Advanced Robotics Research

ADVANCED ROBOTICS RESEARCH Open Access

RESEARCH ARTICLE OPEN ACCESS

Real-Time Holographic Feedback via Mixed Reality Sensorized Laryngoscope Training System Enhances Pediatric Intubation Training

Jiaqi Li[1] | Ryman Hashem[1,2] | Ningzhe Hou[1,3] | Perla Maiolino[1,3] | Louis Halamek[4] | Liang He[1,2]

[1]Department of Engineering Science, University of Oxford, Oxford, UK | [2]The podium institute for sports medicine and technology, University of Oxford, Oxford, UK | [3]Oxford Robotics Institute, University of Oxford, Oxford, UK | [4]Department of Pediatrics, Stanford University, Palo Alto, California, USA

Correspondence: Liang He (liang.he@eng.ox.ac.uk)

Received: 11 January 2026 | **Revised:** 10 May 2026 | **Accepted:** 28 May 2026

Keywords: force and torque visualization | intubation training | medical simulation | mixed reality | real-time holographic feedback | sensorized laryngoscope

ABSTRACT

Pediatric endotracheal intubation (ETI) is a technically demanding procedure in which novices apply excessive force and torque. Traditional manikin training and instructor feedback lack quantitative, real-time performance data. This work introduces a Mixed Reality Sensorized Laryngoscope Training System, integrating a sensorized laryngoscope with head-mounted holographic feedback to visualize torque in real time. In Study 1, six subjects performed a force-tracking task with a simplified laryngoscope while viewing three holographic visualizations (bar, ring, and line graphs). Bar-graph feedback yielded the lowest tracking error ($p < 0.05$). Eye-tracking showed ring graphs elicited significantly more saccades than line graphs ($p < 0.05$), consistent with higher visual demand, while bar and line graphs produced comparable saccade behavior. In Study 2, ten novices, split into mixed-reality and control groups, performed pediatric ETI on a manikin with or without real-time torque feedback during training. In post-training trials without feedback, the feedback-trained group maintained torque more consistently within expert-defined ranges without increasing intubation time. These results suggest real-time mixed-reality feedback can produce lasting improvement in torque regulation during simulated pediatric ETI while preserving procedural efficiency. As a pilot study ($n = 6$ and $n = 10$), these findings motivate larger-scale validation and inform future telesimulation-based training.

1 | Introduction

Endotracheal intubation (ETI) is a critical procedure for airway management in various clinical settings: both emergency and surgical. It involves guiding the endotracheal tube from the oral cavity into the trachea, typically using a laryngoscope to lift the epiglottis and allow passage through the vocal cords (Figure 1). Successful intubation requires precise hand–eye coordination and dexterous manipulation of the laryngoscope; however, novice practitioners often apply excessive force or adopt suboptimal techniques compared to experts [1, 2]. These errors can cause mechanical complications such as oropharyngeal soft-tissue trauma and dental injury, and, in more severe cases, repeated attempts or failed intubations are associated with resultant hypoxemia, cardiovascular instability, or cardiac arrest [3].

Traditional ETI training methods rely on simulated intubation performed on manikins and verbal feedback from expert instructors. While these approaches are useful for teaching procedural steps, they often fall short in conveying the fine motor control and detailed force application needed for safe and effective intubation [4]. To address this, Hou et al. [5] developed the Sensorized Laryngoscope Training System

