, and the force sensor has 0.02-0.06 N resolution.

To test rendering schemes for the human computer interaction, 4 renderings of the virtual probe were tested. For pose teleoperation, an US probe shape was tested, as was the same shape with the middle removed (Fig. 1), to test

[1]https://github.com/doughtmw/ArUcoDetectionHoloLens-Unity

whether occlusion of the real probe by the hologram affects control accuracy. For forces, two schemes were tested: (1) changing the color of the virtual probe continuously from blue (too little force) to green (good force) to red (too much force) and (2) using an error-bar (EB) which grew away from the patient and turned blue (too little force), shrunk down and turned green (good force), and grew toward the patient and turned slowly red (too much force) (Fig. 1). The calculation involved subtracting the measured force from the desired one.

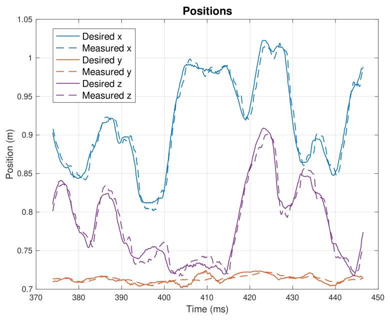
Volunteers ($n = 11$, aged 20-64 years, mean 32. 36% female) were recruited to perform a series of tasks. By design, the volunteers came from a variety of backgrounds including engineering graduate and undergraduate students, a medical student, a physicist, a surgeon, an engineering professor, a lawyer, and an architecture student. They were first given a demonstration of how to use the system and allowed to try it for 1 minute. The tasks included following 2 sequences of prerecorded motions and 2 sequences of prerecorded forces (single parameter tracking), and 4 sequences in which motions (tangent to surface) and forces (normal to rigid surface) were tracked simultaneously (dual parameter tracking). Each sequence was 1-3 minutes long, and each subject carried out the exact same sequences. During each single parameter test, a different one of the four rendering schemes was used. During the dual parameter tests, all four combinations of the rendering schemes were used. The order of rendering schemes was randomized to avoid learning effects or bias due to the prerecorded motions being slightly different from each other.

Next, the subjects performed a frequency response test (Section 3.2) for both pose and force. An equally relevant step response test is described in [18]. The frequency response consisted of a sinusoidal input signal which increased in frequency every 5 oscillations until the follower was unable to track it.

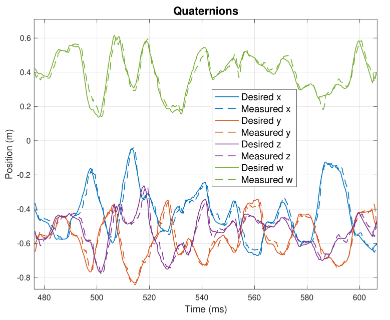
All desired and measured force, position, and orientation data was saved with timestamps at a sampling rate of approximately 36 Hz on the HoloLens to avoid clock drift between devices, and was analyzed using MATLAB (Mathworks Inc.). Desired and measured values were aligned using the timestamps and resampled so each sample lined up in time. Resampling was performed by linear interpolation for position and force data, and spherical interpolation for quaternions (slerp.m in MATLAB). The desired and measured signals were then subtracted from each other element-wise to determine errors. Tracking lags were calculated by finding the time delay that minimized the normalized cross-correlation between the measured and input signal.

In addition to the numerical data, the volunteers filled out a questionnaire. This aimed to determine how the users perceived the tasks in terms of physical and cognitive difficulty, and which rendering schemes they preferred. This could then be compared to the numerical results and gives useful insight into the challenge of human teleoperation from the follower's perspective.

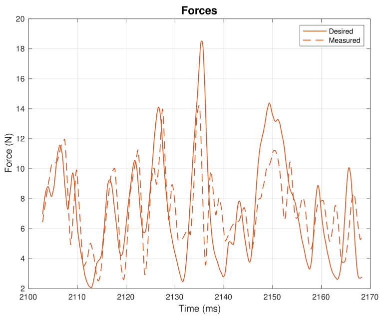
Statistical significance was measured using the two-sample t-test or the two-sample Kolmogorov-Smirnov test, depending on the distribution. Ethical approval for these minimal risk human tests was obtained from the UBC Behavioural Research Ethics Board, with approval number H22-01195.



