Evaluation of Communication and Human Response Latency for (Human) Teleoperation

David G. Black, Dragan Andjelic, Septimiu E. Salcudean, *Fellow, IEEE*

***Abstract*—We previously introduced a novel mixed reality teleguidance system dubbed human teleoperation [1], [2], in which a human (expert) leader and a human (novice) follower are tightly coupled through mixed reality and haptics. Our first evaluation of human teleoperation is in the context of tele ultrasound, in which a sonographer or radiologist's gestures are copied by a remote novice to carry out an ultrasound examination. In this paper, a communication system suitable for implementation of human teleoperation is presented and characterized in various network conditions, over Ethernet, Wi-Fi, 4G LTE, and 5G. To obtain a full understanding of latency in the system, the human response time is additionally characterized through a series of step response tests with 11 volunteers. The step responses were obtained by tracking the position of, and force exerted by, the human hand in response to a change in the mixed reality target. Different rendering methods were evaluated.**

The round-trip communication latency is 40 ± 10 ms over 5G, and down to 1 ± 0.6 ms over Ethernet for typical throughputs. The human response time to a step change in position depends on the step magnitude, but is between 485 to 535 ms, while the reaction time to a change in force is 150 to 200 ms. Both lag times are greatly decreased when tracking a smooth motion. Thus, we demonstrate that the system is network agnostic and can achieve good teleoperation performance and secure, low latency communication in appropriate network conditions. This brings the human teleoperation concept a step closer to human trials in a clinical environment, and the presented tools and concepts are applicable to any high-performance teleoperation system, for example for remote surgery.

***Index Terms*—Teleoperation, Human-Computer Interaction, Augmented Reality, Telemedicine, Communications**

I. Introduction

A. Background

MANY fields including telemedicine, manufacturing, and maintenance profit from remote guidance [3][4].

One particularly relevant procedure to which telehealth can be applied is ultrasound (US). This is useful not only for remote or under-resourced communities [5][6], but for Focused Assessment with Sonography in Trauma (FAST) examinations of trauma patients on ambulances [7], for elderly patients in care homes for whom mobility is difficult [8], for COVID-19 patients [9][10], and even for patients in hospitals when radiologists have to cover call in several hospitals at once. Remote training of sonographers is another popular application [11][12]. Point of Care Ultrasound (POCUS) is becoming increasingly more popular [13]. Existing approaches to tele-ultrasound include robotic teleoperation as well as multimedia applications that combine verbal and graphical guidance on a smartphone or tablet application.

Robotic US systems can provide high precision, low latency, and haptic feedback [14][15][16][17]. One system has demonstrated clinical utility in trials [18], and much recent work has focused on autonomous robotic US [19][20] using combinations of force-based positioning [21][22], depth camera-based trajectory planning [23], and reinforcement learning [24]. A good review of robotic ultrasound systems is found in [25], and more recently in [26]. Despite the large body of literature in this field, the issues of safe human-robot interaction and guaranteed robust autonomy remain difficult, especially from a regulatory perspective. Further limitations include restricted workspaces, time consuming set-up, large physical size that prevents use in ambulances, and cost, especially compared to inexpensive US systems. The questions of cost and complex setup and maintenance in particular make it difficult to deploy such systems in small communities where they are needed.

Conversely, systems sold by Clarius Mobile Health Corp., Butterfly Network, and Philips use a portable US probe with images and video conferencing available via a cloud interface on a mobile phone or tablet application. Though inexpensive and flexible, the desired probe pose and force are given verbally or with some overlays of arrows or pointers on the US image, which is very inefficient, leading to high latency and low precision. These systems are designed more for expert review of images captured by a capable sonographer rather than guidance of a novice.

Robotic teleoperation and video conference-based teleguidance fall on either end of a spectrum from performance to ease of use and deployment, leaving a large gap for solutions that are both flexible and easy to use and precise and efficient. In a previous paper [1], we introduced a novel concept of "Human Teleoperation" through mixed reality (MR) which bridges this gap. In this control framework, the human follower is controlled as a flexible, cognitive robot such that both the input and the actuation are carried out by people, but with near robot-like latency and precision. This allows teleguidance that is more precise, intuitive, and low latency than verbal guidance, yet more flexible, inexpensive, and accessible than robotic teleoperation.

While virtual reality immerses a user in a virtual environment, augmented reality takes the real environment and augments it by adding visual information to it [27]. This augmentation can take place within the real environment itself

David G. Black, Dragan Andjelic, and Septimiu Salcudean are with the Department of Electrical and Computer Engineering, University of British Columbia, BC, Canada e-mail: dgblack@ece.ubc.ca, dragan.andjelic@ubc.ca, tims@ece.ubc.ca.

Support for this Research was provided by NSERC, CFI, MITACS, Rogers, and the C.A. Laszlo Chair held by Prof. Salcudean.

Manuscript received October, 2022.

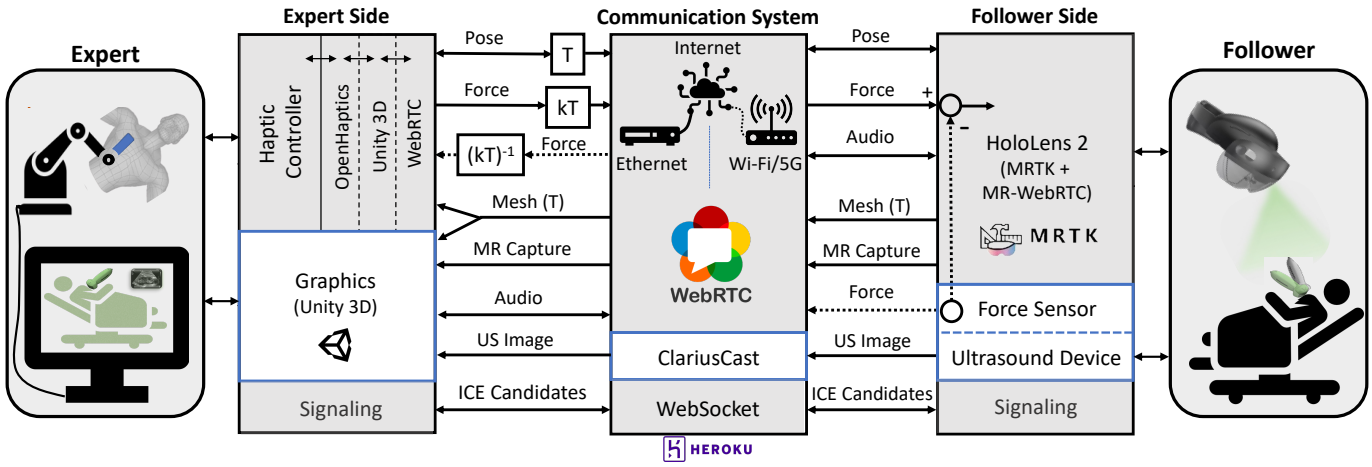


Fig. 1: System Architecture. k is a scaling factor for the force while T is the transform from expert to follower coordinates, obtained from the mesh. The force feedback (dotted lines) has not yet been implemented.

or on a video stream or similar. Mixed reality is a subset of augmented reality in which the visual cues are overlaid onto the real, physical environment using overlays in an optically transparent headset such as the Microsoft HoloLens 2, MagicLeap 2, Nreal Light, and the newly announced Meta Quest Pro [28]. It should be noted that Milgram and Kishino [29] described a "reality-virtuality continuum" in which mixed reality was a superset of augmented reality and augmented virtuality, but this notation is now largely out-of-date after 28 years. The classification was revisited recently in [30], but this does not differentiate between information overlaid onto recorded images or videos and information projected into the real world. Thus, we keep the MR nomenclature described above. Indeed, the ability to project 3D information seamlessly into the real environment is the primary enabling technology for human teleoperation, as this information can be used to guide a novice follower while he/she interacts with the environment, for example for an ultrasound exam.