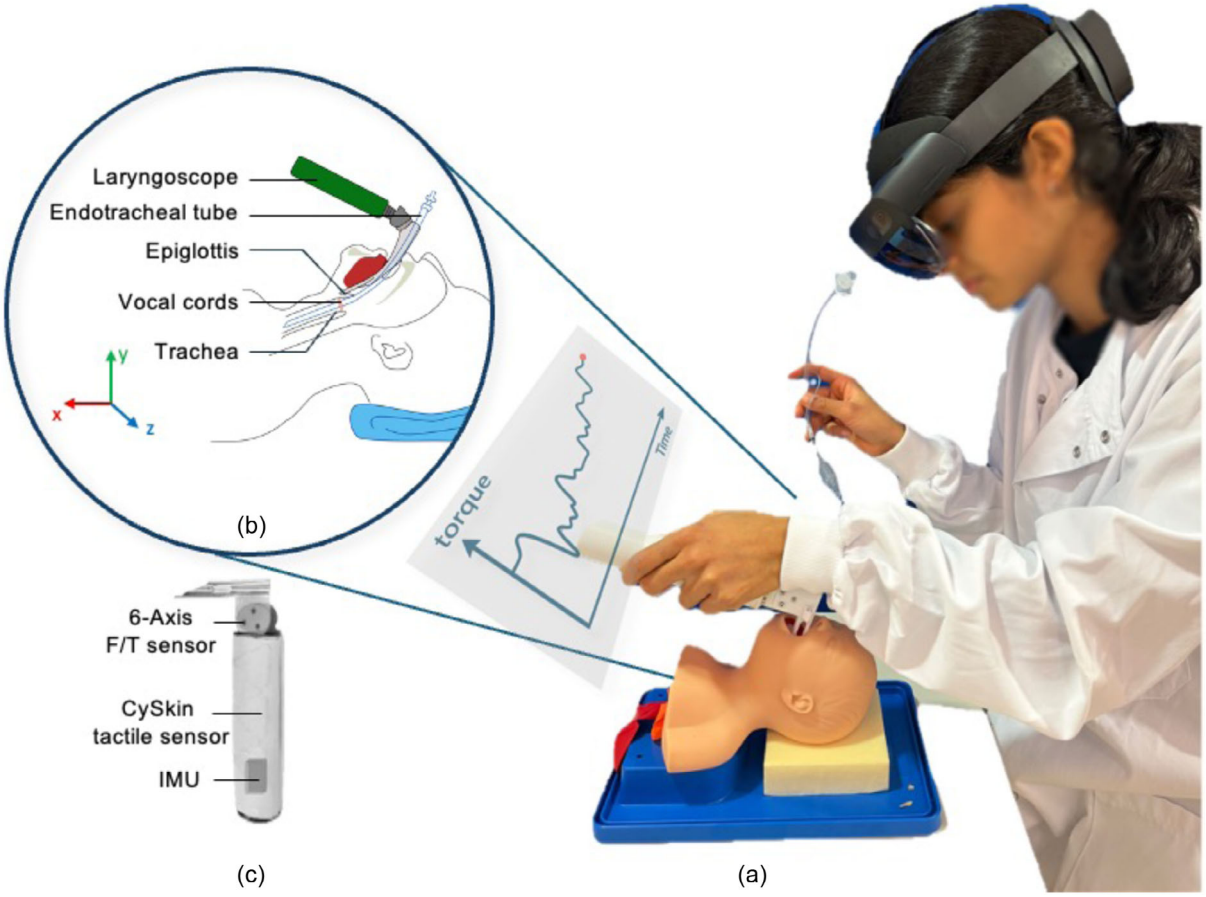

FIGURE 1 | (a) Illustration of the Mixed Reality SLTS (MR-SLTS). Real-time visual feedback of contact force/torque is delivered to the trainee through a head-mounted holographic device, supporting skill acquisition in contact force/torque regulation. (b) The pediatric ETI procedure. The laryngoscope is used to lift the epiglottis and visualize the vocal cords through coordinated rotational and lifting maneuvers, allowing the endotracheal tube to pass through the vocal cords into the trachea. (c) Schematic of the sensorized laryngoscope.

(SLTS), embedding sensors within a laryngoscope to provide real-time quantitative feedback on applied forces and device motion.

While the SLTS enabled objective and quantitative skill assessment, its monitor-based feedback required users to divide attention between the manikin and an external display, motivating the development of a mixed-reality alternative.

To overcome these limitations, this study introduces the MR-SLTS, which delivers live visual feedback directly within the user's field of view using a head-mounted holographic device (Microsoft HoloLens 2). A schematic of the system can be found in Figure 1. By overlaying information onto the physical environment, the system aims to improve force interpretability, minimize attention shifts, and reduce cognitive burden. Two controlled user studies were conducted to evaluate its effectiveness. Study 1 examined how different formats of holographic feedback design influence force calibration and visual attention during a simplified sensorimotor task, while Study 2 assessed the impact of MR torque feedback on pediatric ETI training outcomes. The results provide preliminary evidence for the feasibility of real-time MR feedback for skill acquisition and offer initial design insights for next-generation robotic training simulation.