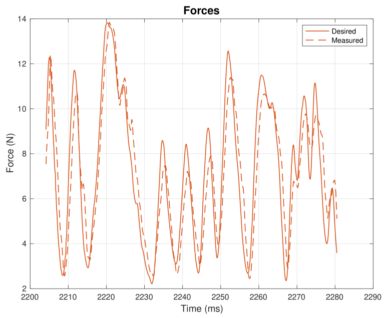
(a) Position



(b) Orientation



(c) Color Force



(d) Error-bar Force

Fig. 2: Example motion tracking test (a,b) and force tracking test (c,d) from one user whose results were approximately average. Solid lines are the input signal while dotted lines are the measured motion or force. The force tracking is plotted for two different rendering schemes, showing clear performance improvement using the error-bar.

Informed consent was obtained from all participants, and all methods were carried out in accordance with relevant regulations. No identifying information was recorded.

3 Results

3.1 Tracking Tests

The first eight tests every volunteer performed involved following prerecorded series of motions and forces. Example position and orientation tracking are shown in Figs. 2a and 2b for an average user. Example force tracking is shown in Figs. 2c and 2d. The single parameter tracking results are outlined in Table 1 while the dual parameter results are in Table 2.

There is no statistically significant difference between the full and partial rendering for pose tracking accuracy or lag, though the mean lag is slightly better with full probe, while the mean tracking accuracy is better with partial probe. This indicates that the full probe does cause some accuracy problems by occluding the real probe, which several users commented on. This is very dependent on the ambient light; in a more well-lit environment, the occlusion was much less. Additionally, the user interface includes an opacity setting in which the user can adjust the alpha level of the probe color. From the questionnaire, testers consistently had no preference between partial and full.

	Full	Partial	Color	Error-bar
Tracking Lag (ms)	346.23 ± 118.15	358.91 ± 57.53	469 ± 107.10	255 ± 118.88
p-value	0.76		< 0.001	
RMS Error	8.1 ± 1.7 mm $7.7 \pm 2.5°$	6.24 ± 1.93 mm $5.93 \pm 1.85°$	1.69 ± 0.43 N	0.99 ± 0.29 N
p-value	0.38		0.001	

Table 1: Single parameter tracking results (average $\pm$ standard deviation)

On the other hand, for force tracking, the error-bar is clearly superior, which can be seen in Fig. 2. On the questionnaire, all users indicated a strong preference for the error-bar, with all users commenting that it was much easier to see when the force error was zero by seeing the error-bar disappear as opposed to trying to see if the probe color was truly green, or had a slight blue or red tinge. Additionally, the tracking lag is significantly better with error-bar than with color force. Hence, error-bar force tracking was used for all the step and frequency response tests.

The next four tests involved tracking force and pose simultaneously. The overall results are found in Table 2. Despite the large differences in the single parameter tracking, none of the differences between the four rendering methods is significant in dual parameter tracking, except that Full+EB is significantly faster than Partial+EB ($p = 0.014$) and is the fastest on average. It is possible that with the Partial+EB rendering, the follower is presented with too much information, which is good for accuracy but impacts the tracking lag.

It is also important to consider the performance difference between single and dual parameter tracking. The overall performance differences are shown

	Full+Color	Full+EB	Partial+Color	Partial+EB
Tracking Lag (ms)	408.72 ± 175.36	345.5 ± 87.60	435.28 ± 219.90	478.13 ± 126.46
RMS Error	8.7 ± 1.6 mm $7.00 \pm 2.31°$ 1.46 ± 0.33N	8.5 ± 1.4 mm $7.27 \pm 2.25°$ 1.25 ± 0.33N	8.9 ± 1.9 mm $8.11 \pm 2.65°$ 1.40 ± 0.23 N	8.6 ± 2.5 mm $7.48 \pm 1.97°$ 1.26 ± 0.20 N