Human trials were carried out in [31] to investigate this ability of humans to act as a robot in tracking an input MR signal, showing promising performance. However, these tests were performed with the expert and follower sides connected over WiFi, on a fast network. Introducing communication latency can have a strong negative effect on teleoperation performance. Li et al. found that performance decreases above 150 ms delay for haptic tasks in teleoperation [32], while Jay and Hubbold determined that delays of 69 ms in visual feedback and 187 ms in haptic feedback are disruptive to a user manipulating a haptic device [33]. The same group later found that delays of 25ms affected specific haptic tasks for collaborative virtual environments [34], but that the effect of delays is highly task-dependent. Though users seem to be more sensitive to delays in visual rather than haptic feedback, when a haptic delay is perceived, performance drops much faster than if the delay is visual [32].

Thus, latency is key for almost all aspects of the teleoperation. The sonographer relies on visual feedback from the ultrasound images, the haptic feedback, and the video stream from the mixed reality headset to decide where to move and what force to apply. Delays in any of the data lead to a very unintuitive experience. Masuda et al. achieved telerobotic ultrasound latency of < 1 second [35], compared to previous experiments where they had 4-5 seconds of latency [36], which they described as "very stressful" [35]. However, even delays of 1 second in force control can cause instability. Niemeyer and Slotine proposed the use of wave variables to maintain stability for time-delayed force reflecting teleoperation [37], which have since been improved for time-varying delays [38], using disturbance observers [39], time domain passivity control [40], μ-synthesis [41], and more.

Given the profound effects of time delays on performance, stability, and controller design, it is important to minimize and then measure and characterize these delays in any system. In this paper, we present a communication system which uses a secure, high-speed, network-agnostic Web Real Time Communication (WebRTC) interface, described in Section II-A. Section II-D describes a number of tests that were performed on the communication system to characterize its performance in different network conditions, including latency tests over Ethernet, WiFi, 4G LTE, and 5G, all with various signal conditions. To our knowledge, no other remote ultrasound system in the literature presents detailed communication system design or thorough characterization thereof in different network conditions that it may practically be exposed to. These tests are equally applicable to any remote teleoperation system.

Since human teleoperation is a human-in-the-loop system, however, the delays associated with the human response time are also critical. These are evaluated through step response tests with 11 subjects. The tests are significant to show that the perceptual and cognitive delay in the human subjects is within the range observed in prior work (cited above) to allow successful human teleoperation. The measured values provide a starting point for control system design, to optimize the human teleoperation system response. To the best of our knowledge, no other study has explored the human response when guided by an MR interface. To perform these tests, the prototype developed in [1] was improved to take direct force input from the expert, and a dummy ultrasound probe

TABLE I: Uplink and downlink throughput on the follower side, which is mobile and thus more bandwidth limited

Type	Uplink	Note	Downlink	Note
Timing	1.28 Kbps	1 8bit long × 20 Hz	3.84 Kbps	3 8bit longs × 20 Hz
Force	16 Kbps	3 32bit floats + 1 64bit timestamp × 100 Hz	16 Kbps	Same as uplink
Pose	28.8 Kbps	7 32bit floats + 1 64bit timestamp × 100 Hz	28.8 Kbps	Same as uplink
Video	≈ 1-2 Mbps	960 × 540 H.264 encoding, 25 Hz, Variable quality	0	No downlink video
Audio	128 Kbps	Typical MP3 bitrate (Part of MPEG-4 stream)	128 Kbps	Same as uplink
US	4.64 Mbps	58KB JPG image × 10 Hz	0	No downlink US
Mesh	2.3 MB	≈12k mesh triangles × 3 points and 3 indices × 32bit floats	0	No downlink mesh
Total	6.81 Mbps	Mesh sent rarely on demand. Peak throughput 9.43 Mbps	180 Kbps	Sum

was designed for the follower, including 6 axis force/torque sensing and 6-DOF position and orientation (pose) tracking (Section II-B). A novel visual control system for the forces was developed, as described in Section II-C. Further rigorous tests of the visual force control and human tracking ability are presented in [31].

B. Human Teleoperation

The human teleoperation system consists of the follower side and the expert side, which communicate over the Internet. The follower, who need not have any US experience, wears an MR headset (Microsoft HoloLens 2) which projects a virtual US transducer into the follower's scene. The expert sonographer controls the virtual probe in real time using a haptic controller (Touch X, 3D Systems, Inc.) to input the desired pose and force. The follower tracks the virtual probe with his/her real probe. The expert, in real time, receives the US images, a video stream of the patient with the virtual and real probes in position (called an MR capture), and is in verbal communication with the follower. Additionally, the follower sends a spatial mesh of the patient, generated by the HoloLens 2, to the expert. The mesh is rendered haptically as a virtual fixture for the haptic device, giving the expert the sensation that they are physically interacting with the tissue. This spatial mapping also provides the expert-to-follower coordinate transform. The system is shown in Fig. 1.

The effectiveness of the approach was demonstrated first using a WebSocket server and Robot Operating System (ROS) on a local wireless network (WLAN), showing large improvement in accuracy and completion time compared to existing teleguidance methods [1]. In the following sections, we describe and characterize a much improved communication system for human teleoperation, testing it in different networks, signal conditions, and data throughputs. An instrumented dummy ultrasound probe is introduced to enable these tests and the development of force control methods, and finally the human-in-the-loop latency is investigated as well.

II. Methods

A. Communication

The US images, video feed, and spatial meshes require a large bandwidth while haptic feedback and MR teleoperation necessitate very low latencies for stable, transparent, and intuitive teleoperation. Accounting of required throughputs is shown in Table I. The follower side is very biased towards uplink although available uplink bandwidth is usually smaller than downlink. Thus, bandwidth is particularly important in this system.

A WebRTC-based system is more suitable to meet these requirements and support tele-US at large distances. This framework provides a direct peer-to-peer connection between the expert and follower, thus removing server-related delays. Data is sent either over data channels (general data) or media channels (video encoded streams), which are built upon Stream Control Transmission Protocol (SCTP) [42] and Realtime Transport Protocol (RTP) [43] respectively. RTP is built on top of UDP (User Datagram Protocol) but highly optimized for real time video communication while SCTP can act much like UDP with some improved features. As with UDP, dropped packets can be ignored for maximum performance. In this case there is no guarantee that packets sent in a specific order will arrive in the same order. However, a second configuration is available which guarantees chronological ordering while not retransmitting dropped packets. Finally, full acknowledgement and retransmission can be configured as well, leading to more TCP-like behaviour. These settings are tested in Section II-D. Generally, however, given the high-performance application, the higher speeds of UDP are preferable to the reliability of TCP. Dropped packets are quickly replaced with new information, and local consistency checks are in place.

A further benefit of WebRTC is that it uses several existing sub-protocols such as Session Description Protocol (SDP) and Interactive Connectivity Establishment (ICE) to establish an optimal connection between two peers over any network and through any router NAT (Network Address Translation) scheme or firewall. This is achieved by having both peers connect to a signaling server and exchange SDP information. Based on this information, they can automatically discover an efficient route through the different network hops between the peers. Once connectivity is established, all data is sent directly peer to peer, and the signaling server is no longer needed. In addition, WebRTC uses Datagram Transport Layer Security (DTLS) to ensure the connection is authenticated and encrypted. Thus, all transported information is secure, making it ideal for this medical application.

We have implemented WebRTC-based communication for the human teleoperation system, using separate data and media channels for each of the rows in Table I, in addition to two control channels that exchange occasional commands. The signaling server is implemented in Python and runs on a password-protected web server hosted on Heroku, a cloud platform. All SDP data is securely encrypted before being sent to the server, and is decrypted by the other peer. The

flexibility of WebRTC allows the system to work without any modification over Ethernet, Wi-Fi including enterprise networks in universities or hospitals, 4G LTE, or 5G.