2 | Related Work

2.1 | Sensorized Training Systems for ETI

Research on ETI training systems has increasingly focused on using sensor technologies to address the complexity of the procedure. These systems typically employ real-time feedback to help trainees perceive forces and positions that are difficult to judge through conventional means. For instance, Zhou et al. [6]. developed a sensor-embedded phantom that provided vibrotactile cues in response to lifting and pushing forces applied to the tongue, which significantly reduced excessive force application.

Other approaches incorporate a broader range of sensing modalities to monitor device positioning and force distribution across the oral cavity. Notable examples include the sensorized video laryngoscope neonatal intubation simulator developed by Covelli et al. [7] and the Difficult Airway Management Simulator Evaluation System by Noh et al. [8].

Expanding upon these efforts, Hou et al. [5] introduced the SLTS, which integrates multiple sensors into a single laryngoscope to provide a more comprehensive assessment of novice performance. The system uses a six-axis force/torque sensor

(Nano 17, ATI Industrial Automation) to measure blade forces, a nine-axis inertial measurement unit (IMU) (GY-95T, Guangyun Electronics) for orientation tracking, and distributed CySkin tactile sensors [9] to monitor grip pressure, all without altering the tool's form factor. Real-time sensor signals were displayed on an external monitor to give users immediate feedback during trials. While preliminary studies confirmed the robustness of sensor integration and data synchronization, subjects reported challenges related to cognitive load due to frequent attention shifts between the manikin and monitor. Additionally, the feedback's numerical presentation made interpretation difficult during the task. These insights motivated the development of the MR-SLTS, which seeks to deliver more intuitive and immersive feedback using a mixed reality interface, as described in the following sections.

2.2 | Mixed Reality in Medical Training

MR has become an increasingly valuable tool in simulation-based medical and surgical training, offering holographic overlays that present procedural information directly within the trainee's field of view. MR systems using optical see-through head-mounted displays (OST-HMDs), such as the HoloLens 2, have been applied to surgical guidance [10], preoperative planning [11], and intraoperative navigation [12]. Studies show that presenting information directly in the user's line of sight reduces the cognitive effort associated with attention switching, thereby improving learning outcomes [13, 14]. More broadly, the sense of presence experienced by users in immersive digital environments has been shown to influence task engagement and performance, and recent work has proposed neurophysiological approaches to quantify this construct [15].

While MR is increasingly used in surgical contexts, relatively few studies have explored its application in airway management training. Bhavsar et al. [16]. demonstrated that the HoloLens 2 could be used effectively to deliver instructional videos, remote coaching, and augmented visual annotations during intubation training. Extending this line of research, our study investigates the impact of using MR to provide real-time quantitative feedback through holographic visualization and its influence on ETI training performance.

2.3 | Eye-Gaze Tracking and Cognitive Load in Medical Training

Medical procedures that require fine motor control and high levels of attention, such as ETI, are particularly susceptible to cognitive overload when trainees must process large volumes of information simultaneously. Eye tracking has been widely used to assess cognitive load in surgical and robotic training environments [17]. Evidence suggests that frequent rapid eye movements (saccades) may fragment visual perception and disrupt the temporal alignment between visual input and motor output, which is critical for efficient sensorimotor coordination [18, 19].

This study investigates how different holographic visualization formats influence gaze fixation patterns and attentional focus. However, saccade-based metrics capture one dimension of cognitive load. Neurophysiological benchmarking studies have shown that oculomotor parameters such as fixation count and saccade frequency, while sensitive to changes in cognitive state, can lag behind EEG-derived indices and may not fully disambiguate workload from other attentional demands without complementary measures [20, 21]. The eye-tracking component of this study is therefore scoped to characterize oculomotor behavior differences across feedback conditions. The resulting insights inform the design of feedback strategies within the MR-SLTS and contribute to a deeper understanding of cognitive load management in immersive training environments.

3 | Design and Methods

3.1 | System Architecture

The architecture of the MR-SLTS extends the previously developed SLTS [5], incorporating a mixed reality interface for enhanced visual feedback as illustrated in Figure 2. The MR-SLTS preserves the original data acquisition and processing pipeline while introducing an MR visual feedback interface via HoloLens 2. The data acquisition, processing, and visualization pipeline of MR-SLTS is illustrated in Figure 2a.

Specifically, sensing signals from the embedded 6-axis F/T sensor (Nano 17, ATI Industrial Automation) and nine-axis IMU (GY-95T, Guangyun Electronics) are collected using an NI DAQ device (USB-6363, National Instruments) and a microcontroller board (Arduino UNO, Arduino S.r.l.), respectively. These signals are then transmitted to a desktop-based server running MATLAB (R2023b, MathWorks), which is responsible for real-time data processing, storage, and bidirectional communication with the HoloLens 2 (Microsoft).