Table 2: Dual parameter tracking results (average $\pm$ standard deviation)

in Table 3, showing some performance detriment from tracking both parameters at once, though only rotation error is statistically significant. Looking at the individual rendering schemes, EB was significantly faster than Full+Color ($p = 0.045$), Partial+Color ($p = 0.050$), and Partial+EB ($p < 0.001$), but not Full+EB. Both Full and Partial probe rendering resulted in significantly faster pose tracking than Partial+EB ($p = 0.032$ and 0.018 respectively). No other difference between single and dual rendering modes was significant for delay or position or orientation tracking. On the other hand, several of the pairings are statistically significant in the force tracking, as seen in Table 4. Here it becomes clear that dual parameter renderings involving color force are significantly worse than other renderings. Conversely, Full+EB performs about as well as tracking just force.

	Single	Dual	p-value
Position Error (mm)	7.9 ± 2.1	8.6 ± 1.9	0.16
Rotation Error (°)	6.08 ± 1.84	7.44 ± 2.18	0.02
Force Error (N)	1.32 ± 0.28	1.34 ± 0.50	0.81
Lag (ms)	362.29 ± 125.73	431.81 ± 183.93	0.062

Table 3: Overall single versus dual parameter tracking

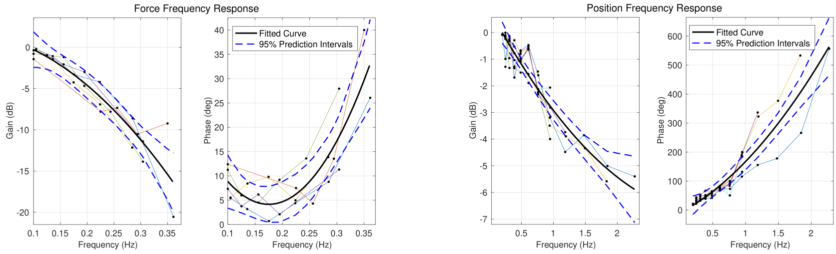
	Full+Color	Full+EB	Partial+Color	Partial+EB
Color	**0.019**	0.27	**0.010**	0.15
EB	**0.011**	0.090	**0.012**	**0.031**

Table 4: p-values of differences between single and dual parameter force tracking RMS error. Significant differences are bold. Full+EB dual parameter rendering is not significantly worse than the single parameter schemes.

All participants stated in the questionnaire that dual parameter tracking constituted a greater cognitive load than single parameter tracking (average score $(4.36 \pm 0.81)/5$ where 5 = much harder and 1 = much easier). Additionally, force tracking was more mentally demanding ($(3.36 \pm 0.50)/5$) than pose tracking $(2.18 \pm 0.60)/5$). The system scored $(2.36 \pm 0.81)/5$ for physical demand (where 5 = very demanding), and no participant became dizzy.

3.2 Frequency Response

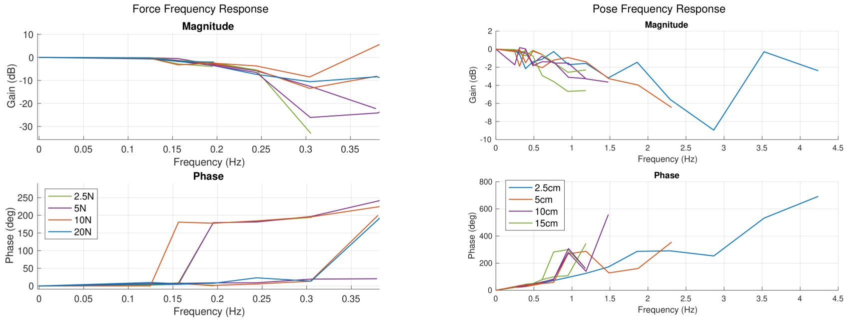
The frequency responses are plotted in Fig. 3, with a fitted average response and 95% confidence interval. The fitted curves are second degree polynomials, which were found to give the best fit. We see that in tracking position, some users managed to get to 2Hz, although with large phase lag. For phase lag less than 180°, users could follow up to 1 Hz. In this range, the phase delay