In collaboration with Rogers Communications, we have set up an antenna which connects to mobile networks and to a Wi-Fi router, which in turn connects to the HoloLens via Wi-Fi or to a PC via Ethernet, thus allowing the HoloLens or PC to communicate over the mobile network. A diagram showing the setup is in Fig. 4. The University of British Columbia was the first campus equipped with a non-standalone (NSA) sub-6GHz 5G network in North America by Rogers, allowing the system to be tested over 4G and 5G. The 5G network in particular holds promise for achieving the required bandwidth and latency, and provides additional features such as multi-access edge computing (MEC), allowing costly computations to be outsourced at very low latency to a server at the base station. Furthermore, 5G can utilize a mm-wave band, leading to vastly improved latencies and throughputs. Testing the benefits of both MEC and mm-wave will constitute future work.

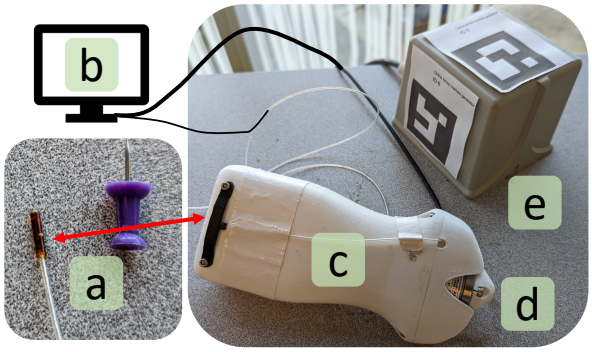


Fig. 2: Instrumented dummy US probe (c) for tests, including pose sensing (a) and force sensing at the tip (d). The pose sensor is shown next to a thumb tack for scale. Both sensors connect to a PC (b), and the electromagnetic transmitter (e) has ArUco markers for registration.

B. Instrumented Test Probe

In order to complete the teleoperation system, force and pose feedback are required from the real ultrasound probe. The measured force is compared to the desired one in order to generate the visual force indicator for the follower to track. Similarly, the measured pose can be compared to the desired one to produce a feedback signal, or it can be used in conjunction with the measured force to estimate the mechanical tissue impedance to feed back to the expert's haptic device. In [31] and Section III-C, the measurements are used to characterize human performance in the system.

To implement pose sensing, several options were explored. An inertial measurement unit (IMU) can provide accelerometer and gyroscope readings which give a good orientation estimate but are subject to large drift and not feasible for position tracking. Optical tracking using an NDI Polaris or similar device is fast and accurate and was tested with our system. However, it loses tracking when the reflective markers are occluded, which happens often during an ultrasound exam. Initial work on a similar infrared-marker-based optical tracking system using the HoloLens IR sensor was carried out in [44]. However, this suffers from some of the same occlusion problems, and it was found that the HoloLens 2 tracking was only accurate to about 3-4 mm and had a relatively low update rate which was not sufficient for this application. Sensor fusion with optical tracking and IMU data has also been explored [45], but adds complexity. We instead utilized an electromagnetic tracking system (NDI driveBAY) which does not rely on line-of-sight and is accurate to about 1.4 mm and $0.5°$. With a readout rate of up to 420 Hz and very small size, it is ideal for this application.

The electromagnetic sensor includes a small sensing element (Fig. 2 a), and a transmitter (Fig. 2 e), which also defines the sensing coordinate frame. ArUco markers [46] are included in known positions on the transmitter, allowing the HoloLens to accurately determine its pose in the HoloLens frame, thus providing the transform from measured force coordinates to desired force coordinates.

For force sensing, an ATI Nano25 6-axis force/torque sensor was used for its high precision (0.02-0.06 N), reliability, and small size. This can be installed between a 3D printed shell and the ultrasound probe as done in [14][47][48][49][50]. For the tests presented here and in [31], it was instead installed at the tip of a 3D printed dummy ultrasound probe to ensure best possible accuracy. The instrumented dummy ultrasound probe is shown in Fig. 2.

Both sensors are connected to a PC, referred to as the sensor PC, which communicates the readings to the HoloLens via WebRTC, over the local WiFi.

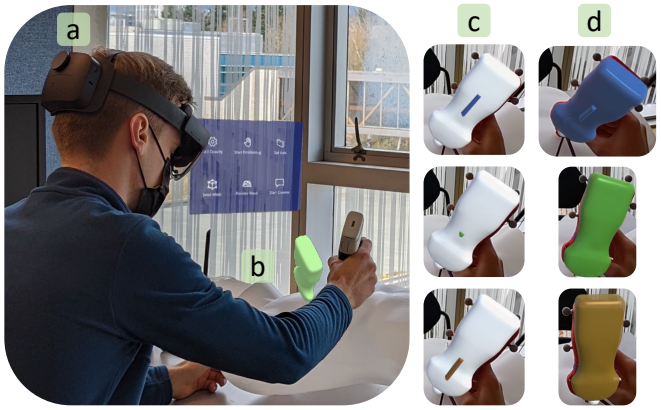


Fig. 3: Follower side showing the follower wearing a HoloLens 2 (a), the holographic user interface (b), and the different force rendering schemes: Error-bar (c), and Color (d)

C. Mixed Reality for Pose and Force Tracking

The primary premise of human teleoperation is efficient tracking of pose and force using MR overlays. The speed and accuracy for tracking step changes is presented in Section III-C. For these tests, a virtual ultrasound probe was projected into the follower's field of view, as shown in Fig. 3. The

follower's goal is to align his/her probe as well as possible with the virtual one, thus matching the desired pose. In some lighting conditions, the virtual probe can occlude the real one, leading to increased position error. Thus, the effectiveness of a full probe rendering was compared to a scheme in which the central part of the virtual probe was removed in [31]. Additionally, the full probe's opacity can be adjusted. For the step response tests in this study, a full probe rendering was used.

Forces are also an important part of US imaging, determining which structures are visible, and ensuring they are not too deformed. To achieve force tracking, several visual force rendering methods were developed and tested in [31]. The two most promising schemes are shown in Fig. 3. First, the expert applies their desired force to the haptic device. Forces in US are typically between 0-20 N [48], but the haptic device is limited to 8 N. Thus, the forces are scaled down, which has the added benefit of decreasing the load on sonographers, who are known to suffer from increased incidence of musculoskeletal injury [51]. The follower's measured forces are then compared to the desired ones to generate an error signal. The virtual US probe then either changes color continuously between blue, green, and red (Fig. 3 d) or an error-bar grows continuously towards or away from the patient and changes color (Fig. 3 c) to indicate too little force, good force, or too much force respectively. The error-bar approach in particular is shown to be very effective [31], and is used in the step response tests.

In most teleoperation tasks involving contact, forces and positions are controlled in orthogonal subspaces; i.e. forces are controlled normal to the surface being contacted, and positions are controlled in the two tangent directions [52]. This applies very well to ultrasound procedures as well, so the step response tests were performed on a flat, rigid surface with forces normal to the surface and motions tangent to it.

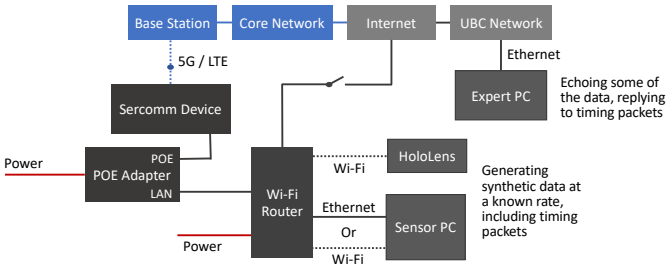


Fig. 4: Test setup for testing the communication system. POE = Power Over Ethernet. The WiFi router can either be connected to the mobile network or directly to the Internet via a wired connection. The HoloLens and sensor PC can similarly connect to the router via WiFi or Ethernet.

D. Communication System Latency Tests

While the human-computer interaction and human tracking performance are characterized in [31], and the force feedback and practical use in a clinical environment will be evaluated carefully in future work, this paper focuses on latency, both in the communication system and in the human response time. We performed a number of tests to determine system performance over different networks and in various conditions.

To perform these tests, the human teleoperation system was modified to send synthetic data of a specific size, generated randomly and sent at a set rate. Thus, the throughput could be adjusted. A diagram of the test setup is shown in Fig. 4. The data was communicated constantly for 4 minutes during each test. A separate data channel with timing packets was set up to measure the latency. In fact, the round trip time (RTT) was measured instead of latency because clock drift between two devices can easily be of the same scale as the communication latency, making direct measurement of latency impractical. Instead, a specific procedure was devised to cancel out the clock drift, as follows.