The HoloLens 2, wirelessly interfaced with the MATLAB server via the TCP protocol, subscribes to the processed sensing data and renders corresponding holographic visualizations in the user's field of view. In parallel, the HoloLens 2 also collects eye-gaze data throughout the training session, which is transmitted back to the MATLAB server for synchronized storage and later analysis.

3.2 | Development of the MR Feedback Platform

The MR-SLTS software platform was developed using Unity (Unity 2022.3.20f1, Unity Technologies), a widely used engine for creating holographic applications tailored to the HoloLens 2 headset. The development process was supported by the Mixed Reality Toolkit (MRTK) (version 2.8.3, Microsoft), an open source framework from Microsoft that streamlines the creation of mixed and augmented reality applications by providing prebuilt components and interaction models optimized for HoloLens devices.

Visual Studio 2022 (Microsoft) was used to build, debug, and deploy the application to the HoloLens 2. The project also leveraged the Universal Windows Platform Software Development Kits (SDKs), which offer the runtime libraries and APIs necessary

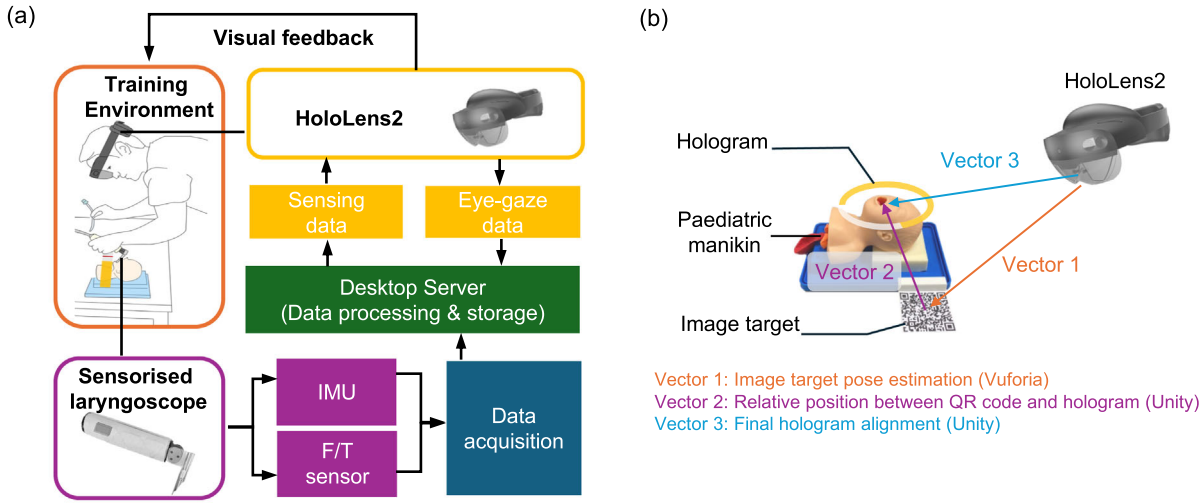

FIGURE 2 | (a) Overview of the MR-SLTS data pipeline. Sensing signals from the F/T sensor and IMU are acquired via NI DAQ and Arduino, processed on a desktop-based MATLAB server, and streamed to the HoloLens 2 for real-time visual feedback in the training environment, with gaze data sent back for synchronized analysis. (b) QR code–based tracking and hologram rendering process. Vuforia detects the QR code and estimates its pose (Vector 1). Unity applies the predefined relative transform (Vector 2). Unity then computes the final hologram position (Vector 3), ensuring consistent alignment above the manikin's face across sessions.

for running Unity-based applications on Windows-compatible MR hardware.

These tools collectively enabled the creation of an interactive Unity application capable of rendering real-time numerical feedback within the HoloLens 2's MR environment.

3.2.1 | Visual Feedback Registration

In the MR-SLTS design, the precision requirement for holographic visual feedback registration is relatively modest, with deviations below 5 mm considered acceptable based on informal usability testing. This tolerance reflects the acceptable visual displacement between the intended hologram location and its perceived position in the user's field of view. Since the feedback is not anatomically anchored but is instead presented near the manikin to support user interpretation, such minor deviations do not impair usability.