(a) Force (b) Position

Fig. 3: Frequency Responses. The black line is the fitted curve, and the dotted line is 95% CI. Phase refers to phase lag, where the follower lags the expert.

trend was uniform among users and the gain was > -3 dB. Tracking forces was substantially more difficult despite not actually having to move. This is likely due to the less directly intuitive visual force control. While phase delay remained fairly small, it quickly became impossible to follow signals faster than about 0.35 Hz. There was a strong trend among all participants where phase angle initially decreased slightly with frequency. This is possibly because users were initially careful to be as accurate as possible, but quickly switched their focus to following quickly enough. It is unlikely that the users learned and improved during the low frequency, since these tests were carried out last. Force tracking gain decreased rapidly with frequency as users failed to match the desired force quickly enough before it changed again. Relatively good force performance could be achieved at less than about 0.25 Hz.



(a) Force (b) Position

Fig. 4: Frequency Responses at different input signal amplitudes. For position, smaller signal amplitudes are generally easier to follow, while the trend is less clear for force.

We next look at the dependency of the frequency response on the input signal amplitude. Stocco found that the human hand force bandwidth was 7 Hz for small motions but dropped past 5 Hz for motions of 160 mm [19]. From preliminary tests with a subset of the volunteers, we obtain the responses in Fig. 4. We see that the users are able to continue following positions for longer as the motion amplitude decreases, but that the trend in gain and phase is about the same, irrespective of amplitude. This implies that the users are physically able to make fast enough motions (as shown in [19]), but that their ability to follow is cognitively or visually limited as the virtual probe begins to move too fast. With force tracking we see the opposite trend because no motion is required. Instead, as the desired force differences become very small and approach the human hand's JND, it becomes hard to match them precisely, so the smaller forces are harder to follow.

Closely related tests of step response were carried out with the same volunteers and measurement system and are presented in [18]. Due to space constraints, they could not be included here, but human response time was found to be 628.3 ± 102.3 ms and 171.5 ± 85.9 ms, with steady state RMS error of 2.8 ± 2.1 mm and 0.26 ± 0.16 N for position and force respectively.

4 Discussion

In this paper we have introduced a measurement system based on the human teleoperation concept from [1], and used it to perform a number of measurements of human performance in mixed reality pose and force teleoperation. Additionally, different rendering schemes were compared to determine the best human-computer interface.

It was found that the error-bar force rendering was superior to color-based force, and there was little difference between full and partial probe rendering, but that the opacity of the probe is a matter of personal preference and should thus continue to be adjustable. With full probe + error-bar rendering, tracking both pose and force simultaneously did not lead to a statistically significant decrease in performance, though it was more mentally demanding. This shows that high performing human teleoperation, in which poses and forces are tracked simultaneously, accurately, and quickly, is feasible. For US procedures, motions are generally slow and do not coincide temporally with large force variations, so the requirement for simultaneous pose and force tracking is much less strict than those considered in this paper. However, for other applications of human teleoperation, requirements may be more stringent. While this study shows the potential for good performance in human teleoperation, it does not assess the feasibility in real-world scenarios including sonographers and patients. Future work will include a clinical human study to build off the preliminary work in [1].

The visual force renderings tested here are not the only possibilities. For example, a second virtual probe could be rendered, offset from the primary one in a direction and by an offset proportional to the force error. Alternatively, an arrow could be used instead of the error-bar. Both of these methods include directional information which was lacking in the renderings tested here. However, forces in US are generally normal to the surface of the patient. Thus, single-axis force rendering may be sufficient or even preferable as excess information may slow down the follower, as seen in the Partial+EB rendering in Section 3.1. Having a second virtual probe could also lead to confusion. This was the motivation for the two tested force rendering schemes. They are relatively simple and allow the user to keep their eyes on the US probe.