Every 50 ms, a microsecond-resolution, 64-bit timestamp, t_1, is measured on the follower side and sent to the expert side. Immediately upon receipt of the message, the expert side measures its own timestamp, t_2, appends it to the packet, and prepares to send it back. Directly before sending, another timestamp, t_3, is measured and appended. When the follower receives the response, it immediately measures a fourth timestamp, t_4. The RTT can then be calculated as

$$RTT = (t_2 - t_1) + (t_4 - t_3) \tag{1}$$

If we consider a clock drift of δt, and denote times in the follower clock with a prime (eg. t'), then in the follower clock, $t'_2 = t_2 + \delta t$ and $t'_3 = t_3 + \delta t$. The RTT from (1) then becomes

$$\begin{aligned} RTT &= [(t'_2 - \delta t) - t'_1] + [t'_4 - (t'_3 - \delta t)] \\ &= [t'_2 - t'_1] + [t'_4 - t'_3] + \delta t - \delta t \\ &= [t'_2 - t'_1] + [t'_4 - t'_3] \end{aligned}$$

Thus, the clock drift is effectively canceled out. For an approximate latency figure, one can take $RTT/2$. Since most networks are faster for downlink than uplink, however, this is not necessarily a good approximation, so we use RTT for the remainder of the paper.

Different amounts of data were sent over different WebRTC channels to simulate the data in the teleoperation. In total, 9 different throughputs were tested, each in 7 different network conditions. These are outlined in Table II and III respectively. During testing, SINR values varied randomly by a few points. The different conditions were achieved by testing in different locations. The expert side was stationary in one building, while the follower side was moved to a lab two buildings down for some of the tests.

No test was performed in poor 5G signal because the network automatically switched to 4G in this case. Indeed, in an NSA network, as was available for our tests, a given user equipment (UE) device connects to the nearest base station via a relatively static LTE primary carrier, but may dynamically send data over 4G or 5G, depending on the UE's current throughput and latency requirements. It is not known or controllable by the user which network the data is ultimately sent over. However, in our testing, data was sent at a high rate and throughput and the 5G carrier had a far higher signal to interference plus noise ratio (SINR) than the LTE, so it can safely be assumed that the data was sent over 5G. In the 5G

TABLE II: Latency tests were performed with these throughputs. The last 5 rows test possible sizes of US stream, where 6.69 Mbps constitutes sending the US with just jpeg compression.

Throughput	Description
1.28 Kbps	Just timing packets
46.08 Kbps	+ Force and Pose
1.17 Mbps	+ Video and Audio (lower quality)
2.17 Mbps	+ Video and Audio (higher quality)
2.57 Mbps	+ US (5KB)
2.97 Mbps	+ US (10KB)
4.17 Mbps	+ US (25KB)
4.97 Mbps	+ US (35KB)
6.81 Mbps	+ US (58KB)

TABLE III: Latency tests were performed in these network conditions to simulate conditions that would be encountered in the field. SINR = Signal to Interference plus Noise Ratio, RSRP = Reference Signal Received Power, RSSI = Received Signal Strength Indicator.

Network	SINR (dB)	RSRP (dB)	RSSI (dB)	Short Name
Ethernet	-	-	-	Ethernet
WiFi	-	-	-	WiFi
5G NR	28	-77	-49	NR Good
5G NR	14	-87	-65	NR Mid
4G LTE	27	-63	-33	LTE Good
4G LTE	19	-94	-66	LTE Mid
4G LTE	4	-89	-52	LTE Poor

tests, the LTE SINR was around 2-3dB. To test 4G latency, the antenna was configured not to connect to 5G. Note, it was not possible to configure the antenna to connect only to 5G as no Stand Alone (SA) 5G network was available.

The Ethernet and WiFi tests in Table III refer to the follower being connected directly to the Internet via Ethernet or WiFi. The expert PC is always connected via Ethernet. Fig. 4 shows the path taken by data between the expert and follower. Notice that when communicating over 4G or 5G with the HoloLens, there is first a hop over WiFi; i.e. the HoloLens connects to the RF antenna via WiFi. To establish the delay associated with this hop, an equivalent C# program was written for the sensor PC, which was attached to the antenna via Ethernet. All mobile network tests in Table III were carried out over Ethernet, and then an additional set of tests was performed to determine the added latency from the WiFi hop.

When data is sent over a data channel, it is first added to the channel's send queue, which tends to fill if there is network congestion, leading to packet delays. We therefore experimented with splitting the US channel which had a large throughput into 2 smaller channels. This test was repeated at 2.17 Mbps and 4.17 Mbps in medium 5G conditions (Table III) to determine the effect.

Furthermore, as mentioned in Section II-A, the underlying transmission protocol can be configured as reliable (TCP-like retransmission of dropped packets), ordered (packets guaranteed to arrive in order), or none. We performed tests at 46.1Kbps and 2.17 Mbps throughput in medium 5G conditions (Table III) using each mode to determine the effect on latency.

E. Human Response Tests

Finally, initial results in [1] showed that the system's latency was limited not by the communication latency, but rather by the reaction time of the follower. To quantify this carefully, a series of step response tests were carried out for force and position tracking of the follower. Using the experimental setup described in [31], $n = 11$ healthy volunteers aged 20-64 (mean age 32) were asked to track the virtual US probe with the instrumented probe from Section II-B. The step response consisted simply of a series of step changes in pose or force. The directions of the position jumps and interval between steps were randomized to avoid the subject learning and anticipating where or when the next step would occur. Each input signal was generated on the expert side and sent via the described communication system to the follower, where it was rendered by the MR headset and tracked by the follower. For force, the error-bar rendering was used (Fig. 3), and the follower held the dummy probe against a rigid table. The desired forces were normal to the surface. All desired and measured positions and forces were logged on the HoloLens with timestamps to allow precise comparison.

All of the step response tests were performed at a constant amplitude of 10 cm and 11 N for position and force respectively. However, we expect the response to be slightly different at different amplitudes. This was tested in a few subjects by having them perform the step response tests at four different input amplitudes: 2.5, 5, 10, and 15 cm, and 3, 6, 12, and 18 N.

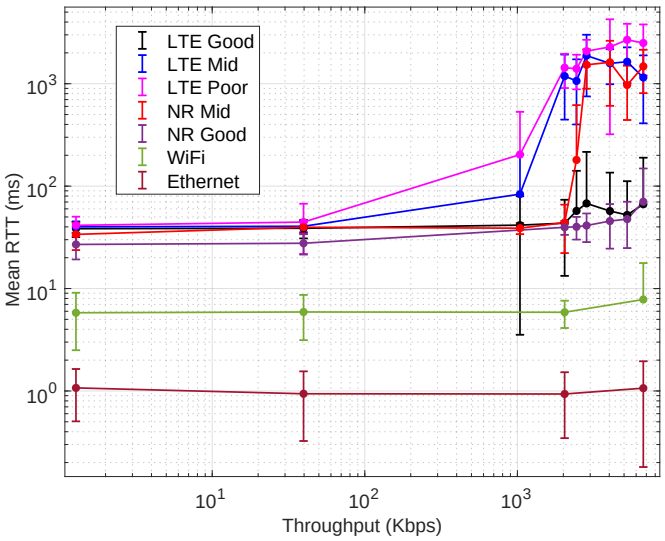
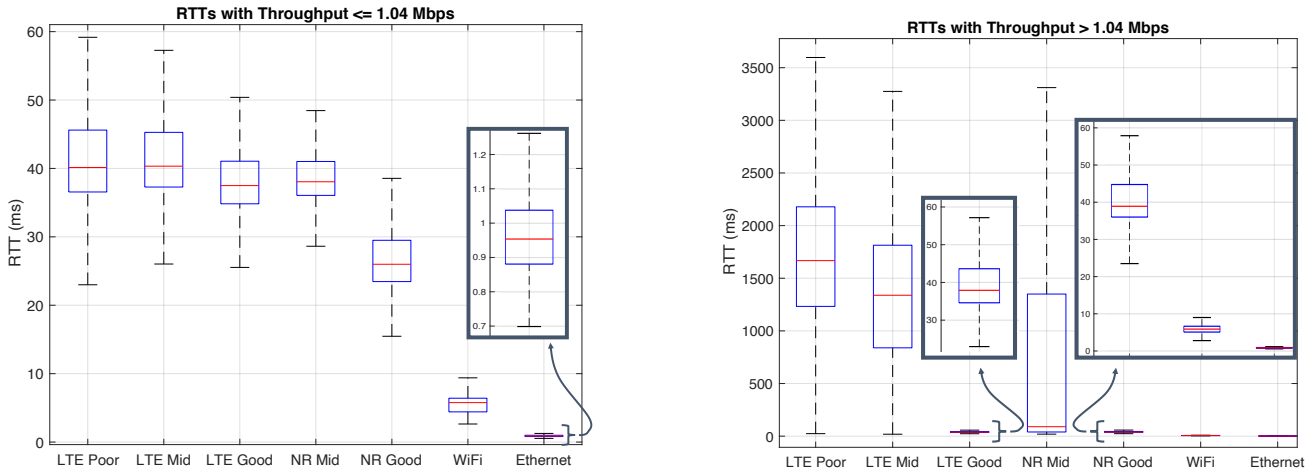