To ensure robust and repeatable spatial anchoring of holograms, the Vuforia Engine SDK (version 10.21.3, PTC) was employed to track a feature-rich 8 × 8 cm QR-code image target. By relying on 2D image target detection, which offers higher precision and lower latency than 3D object detection, the system maintained a stable spatial reference for registration.

The QR-code target was mounted on a 3D-printed tracking guide attached to the lower-left corner of the pediatric manikin. Figure 2b illustrates the tracking and hologram-rendering workflow. Details of the procedure can also be found in the supplementary material. The Vuforia Engine first detects the QR code and estimates its pose (Vector 1) within the HoloLens spatial-mapping coordinate system. Unity then applies the predefined relative transform between the image target and the hologram (Vector 2) to render the overlay in a consistent position and orientation above the manikin's face (Vector 3). This configuration ensured that the hologram remained spatially aligned both within a single session and across repeated training sessions, even when the manikin was physically repositioned.

3.2.2 | Holographic Feedback Design

To minimize cognitive load during ETI, MR-SLTS simplifies how quantitative sensor data are presented. Previous user studies indicated that the error bar format used in the design of SLTS was difficult to interpret in real time [5], often overwhelming trainees during motor-intensive tasks [22, 23].

Comparative analysis between experts and novices has identified torque about the z-axis (coordinate system defined in Figure 1a) as a key indicator of improper wrist rotation in novices [24], and one that improves measurably with training. This torque corresponds to the rotational force applied when lifting the laryngoscope blade, which, if excessive or misaligned, can lead to patient injury or failed intubation. Measurements of linear force are confounded by the relatively stiff jaw, upper airway, and tracheal structures, as well as variable contact conditions, whereas torque more directly reflects the operator's wrist rotation and blade manipulation. Due to its clear interpretability and sensitivity to skill progression, this single parameter was selected as the target for real-time visual feedback. Focusing on one dimension also reduces information overload and helps trainees concentrate on a specific skill.

To deliver continuous and easily interpreted visual feedback, three feedback formats were prototyped: a bar graph, which emphasizes simplicity but lacks temporal context; a ring graph, which conveys proportions effectively but similarly omits temporal changes; and a line graph, which illustrates trends over time but is visually more complex. The selected visualization was spatially calibrated to remain within the user's consistent field of view relative to the manikin, ensuring accessibility and reducing distraction during ETI trials. The holographic feedback designs are shown in Figure 3.

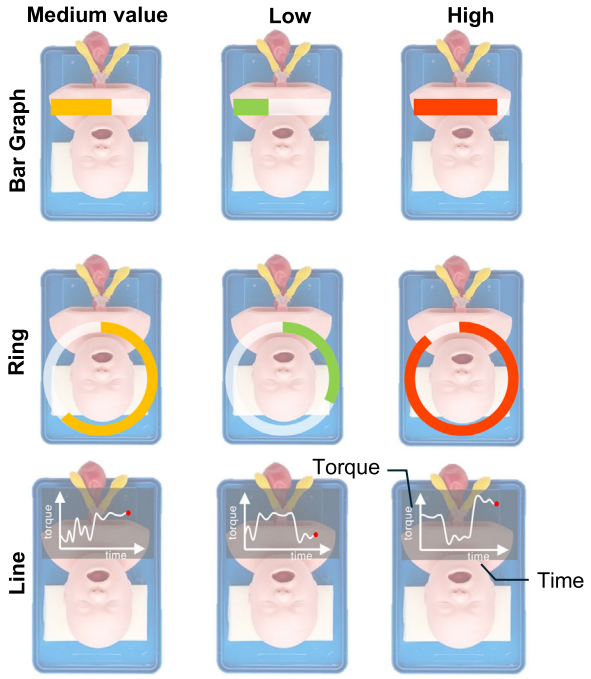

FIGURE 3 | Comparison of visual feedback formats. Each row corresponds to one feedback format: bar (top), ring (middle), and line graph (bottom). Each column shows the same sensor reading at the same time point rendered in the three formats. This arrangement illustrates the different ways in which identical data are conveyed, highlighting trade-offs between simplicity, proportional representation, and temporal trend visualization.

3.2.3 | Wireless Communication and Latency Considerations

Reliable and low-latency communication between the HoloLens 2 and the desktop server is essential for effective real-time visual feedback. Several wireless protocols were considered, including TCP, UDP, HTTP, and Bluetooth-based alternatives. TCP over WiFi was ultimately selected for its high reliability, consistent data ordering, and well-documented implementation support in both Unity (via MRTK) and MATLAB.