A third alternative which will be explored in future work is to offset the virtual probe itself from the desired position by an amount proportional to the force error and an estimated impedance of the patient tissue. In this way, by matching the rendered pose, the follower would automatically match the desired force as well, while effectively only tracking a single parameter at a time. However, for relatively stiff tissue, this would rely on the follower tracking very small differences in position. As shown in [18], the position resolution is approximately 3 mm. Finally, it is possible that using a color map other than red/green/blue would lead to better tracking, depending on the preferential sensitivity of the human eye to certain changes in color [31].

Conversely, it has been shown that the human auditory system is faster than the visual [25, 32]. Brooks describes a visual perception bandwidth of 50 Hz compared to auditory perception at 20 kHz [24]. Thus, force feed-forward could potentially also be achieved through auditory signals. However, unfamiliar auditory signals are known to inhibit visual processing abilities [33]. Thus, this may not lead to better overall performance.

In the frequency response, the motion of the virtual probe repeated sinusoidally, so, barring the constant increases in frequency, the follower could anticipate where the probe would move next, which could affect the performance. We considered having a frequency response test in which the direction changed after every period. However, this instant change in direction introduces infinite frequencies into the signal and has an unknown effect on the response, likely leading to artifacts and less comparability to other frequency response tests in literature. Instead, the ability of a person to anticipate motions may be a benefit of human teleoperation over non-intelligent robots.

Several frequency response measures are outlined in Section 1 which can be compared to the values found in this study. The bandwidth of the human hand was found to be approximately 5-7 Hz [19, 24]. While this measures how fast a human hand can move or control forces, our study explored the ability of people to follow a desired signal, which is likely limited by the visual and cognitive systems rather than the hand. Indeed, for decent performance, the volunteers tracked up to 1 Hz input signals. This is approximately in the same order of magnitude as the maximum hand capabilities.

Several users commented that they were becoming fatigued by the end of the tests, after prolonged intense focus and visual stimulation. It is therefore likely that a well rested group of volunteers performing the frequency and step responses again would perform better, although the difference would likely be relatively small. This is also one reason why we did not test further rendering schemes. Additionally, it has been shown that this type of behaviour is a learned skill [34, 35], so with training we might also expect improved results.

The described tests were performed on a TUS system. Correct pose and force are important in US to obtain good quality images, and are very difficult for a novice user to achieve without tightly-coupled expert guidance. While pose ensures the correct anatomy is visualized, force is equally important for maintaining proper contact. The sonographer must avoid deformation of some anatomies or push down surprisingly hard under the ribs to view others. The force control thus requires accurate feedback of the forces or replication of the patient on the expert side, which will be the topic of future work and is discussed in [1]. This paper focuses on the human MR tracking ability in TUS, but the only US-specific factor in the presented tests is the shape of the virtual probe. Additionally, most motion and force sequences were faster and more difficult to follow than typical US exams. Thus, the results are expected to generalize well to other applications of human teleoperation. Indeed, there are many other applications where such teleoperation could be useful, including remote maintenance, inspection, and teaching.

5 Conclusion

In this paper we have measured human performance in tracking a desired pose and force sequence through an MR interface. While tracking was found to be strongly dependent on the rendering scheme, humans can track pose and force simultaneously with a lag of 345.5 ± 87.6 ms, and RMS tracking errors of 8.5 ± 1.4 mm, $7.27 \pm 2.25°$, and 1.25 ± 0.33 N. Steady state errors are significantly better, at 2.8 ± 2.1 mm and 0.26 ± 0.16 N. Tracking of signals with good performance is possible up to a bandwidth of about 1 Hz for position and 0.25 Hz for forces, both of which depend on the magnitude of the input signal. Rendering the full ultrasound probe with error-bar based force feedback was most effective, and tracking both force and pose simultaneously was cognitively more difficult but did not lead to statistically significant degradation in performance. Ultimately, these values and results can serve as a guide for the design of future MR (or AR/VR) interfaces, and demonstrate that human teleoperation for teleultrasound and likely other applications can provide good performance.