Fig. 5: RTT versus throughput in different signal conditions

III. Results

A. Communication System Throughput and Latency

The results of the tests in good network conditions are found in Table IV, showing the difference between Ethernet, WiFi, 4G, and 5G. With Ethernet, it would even be possible to maintain a 1 kHz control loop for bilateral teleoperation with force feedback. The WiFi link adds about 4-6 ms delay



(a) $RTT \leq 1.04$ Mbps (b) $RTT > 1.04$ Mbps

Fig. 6: RTT with large and small throughput in different network conditions

TABLE IV: RTT (ms) versus throughput in good signal conditions for different networks.

	Ethernet	WiFi	4G LTE	5G
1.28 Kbps	1.07 ± 0.57	5.80 ± 3.30	38.41 ± 6.63	26.95 ± 7.72
46.08 Kbps	0.94 ± 0.61	5.90 ± 2.77	38.77 ± 7.92	27.67 ± 6.21
1.17 Mbps	-	-	41.6 ± 38.06	-
2.17 Mbps	0.93 ± 0.59	5.87 ± 1.75	43.49 ± 30.18	39.61 ± 6.14
2.57 Mbps	-	-	57.30 ± 83.88	40.09 ± 10.02
2.97 Mbps	-	-	67.78 ± 148.30	41.23 ± 12.85
4.17 Mbps	-	-	57.13 ± 78.41	45.61 ± 21.07
4.97 Mbps	-	-	52.30 ± 59.30	47.64 ± 22.81
6.81 Mbps	1.07 ± 0.88	7.82 ± 9.90	66.58 ± 123.00	70.44 ± 78.29

on top of the Ethernet but is still very fast. The mobile networks are slower but still fast, with 5G being about 5-10ms faster than 4G. Both become significantly slower as throughput grows very large, but remain for the most part below the thresholds for human notice in haptic or visual feedback cited in Section I-A. Further tests in medium to poor network conditions are shown in Fig. 5.

In these tests the network demonstrates similar behaviour to good conditions for low throughputs. However, at about 1 Mbps the RTT makes a sudden, large jump of more than an order of magnitude, then stays relatively constant. In medium 5G, the jump occurs later, at about 2 Mbps. Fig. 6 compares the different networks and network conditions for the two distinct cases, before and after the jump in RTT. We see that at low throughputs, good LTE and medium NR have similar performance, but that the medium 5G has a long tail towards large RTTs as throughput increases, though the median remains relatively low. Good 5G has a significantly better performance at all throughputs than LTE ($p < 0.001$), and within LTE, variance decreases with improving signal condition.

TABLE V: Effect of splitting high-volume channel into two

Throughput	Split RTT	Non-split RTT	p-value
2.17 Mbps	2491 ms	2478 ms	0.58
4.17 Mbps	2763 ms	2609 ms	< 0.001

TABLE VI: Effect of packet reliability on RTT. All pairings are significantly different with $p < 0.001$ except None and Ordered for low throughput.

Throughput	None	Ordered	Reliable
46.1 Kbps	60.66 ± 69.91	55.34 ± 20.84	67.53 ± 77.29
2.17 Mbps	2478.3 ± 1199.4	2497.7 ± 4191.5	2634.6 ± 1809.0

B. *Queuing and Reliability*

The large latencies are in part due to long queueing delays in the send buffers as network congestion increases. A test was performed to determine if splitting the US channel into two would improve performance by adding a second send buffer. The results are shown in Table V. Interestingly, splitting leads to significantly worse performance, and this does not even take into account the overhead of synchronizing packets and recombining the image on the expert side.

Finally, the effect of packet reliability was tested and is shown in Table VI and Fig. 7. At small throughputs there is relatively little difference, though reliable packets are still significantly slower than the other two. At large throughputs, however, there is a marked difference. As seen in Fig. 7, the reliable mode leads to massive delays which are unacceptable in this system. Conversely, ordered and unordered means remain very similar, though with ordered packets the median is lower and the variance is much larger, as seen by the relatively large number of outliers. Thus, as hypothesized in II-A, a UDP-like, communication without retransmission is best for this system where low latency and high data rate are key.

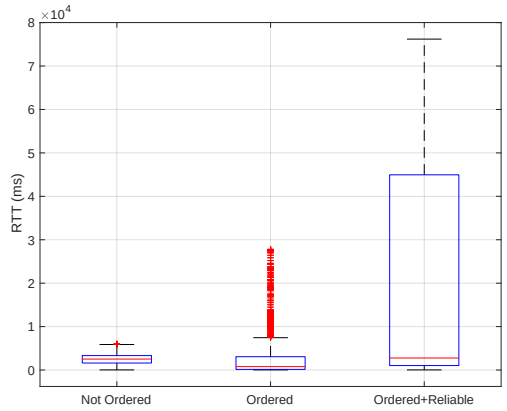


Fig. 7: Effect of packet reliability on RTT for 2.17 Mbps throughput in 5G with medium signal quality.

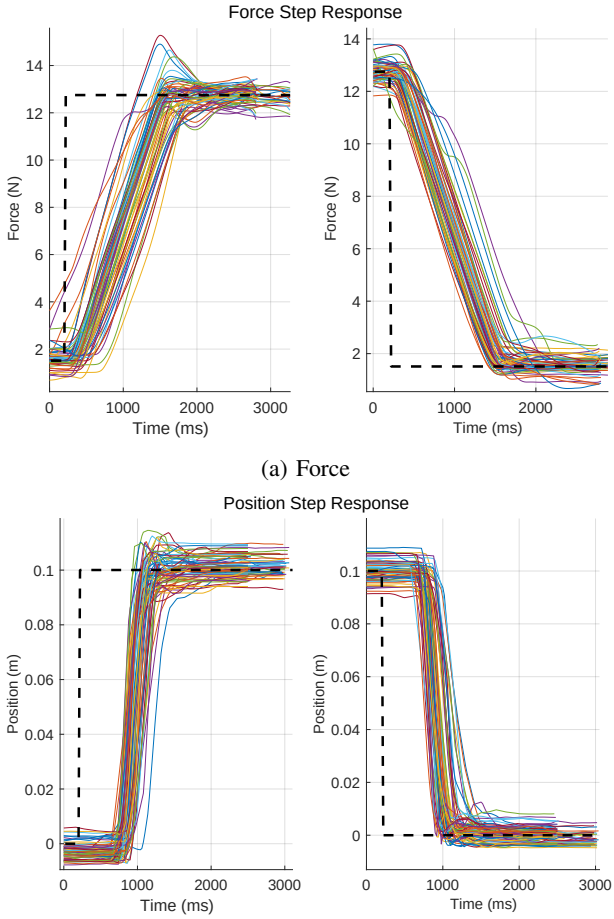
C. Human Step Response

In total, 418 step responses, plotted in Fig. 8, were measured and analyzed. We can define reaction time (RT) as the time delay between the initiation of the desired step and when the user starts moving, specifically when the second derivative is maximum. We define rise time to be the time taken between starting the motion and finishing it. The reaction times, rise times, and steady-state errors are listed in Table VII.