Although UDP may offer lower latency, our preliminary tests showed that TCP communication incurred an average delay of only 4.17 ms per transmission. The HoloLens 2 introduced an additional display latency of 16 ms [25], while the improved sensor data readout rate of 70 Hz corresponded to a delay of approximately 14 ms. Together, these factors kept the total system latency well below the 100 ms threshold commonly cited as acceptable for surgical and motor feedback tasks [26, 27]. Consequently, the holographic feedback appeared in near real time and did not perceptibly disrupt trainee performance.

3.2.4 | Eye-Gaze Tracking Integration

The HoloLens 2 provides built-in eye-tracking capabilities, which were utilized through the MRTK in Unity to collect eye-gaze direction and head-movement data during intubation trials. These data were transmitted in real time via TCP communication to a desktop-based server for storage and postanalysis. This setup enabled detailed examination of trainee visual attention patterns, including fixation behavior and saccade frequency.

4 | Experimental Protocol and Setup

This research comprises two sequential studies designed to inform and evaluate the holographic feedback design used in MR-SLTS. Study 1 compares how different visualization formats (e.g., bar, ring, and line graphs) convey real-time quantitative torque data during a simplified intubation-related task in an MR environment. The goal is to identify the most effective and intuitive format based on user performance. The outcome of Study 1 informs the design of Study 2, which evaluates the training efficacy of the selected holographic feedback format. In Study 2, subjects engage with the full MR-SLTS system in a realistic pediatric intubation scenario to assess how the chosen visualization supports skill acquisition. An overview of the study timeline and design is provided in Figure S1.

4.1 | Study 1: Holographic Feedback Comparison

Study 1 aimed to evaluate the impact of different holographic feedback formats, including bar, ring, and line graphs, on users' performance in a simplified sensorimotor coordination task using HoloLens 2. The findings would inform the selection of a candidate feedback format for the MR-SLTS system, later evaluated in Study 2.

4.1.1 | Experimental Setup and Protocols

Six subjects with no prior experience with OST-HMDs were recruited. The sample included both left- and right-handed individuals. Subjects who normally used corrective lenses or contact lenses wore them during the experiments. Although clinical laryngoscopes are conventionally designed for left-handed use, subjects were permitted to use their preferred hand when conducting the experiments. All subjects provided their informed consent under an approved ethics protocol (CUREC R91525/RE001).

The experimental setup mirrored the MR-SLTS architecture in simplified form. A 3D-printed laryngoscope prototype with a single rotational degree of freedom was mounted on a support stand, as shown in Figure 4a. Compression forces applied by the blade were captured using a FlexiForce sensor (A201, 25 lb, Tekscan). A single-axis thin-film sensor was selected for this location, as only the normal contact force was of interest, and its thin profile allowed flush mounting on the flat support surface. A 10 × 15 × 3 mm sheet of platinum-cure silicone elastomer (Dragon Skin 20, Smooth-On, Shore A 20) was placed on the lingual contact surface of the 3D-printed blade—the surface that engages the tongue during jaw lifting in the ETI procedure. The elastomer compensated for surface irregularities of the 3D-printed blade and ensured uniform force transmission to the sensor. The sensor was affixed to the vertical surface of the support stand, aligned with the blade contact area. Force data was streamed via a microcontroller (Arduino MEGA, Arduino S.r.l.) to a laptop-based MATLAB server. Holographic feedback was rendered on