Declarations

Funding

We gratefully acknowledge scholarship support from the Vanier Canada Graduate Scholarships program and financial support from the National Science and Engineering Research Council of Canada (NSERC, grant number RGPIN-2016-04618) and the Charles Laszlo Chair in Biomedical Engineering, as well as partial funding, equipment, and technical support for the communication system from Rogers Communications.

Potential Conflicts of Interest

The authors have no relevant financial or non-financial interests to disclose beyond the funding sources listed above.

Ethics Approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the University of British Columbia Behavioural Research Ethics Board with approval number H22-01195 (June 20, 2022)

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

References

[1] Black, D., Yazdi, Y.O., Hosseinabadi, A.H.H., Salcudean, S.: Human teleoperation - a haptically enabled mixed reality system for teleultrasound. Human Computer Interaction (2022)

[2] Yamamoto, T., Otsuki, M., Kuzuoka, H., Suzuki, Y.: Tele-guidance system to support anticipation during communication. Multimodal Technologies and Interaction **2**(3), 55 (2018)

[3] Gajarawala, S.N., Pelkowski, J.N.: Telehealth benefits and barriers. The Journal for Nurse Practitioners **17**(2), 218–221 (2021)

[4] Jemal, K., Ayana, D., Tadesse, F., Adefris, M., Awol, M., Tesema, M., Dagne, B., Abeje, S., Bantie, A., Butler, M., Nwoke, C., Kanyuka, Z., Adams, S.J., Mendez, I.: Implementation and evaluation of a pilot antenatal ultrasound imaging programme using tele-ultrasound in ethiopia. Journal of Telemedicine and Telecare (2022)

[5] Kaneko, T., Kagiyama, N., Nakamura, Y., Hirasawa, T., Murata, A., Morimoto, R., Miyazaki, S., Minamino, T.: Effectiveness of real-time tele-ultrasound for echocardiography in resource-limited medical teams. Journal of Echocardiography **20**(1), 16–23 (2022)

[6] Uschnig, C., Recker, F., Blaivas, M., Dong, Y., Dietrich, C.F.: Tele-ultrasound in the era of covid-19: A practical guide. Ultrasound in Medicine & Biology (2022)

[7] Drake, A.E., Hy, J., MacDougall, G.A., Holmes, B., Icken, L., Schrock, J.W., Jones, R.A.: Innovations with tele-ultrasound in education sonography: the use of tele-ultrasound to train novice scanners. The ultrasound journal **13**(1), 1–8 (2021)

[8] Soni, N.J., Boyd, J.S., Mints, G., Proud, K.C., Jensen, T.P., Liu, G., Mathews, B.K., Schott, C.K., Kurian, L., LoPresti, C.M., Andrus, P., Nathanson, R., Smith, N., Haro, E.K., Mader, M.J., Pugh, J., Restrepo, M.I., Lucas, B.P.: Comparison of in-person versus tele-ultrasound point-of-care ultrasound training during the covid-19 pandemic. The ultrasound journal **13**(1), 1–7 (2021)

[9] Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. ACM SIGGRAPH Computer Graphics **26**(2), 203–210 (1992)

[10] Rosenthal, M., State, A., Lee, J., Hirota, G., Ackerman, J., Keller, K., Pisano, E.D., Jiroutek, M., Muller, K., Fuchs, H.: Augmented reality guidance for needle biopsies: an initial randomized, controlled trial in phantoms. Medical Image Analysis **6**(3), 313–320 (2002)

[11] Kalia, M., Avinash, A., Navab, N., Salcudean, S.: Preclinical evaluation of a markerless, real-time, augmented reality guidance system for robot-assisted radical prostatectomy. International Journal of Computer Assisted Radiology and Surgery **16**(7), 1181–1188 (2021)

[12] Navab, N., Traub, J., Sielhorst, T., Feuerstein, M., Bichlmeier, C.: Action-and workflow-driven augmented reality for computer-aided medical procedures. IEEE computer graphics and applications **27**(5), 10–14 (2007)

[13] Kolagunda, A., Sorensen, S., Mehralivand, S., Saponaro, P., Treible, W., Turkbey, B., Pinto, P., Choyke, P., Kambhamettu, C.: A mixed reality guidance system for robot assisted laparoscopic radical prostatectomy. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, pp. 164–174. Springer, ??? (2018)