The RTs in the down steps are faster on average than on the up steps (Table VII). In the tests, the probe was moved from a central position to a point 10 cm away in a random direction, then always back to the central position. Similarly, the force always returned to the same low value. Despite differences in the interval between steps, returning to a more familiar position or force decreased the reaction time significantly ($p = 0.047$ for position, $p < 0.001$ for force). This implies that part of the reaction time involves processing in which direction to move, or how hard to press.

On the steps up in particular, the fastest responses also have larger overshoot and more oscillation. Clearly, these users adopted more aggressive, higher-gain controllers. All such users were young (< 25). Indeed, there is a positive correlation between age and RT in the step responses, with correlation coefficient 0.5 ($p < 0.001$). As expected, older participants reacted more slowly. However, there was no similar correlation between age and rise time (correlation 0.06), so the limitation appears to be cognitive, not physical.

The force RTs were much faster than the position RTs ($p < 0.001$), likely because no motion was required, only a change in force. On the contrary, rise times were much slower for forces ($p < 0.001$), because the followers had to rely entirely on the visual feedback rather than having an intuitive feel for whether they had achieved the desired value, unlike for the position tracking. This is discussed more in Section IV.



(a) Force

(b) Position

Fig. 8: Step Responses. Dotted line is input signal.

The step responses were repeated in a few subjects at four different input amplitudes, to determine the effect. At larger magnitudes, the rise time is of course larger, since the person has to move further (correlation 0.98, $p = 0.018$). However, the RT is also slower (correlation 0.93, $p = 0.07$). These are shown in Table VIII and agree well with Fitt's Law, which describes a relation between reaction time and motion amplitude [53]. The same trend is not present in the force step response, where no physical motion was required.

TABLE VIII: Reaction time versus step amplitude

Magnitude (cm)	Reaction Time (ms)	Rise Time (ms)
2.5	485.12 ± 14.14	200.88 ± 14.15
5	465.42 ± 14.12	250.76 ± 56.58
10	525.00 ± 42.43	270.01 ± 57.65
15	565 ± 127.38	280.36 ± 212.30

TABLE VII: Step response results

	Reaction Time (ms)	Rise Time (ms)	Steady State Error
Force Up	201.09 ± 95.67	1101.23 ± 361.82	2.8 ± 2.1 mm
Force Down	141.91 ± 76.13	1047.88 ± 361.54	
Position Up	535.00 ± 96.32	199.06 ± 104.88	0.26 ± 0.19 N
Position Down	521.57 ± 108.17	186.70 ± 107.66	

IV. Discussion and Conclusions

In the original system described in [1], latency tests on the ROS WebSocket implementation showed on average 11.4 ms delay for pose and force transmission over a local network. In contrast, the new system has delays of 1.07 ± 0.57 ms over Ethernet or 5.80 ± 3.30 ms over WiFi on average for the same local network. In addition, the latency of the audio/video data over WebRTC is about the same whereas with Windows Device Portal it was found to be ≥ 4 seconds, and with ROS it was infeasibly slow. Furthermore, the new system can run remotely over the Internet, is secure, and works on mobile networks as well. It achieves RTTs of $38 - 67$ ms over 4G LTE and $27 - 70$ ms over 5G, depending on data throughput. The presented system is thus a major improvement over the original prototype.

From the results we can conclude that the communication should be run without retransmission, and likely without ordering guarantees. High-volume data channels should not be split, although the US images require further consideration (see below). For optimal performance WiFi should be used when possible, or even over Ethernet via a USB-C adapter on the HoloLens 2. WiFi adds 4-6 ms latency over the 1 ms latency achieved by Ethernet. Both 5G and 4G offer high performance as well when needed, though more care is required. In worse network connectivity, some parts of the system, for example the video conference, may have to be turned off, and the video and US quality should be adjusted dynamically. This is already the case for the video stream, but it needs to be implemented for the US. Only in poor LTE connectivity with SINR < 4dB or so is the teleoperation with transmission of reasonable quality US images not feasible.

In all cases, for good network conditions the teleoperation latency is strongly dominated by the human response time, which is between 150-550 ms. For poor network conditions and large throughputs, however, this relation can reverse, which leads to very unintuitive teleoperation. This condition should be avoided. The step response RTs match well with previously proposed values for human visual system RT. Badau et al. describe three different RTs - simple RT, recognition RT, and cognitive RT - which have significantly different values [54]. Simple RT is for tasks where the subject sees an indicator and pushes a button, whereas in recognition RT the user has to recognize a specific object among a collection of shapes and locate and click on the object. This explains the difference between the force and pose RTs in Section III-C. Force is a simple RT: the user sees the error-bar change and pushes down, which happens very fast. On the other hand, pose involves a recognition RT and is thus slower: the user has to recognize which direction the probe moved in, and follow it. In this way, much of the processing of where to move occurs before initiating the motion for pose tracking, while for force tracking deciding how hard to press occurs during the motion. Hence, the rise time for pose is much faster than force.

It was also found that younger users were faster and in some cases adopted a more aggressive controller with overshoot. This precisely mirrors what is found in [55] and [56]. Finally, Carlton argues that the reaction time approach to studying processing delays is not appropriate when visual information constitutes feedback from continuous motion [53]. This suggests that better performance can be expected during teleoperation when motions are relatively smooth and continuous, as opposed to the large steps shown here. Indeed, our results regarding tracking delays for continuous motions and force sequences in [31] are much faster than the discrete reaction times from the step responses (Table VII), and are more realistic representations of an ultrasound exam. Nonetheless, the step response tests presented here represent a worst-case response time, which is important to know. Furthermore, the response time and accuracy is dependent on the rendering method used to show the desired position and force. In [31] we tested four different rendering schemes and present here step response results using only the best two. However, better schemes likely exist, so the results presented here constitute a baseline.

Although the transmission protocol decided upon in the above tests achieves its performance by ignoring dropped packets, there is still retransmission at a lower level. The mobile network itself can run in acknowledge or non-acknowledge mode, in which dropped or corrupt packets are retransmitted or not, respectively. The Rogers network used in the tests runs using a default "best effort" quality of service (QoS), which includes acknowledge mode. As it is a public network, we were unable to change this or test its effect. Furthermore, again since it is a public network, the tests were subject to the amount of traffic currently loading the network from students, faculty, and staff on the university campus. For this reason, all tests were performed early in the morning when few students were present. However, configuring the network to treat packets from this critical medical application with a different QoS - i.e. with higher priority and without acknowledge mode - would further increase performance and reliability. Though performance was already sufficient for human teleoperation, and thus no special QoS configuration is required, the performance of haptic feedback could likely benefit.

A limitation of this study is that all the network tests were carried out on a single network, which is subject to certain configurations as explained above. Different networks in different locations will lead to slightly different performance. Similarly, the expert PC was connected to the Internet via an institutional enterprise network, which likely adds some latency. Further, the human study was limited. Though the volunteers represented a mix of sexes and ages, further tests should be performed on a larger, more diverse population of novices and a specialized population of sonographers. While the tests presented here aimed to ascertain human performance limitations, specific performance tests should be carried out for ultrasound, using realistic motion ranges from sonographers and radiologists for standardized exams [57].

Currently, the US images are streamed with jpeg compression from the Clarius C3HD3 device to the sensor PC using the ClariusCast API. From here, they are forwarded to the expert and follower via WebRTC. However, as seen in the results, the large amount of throughput required for this can seriously affect the communication latency. Sending individual

jpeg images is highly inefficient, especially considering that the US image does not change much from frame to frame. Thus, future work will investigate sending difference images between frames, with video encoding such as H.264 or H.265 and variable quality depending on the connection. This will dramatically reduce the required throughput.