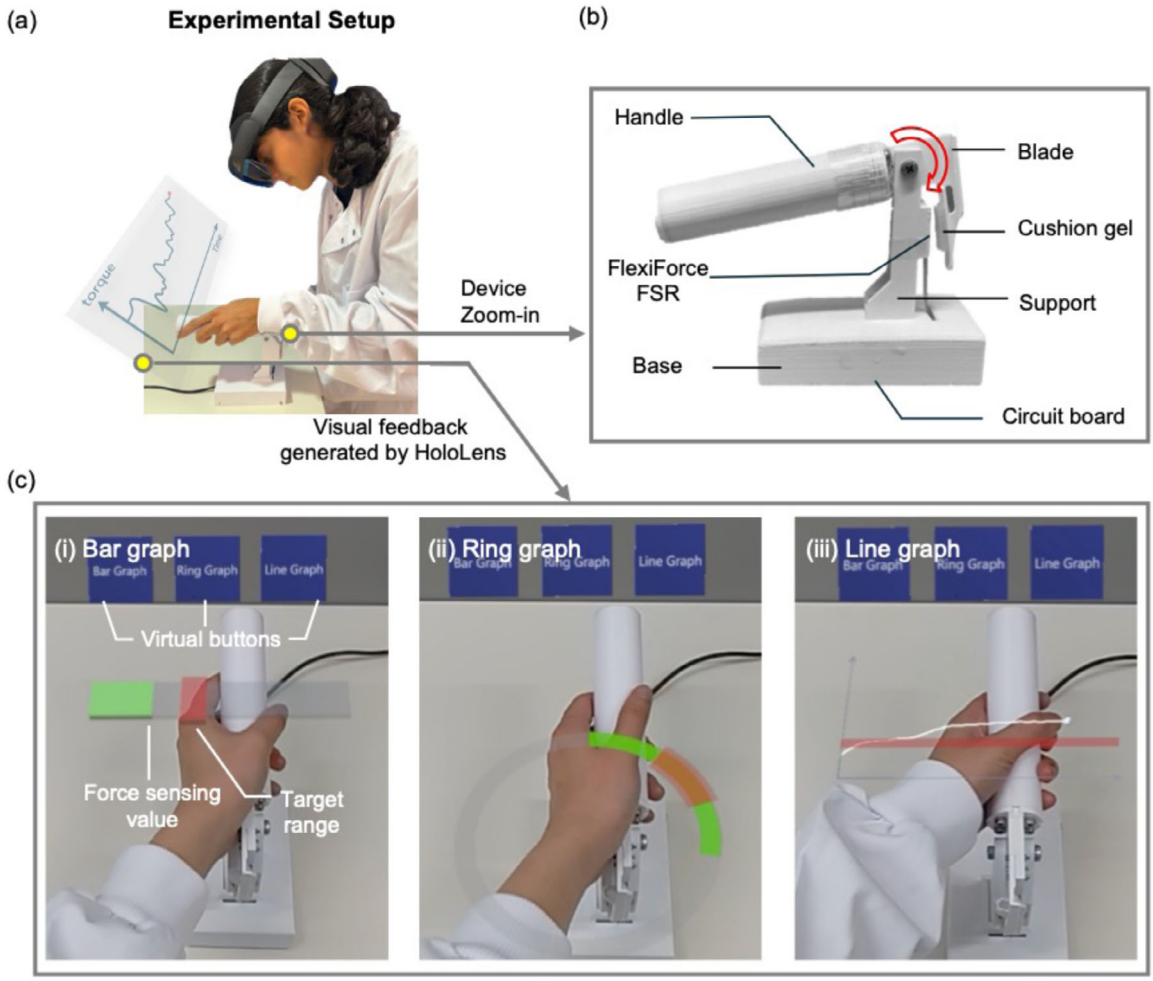

FIGURE 4 | (a) 3D-printed laryngoscope prototype with compression force measured by a FlexiForce sensor. (b) Experimental setup schematic. (c) HoloLens screenshots showing the subject's view for (i) bar graph, (ii) ring graph, and (iii) line graph.

HoloLens 2 using Unity with the MRTK framework. Eye-gaze data and force outputs were synchronized via bidirectional TCP communication for storage and analysis. The experimental setup for Study 1 is illustrated in Figure 4b.

Each graphical format (bar, ring, and line) was displayed above the laryngoscope in the user's field of view (HoloLens 2 photoshots shown in Figure 4c), with position calibrated using a QR code image target. Prior to trials, each subject completed eye-gaze calibration to ensure accurate data collection.

Each subject completed 21 trials, including six initial training trials excluded from analysis. The remaining 15 test trials consisted of five trials per feedback format, presented in randomized order. Each 20-s trial required subjects to adjust the applied force to remain within a time-varying target range $[L_{\min}, L_{\max}]$, defined by step-changing upper and lower limits designed to approximate the force regulation demands relevant to ETI, though in a simplified single-axis context (see Figure S2). Force data were normalized to a 0–100 scale, where 100 corresponded to a fixed maximum sensor reading established by the experimenter prior to data collection. Targets set between 10 and 45 to reflect delicate procedural demands. Eye-gaze direction and applied force were recorded continuously for post-trial analysis.