[14] Leuze, C., Yang, G., Hargreaves, B., Daniel, B., McNab, J.A.: Mixed-reality guidance for brain stimulation treatment of depression. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 377–380 (2018). IEEE

[15] Nee, A.Y., Ong, S., Chryssolouris, G., Mourtzis, D.: Augmented reality applications in design and manufacturing. CIRP annals **61**(2), 657–679 (2012)

[16] Mourtzis, D., Zogopoulos, V., Vlachou, E.: Augmented reality application to support remote maintenance as a service in the robotics industry. Procedia Cirp **63**, 46–51 (2017)

[17] Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A.,

Degtyarev, Y., Kim, D., Davidson, P.L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Qin, C., Chou, P.A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S.B., Kohli, P., Lutchyn, Y., Keskin, C., Izadi, S.: Holoportation: Virtual 3d teleportation in real-time. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 741–754 (2016)

[18] Black, D., Salcudean, S.: A mixed reality system for human teleoperation in tele-ultrasound: Communication performance. TechRxiv (2022)

[19] Stocco, L., Salcudean, S.E.: A coarse-fine approach to force-reflecting hand controller design. In: Proceedings of IEEE International Conference on Robotics and Automation, vol. 1, pp. 404–410 (1996). IEEE

[20] Adams, S.K., Peterson, P.J.: Maximum voluntary hand grip torque for circular electrical connectors. Proceedings of the Human Factors Society Annual Meeting **30**(8), 847–851 (1986)

[21] Zadeh, M.H., Wang, D., Kubica, E.: Human factors for designing a haptic interface for interaction with a virtual environment. In: 2007 IEEE International Workshop on Haptic, Audio and Visual Environments and Games, pp. 21–26 (2007). IEEE

[22] Tan, H.Z., Srinivasan, M.A., Eberman, B., Cheng, B.: Human factors for the design of force-reflecting haptic interfaces. Dynamic Systems and Control **55**(1), 353–359 (1994)

[23] Kaber, D.B., Zhang, T.: Human factors in virtual reality system design for mobility and haptic task performance. Reviews of Human Factors and Ergonomics **7**(1), 323–366 (2011)

[24] Brooks, T.L.: Telerobotic response requirements. In: 1990 IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings, pp. 113–120 (1990). IEEE

[25] Wargo, M.J.: Human operator response speed, frequency, and flexibility: A review and analysis. Human factors **9**(3), 221–238 (1967)

[26] Keele, S.W., Posner, M.I.: Processing of visual feedback in rapid movements. Journal of experimental psychology **77**(1), 155 (1968)

[27] Beggs, W., Howarth, C.: Movement control in a repetitive motor task. Nature **225**(5234), 752–753 (1970)

[28] Liu, H., Agam, Y., Madsen, J.R., Kreiman, G.: Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron **62**(2), 281–290 (2009)

[29] Black, D., Salcudean, S.: A mixed reality system for human teleoperation in tele-ultrasound. Technical report, EasyChair (2022)

[30] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition **47**(6) (2014)

[31] Jameson, D., Hurvich, L.M.: Theory of brightness and color contrast in human vision. Vision research **4**(1-2), 135–154 (1964)

[32] Jain, A., Bansal, R., Kumar, A., Singh, K.: A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. International Journal of Applied and Basic Medical Research **5**(2), 124 (2015)

[33] Robinson, C.W., Sloutsky, V.M.: Visual processing speed: Effects of auditory input on visual processing. Developmental science **10**(6) (2007)

[34] Lintern, G., Kennedy, R.S.: Video game as a covariate for carrier landing research. Perceptual and Motor Skills **58**(1), 167–172 (1984)

[35] Roenker, D.L., Cissell, G.M., Ball, K.K., Wadley, V.G., Edwards, J.D.: Speed-of-processing and driving simulator training result in improved driving performance. Human factors **45**(2), 218–233 (2003)