Future work will also involve performing human trials with patients in the community and expert sonographers at Vancouver General Hospital to establish the practicality of the system. We are also developing miniaturized force sensing transducers which can be integrated in a low profile shell on an US probe to provide force feedback without disrupting the ultrasound imaging. Using the measured forces we can study stable and transparent force reflection in bilateral teleoperation under time delays imposed by the human response time. Furthermore, the human-computer interface can be optimized, and reinforcement learning for autonomous US guidance can be explored. This constitutes an exciting avenue for autonomy since there is no possibility of dangerous or unpredictable robot actions as the AI would control only the virtual probe.

ACKNOWLEDGMENT

We gratefully acknowledge scholarship support from the Vanier Canada Graduate Scholarships program, infrastructure support from CFI and funding support from NSERC and the Charles Laszlo Chair in Biomedical Engineering, as well as infrastructure, technical, and funding support for the communication system from Rogers Communications and MITACS.

REFERENCES

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

[2] D. Black and S. Salcudean, "A mixed reality system for human teleoperation in tele-ultrasound." Hamlyn Symposium for Medical Robotics, July 2022, pp. 91–92.

[3] S. N. Gajarawala and J. N. Pelkowski, "Telehealth benefits and barriers," *The Journal for Nurse Practitioners*, vol. 17, no. 2, pp. 218–221, 2021.

[4] T. Yamamoto, M. Otsuki, H. Kuzuoka, and Y. Suzuki, "Tele-guidance system to support anticipation during communication," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 55, 2018.

[5] T. Kaneko, N. Kagiyama, Y. Nakamura, T. Hirasawa, A. Murata, R. Morimoto, S. Miyazaki, and T. Minamino, "Effectiveness of real-time tele-ultrasound for echocardiography in resource-limited medical teams," *Journal of Echocardiography*, vol. 20, no. 1, pp. 16–23, 2022.

[6] K. Jemal, D. Ayana, F. Tadesse, M. Adefris, M. Awol, M. Tesema, B. Dagne, S. Abeje, A. Bantie, M. Butler *et al.*, "Implementation and evaluation of a pilot antenatal ultrasound imaging programme using tele-ultrasound in ethiopia," *Journal of Telemedicine and Telecare*, p. 1357633X221115746, 2022.

[7] J. Montoya, S. Stawicki, D. C. Evans, D. Bahner, S. Sparks, R. Sharpe, and J. Cipolla, "From fast to e-fast: an overview of the evolution of ultrasound-based traumatic injury assessment," *European journal of trauma and emergency surgery*, vol. 42, no. 2, pp. 119–126, 2016.

[8] P. J. Mariani and J. A. Setla, "Palliative ultrasound for home care hospice patients," *Academic Emergency Medicine*, vol. 17, no. 3, pp. 293–296, 2010.

[9] N. Smallwood, A. Walden, P. Parulekar, and M. Dachsel, "Should point-of-care ultrasound become part of healthcare worker testing for covid?" *Clinical Medicine*, vol. 20, no. 5, p. 486, 2020.

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

[11] A. E. Drake, J. Hy, G. A. MacDougall, B. Holmes, L. Icken, J. W. Schrock, and R. A. Jones, "Innovations with tele-ultrasound in education sonography: the use of tele-ultrasound to train novice scanners," *The ultrasound journal*, vol. 13, no. 1, pp. 1–8, 2021.

[12] N. J. Soni, J. S. Boyd, G. Mints, K. C. Proud, T. P. Jensen, G. Liu, B. K. Mathews, C. K. Schott, L. Kurian, C. M. LoPresti *et al.*, "Comparison of in-person versus tele-ultrasound point-of-care ultrasound training during the covid-19 pandemic," *The ultrasound journal*, vol. 13, no. 1, pp. 1–7, 2021.

[13] J. Torres-Macho, T. Aro, I. Bruckner, C. Cogliati, O. Gilja, A. Gurghean, E. Karlafti, M. Krsek, Z. Monhart, A. Müller-Marbach *et al.*, "Point-of-care ultrasound in internal medicine: A position paper by the ultrasound working group of the european federation of internal medicine," *European journal of internal medicine*, vol. 73, pp. 67–71, 2020.

[14] S. E. Salcudean, G. Bell, S. Bachmann, W.-H. Zhu, P. Abolmaesumi, and P. D. Lawrence, "Robot-assisted diagnostic ultrasound–design and feasibility experiments," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1999, pp. 1062–1071.

[15] M. Akbari, J. Carriere, T. Meyer, R. Sloboda, S. Husain, N. Usmani, and M. Tavakoli, "Robotic ultrasound scanning with real-time image-based force adjustment: Quick response for enabling physical distancing during the covid-19 pandemic," *Frontiers in Robotics and AI*, vol. 8, 2021.

[16] C. Delgorge, F. Courrèges, L. A. Bassit, C. Novales, C. Rosenberger, N. Smith-Guerin, C. Brù, R. Gilabert, M. Vannoni, G. Poisson *et al.*, "A tele-operated mobile ultrasound scanner using a light-weight robot," *IEEE transactions on information technology in biomedicine*, vol. 9, no. 1, pp. 50–58, 2005.

[17] K. Mathiassen, J. E. Fjellin, K. Glette, P. K. Hol, and O. J. Elle, "An ultrasound robotic system using the commercial robot ur5," *Frontiers in Robotics and AI*, vol. 3, p. 1, 2016.

[18] P. Vieyres, G. Poisson, and et al., "A tele-operated robotic system for mobile tele-echography: The otelo project," in *M-health*. Springer, 2006, pp. 461–473.

[19] K. Li, Y. Xu, and M. Q.-H. Meng, "An overview of systems and techniques for autonomous robotic ultrasound acquisitions," *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 2, pp. 510–524, 2021.

[20] S. Wang, J. Housden, Y. Noh, D. Singh, and A. Singh, "Robotic-assisted ultrasound for fetal imaging: Evolution from single-arm to dual-arm system," in *Annual Conference Towards Autonomous Robotic Systems*, vol. 11650, July 2019, pp. 27–38.

[21] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab, "Automatic force-based probe positioning for precise robotic ultrasound acquisition," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 11 200–11 211, 2020.

[22] S. Virga, O. Zettinig, M. Esposito, K. Pfister, B. Frisch, T. Neff, N. Navab, and C. Hennersperger, "Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms," in *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2016, pp. 508–513.

[23] Q. Huang, J. Lan, and X. Li, "Robotic arm based automatic ultrasound scanning for three-dimensional imaging," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1173–1182, 2018.

[24] G. Ning, X. Zhang, and H. Liao, "Autonomic robotic ultrasound imaging system based on reinforcement learning," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 9, pp. 2787–2797, 2021.

[25] A. M. Priester, S. Natarajan, and M. O. Culjat, "Robotic ultrasound systems in medicine," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 60, no. 3, pp. 507–523, 2013.

[26] S. E. Salcudean, H. Moradi, D. G. Black, and N. Navab, "Robot-assisted medical imaging: A review," *Proceedings of the IEEE*, 2022.

[27] J. Carmigniani and B. Furht, "Augmented reality: an overview," *Handbook of augmented reality*, pp. 3–46, 2011.

[28] S. Rokhsaritalemi, A. Sadeghi-Niaraki, and S.-M. Choi, "A review on mixed reality: Current trends, challenges and prospects," *Applied Sciences*, vol. 10, no. 2, p. 636, 2020.

[29] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[30] R. Skarbez, M. Smith, and M. Whitton, "Revisiting milgram and kishino's reality-virtuality continuum. front," *Presence and Beyond: Evaluating User Experience in AR/MR/VR*, p. 8, 2022.

[31] D. Black and S. Salcudean, "Human performance in augmented reality ultrasound teleoperation," *TechRxiv*, Oct 2022.

[32] D. B. Kaber and T. Zhang, "Human factors in virtual reality system design for mobility and haptic task performance," *Reviews of Human Factors and Ergonomics*, vol. 7, no. 1, pp. 323–366, 2011.

[33] C. Jay and R. Hubbold, "Delayed visual and haptic feedback in a reciprocal tapping task," in *First joint eurohaptics conference and*

symposium on haptic interfaces for virtual environment and teleoperator systems. World Haptics Conference. IEEE, 2005, pp. 655–656.

[34] C. Jay, M. Glencross, and R. Hubbold, "Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 14, no. 2, pp. 8–es, 2007.

[35] K. Masuda, N. Tateishi, Y. Suzuki, E. Kimura, Y. Wie, and K. Ishihara, "Experiment of wireless tele-echography system by controlling echographic diagnosis robot," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2002, pp. 130–137.

[36] K. Masuda, E. Kimura, N. Tateishi, and K. Ishihara, "Three dimensional motion mechanism of ultrasound probe and its application for tele-echography system," in *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, vol. 2. IEEE, 2001, pp. 1112–1116.

[37] G. Niemeyer and J.-J. Slotine, "Stable adaptive teleoperation," *IEEE Journal of oceanic engineering*, vol. 16, no. 1, pp. 152–162, 1991.

[38] Y. Yokokohji, T. Imaida, and T. Yoshikawa, "Bilateral teleoperation under time-varying communication delay," in *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No.99CH36289)*, vol. 3, 1999, pp. 1854–1859 vol.3.

[39] K. Natori, T. Tsuji, K. Ohnishi, A. Hace, and K. Jezernik, "Time-delay compensation by communication disturbance observer for bilateral teleoperation under time-varying delay," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 3, pp. 1050–1062, 2010.

[40] Y. Ye, Y.-J. Pan, and T. Hilliard, "Bilateral teleoperation with time-varying delay: A communication channel passification approach," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 4, pp. 1431–1434, 2013.

[41] G. Leung, B. Francis, and J. Apkarian, "Bilateral controller for teleoperators with time delay via /spl mu/-synthesis," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 1, pp. 105–116, 1995.

[42] "Rfc 4960: Stream control transmission protocol," Sept 2007. [Online]. Available: https://www.rfc-editor.org/rfc/rfc4960

[43] "Rfc 1889: Rtp: A transport protocol for real-time applications," Jan 1996. [Online]. Available: https://www.rfc-editor.org/rfc/rfc1889

[44] H. Iqbal, F. Tatti, and F. R. y Baena, "Augmented reality in robotic assisted orthopaedic surgery: A pilot study," *Journal of Biomedical Informatics*, vol. 120, p. 103841, 2021.

[45] C. He, P. Kazanzides, H. T. Sen, S. Kim, and Y. Liu, "An inertial and optical sensor fusion approach for six degree-of-freedom pose estimation," *Sensors*, vol. 15, no. 7, pp. 16 448–16 465, 2015.

[46] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[47] T. Schimmoeller, R. Colbrunn, T. Nagle, M. Lobosky, E. E. Neumann, T. M. Owings, B. Landis, J. E. Jelovsek, and A. Erdemir, "Instrumentation of off-the-shelf ultrasound system for measurement of probe forces during freehand imaging," *Journal of biomechanics*, vol. 83, pp. 117–124, 2019.

[48] M. W. Gilbertson and B. W. Anthony, "An ergonomic, instrumented ultrasound probe for 6-axis force/torque measurement," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 140–143.

[49] M. O. Harris-Love, C. Ismail, R. Monfaredi, H. J. Hernandez, D. Pennington, P. Woletz, V. McIntosh, B. Adams, and M. R. Blackman, "Interrater reliability of quantitative ultrasound using force feedback among examiners with varied levels of experience," *PeerJ*, vol. 4, p. e2146, 2016.

[50] G. P. Mylonas, P. Giataganas, M. Chaudery, V. Vitiello, A. Darzi, and G.-Z. Yang, "Autonomous efast ultrasound scanning by a robotic manipulator using learning from demonstrations," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 3251–3256.

[51] M. Muir, P. Hrynkow, R. Chase, D. Boyce, and D. Mclean, "The nature, cause, and extent of occupational musculoskeletal injuries among sonographers: recommendations for treatment and prevention," *Journal of Diagnostic Medical Sonography*, vol. 20, no. 5, pp. 317–325, 2004.

[52] L. Whitcomb, S. Arimoto, T. Naniwa, and F. Ozaki, "Adaptive model-based hybrid control of geometrically constrained robot arms," *IEEE Transactions on Robotics and Automation*, vol. 13, no. 1, pp. 105–116, 1997.

[53] L. G. Carlton, "Visual processing time and the control of movement," in *Advances in psychology*. Elsevier, 1992, vol. 85, pp. 3–31.

[54] D. Badau, B. Baydil, and A. Badau, "Differences among three measures of reaction time based on hand laterality in individual sports," *Sports*, vol. 6, no. 2, p. 45, 2018.

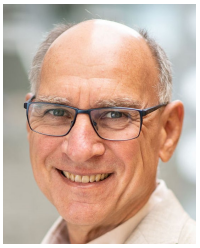
[55] G. E. Brogmus, "Effects of age and sex on speed and accuracy of hand movements: And the refinements they suggest for fitts' law," in *Proceedings of the Human Factors Society Annual Meeting*, vol. 35, no. 3. SAGE Publications Sage CA: Los Angeles, CA, 1991, pp. 208–212.

[56] T. Darbutas, V. Juodžbalienė, A. Skurvydas, and A. Kriščiūnas, "Dependence of reaction time and movement speed on task complexity and age," *Medicina*, vol. 49, no. 1, p. 4, 2013.

[57] L. Drukker, H. Sharma, R. Droste, M. Alsharid, P. Chatelain, J. A. Noble, and A. T. Papageorghiou, "Transforming obstetric ultrasound into data science using eye tracking, voice recording, transducer motion and ultrasound video," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.

David G. Black was born in Mainz, Germany, in 1998. He completed a BASc in engineering physics at the University of British Columbia (UBC), Canada, in 2021. He is currently a PhD student and Vanier Scholar in electrical and computer engineering at UBC. During his studies, he has worked as a Robotics Engineer Intern at A&K Robotics, Vancouver, Canada and a Research Student at the Robotics and Control Laboratory (RCL) at UBC, and at the BC Cancer Research Centre. From 2018 to 2019 he worked as a systems engineer in Advanced Development at Carl Zeiss Meditec AG, Oberkochen, Germany, and has continued as a consultant and collaborator since 2019.

Dragan Andjelic was born in Foča, Bosnia and Herzegovina. He completed a BASc in electrical engineering at the University of Belgrade, Serbia, in 1983. He is currently working on the 5G cellular network integration at UBC with research projects, including defining network requirements, designing research systems, and development/selection of required network tools. His experience includes working for Motorola Ltd. and Viavi Solutions Inc. as a Principal Engineer, where he was responsible for the design and development of different cellular system protocols, functions, and network services. During that time he was also supporting company involvement in 3GPP system specification and analysis of new features introduced by the new 3GPP specification releases.



Septimu E. Salcudean was born in Cluj, Romania. He received the BEng (Hons.) and MEng degrees in from McGill University, Montreal, Quebc, Canada in 1979 and 1981, respectively, and his PhD degree from the University of California, Berkeley, USA in 1986, all in electrical engineering.

He was a Research Staff Member at the IBM T.J. Watson Research Center from 1986 to 1989. He then joined the University of British Columbia (UBC) and currently is a Professor in the Department of Electrical and Computer Engineering, where he holds the C.A. Laszlo Chair in Biomedical Engineering and a Canada Research Chair. He has courtesy appointments with the UBC School of Biomedical Engineering and the Vancouver Prostate Centre. He has been a co-organizer of the Haptics Symposium, a Technical Editor and Senior Editor of the IEEE Transactions on Robotics and Automation, and on the program committees of the ICRA, MICCAI and IPCAI Conferences. He is currently on the steering committee of the IPCAI conference and on the Editorial Board of the International Journal of Robotics Research. He is a Fellow of the IEEE, a Fellow of MICCAI and of the Canadian Academy of Engineering.