4.1.2 | Performance Metrics and Data Processing

The performance of the subjects was quantified using three metrics.

First, the tracking error was defined by comparing the applied force M with the target force range $L_{\min}$ and $L_{\max}$. At each time step, the instantaneous error was calculated as

$$\text{error}(\text{i}) = \begin{cases} \text{M}(\text{i}) - L_{\max}(\text{i}) & \text{if } \text{M}(\text{i}) > L_{\max}(\text{i}) \\ L_{\min}(\text{i}) - \text{M}(\text{i}) & \text{if } \text{M}(\text{i}) < L_{\min}(\text{i}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The root–mean-square error (RMSE) was then computed over the valid duration of each trial (excluding the first 2 s) and averaged across trials per feedback type.

Second, the eye-gaze shift distance defined the displacement between consecutive eye-gaze direction vectors $\rightarrow g_i = (x_i, y_i, z_i)$, and it was calculated as

$$G_{\text{shift}}(\text{i}) = \sqrt{(x_{\text{i}+1} - x_{\text{i}})^2 + (y_{\text{i}+1} - y_{\text{i}})^2 + (z_{\text{i}+1} - z_{\text{i}})^2} \quad (2)$$

A peak detection algorithm was applied on G_{shift} to identify gaze shifts exceeding 0.03 m, which were classified as saccades.

Third, the response time and fluctuation periods were computed using a threshold-based algorithm. The algorithm identifies the force calibration state as stabilized when the applied force remains within the target range continuously for more than 0.6 s after each change of target range, as illustrated in Figure S3. Applied to the force-sensing data, it segments periods of force fluctuation (i.e., active adjustment) and stabilization. From this, the average and median fluctuation durations following target changes were calculated, providing a measure of response speed.

All data were processed with Python. Statistical comparisons between feedback types were performed using paired t-tests or Mann–Whitney U tests, depending on normality.

4.2 | Study 2: MR-SLTS Evaluation

Study 2 aimed to evaluate the effect of real-time holographic visual feedback on z-torque in improving novice performance during ETI training. It was hypothesized that graphical torque feedback would help reduce excessive force application without significantly increasing procedure time or eye distraction.

4.2.1 | Experimental Setup and Protocols

A randomized controlled experiment was conducted with ten subjects (aged 22–27) with no prior ETI experience. Subjects were assigned to either a feedback group (received real-time Z-torque feedback via a bar graph in the HoloLens 2 display) or a control group (no feedback, but wore HoloLens for gaze data collection). Ethical approval was obtained from the Central University Research Ethics Committee (CUREC, R91525/RE001).

Experiments were conducted using the full MR-SLTS system, illustrated in Figure 5a. The sensorized laryngoscope transmitted z-torque and angular position data to a MATLAB-based server, which streamed processed values to the HoloLens 2 via TCP. CySkin tactile sensors were disabled to improve data throughput. Feedback consisted of a bar graph displaying real-time Z-torque values in the HoloLens 2 view. This format was selected based on the optimal results obtained in Study 1. Only negative torque was visualized, as it corresponds to the lifting action of the tongue during laryngoscopy. A red threshold line was included to indicate the maximum allowed z-torque derived from expert performance reported in prior literature (−370.65 Nmm), based on measurements from three licensed surgeons [5]. It should be noted that the small expert sample may constrain the robustness of this benchmark. Figure 5b shows the in-situ visualization as seen by subjects during training trials.

After an instructional video and three baseline trials with a standard laryngoscope, subjects completed eight training trials (trials 1–8) using MR-SLTS. Each ETI trial required them to insert an endotracheal tube into the manikin's trachea within a 30 s time limit. A trial was considered successful if, upon blowing air into the endotracheal tube using a resuscitation bag, the lung-representing balloons could be visibly inflated. Between trials, tube placement was verified, and the laryngoscope was recalibrated using IMU data. Feedback was enabled only for the experimental group. After a 5-min rest period, all subjects completed three post-training trials (trials 9–11) without visual feedback.

Force/torque, eye-gaze, and laryngoscope angle data were recorded for all trials and used to evaluate differences in training outcomes between groups.

4.2.2 | Performance Metrics and Data Processing

The ETI performance in Study 2 was evaluated using multiple quantitative metrics.

First, a Z-torque analysis was conducted using post-training trials (trials 9–11) to assess whether feedback-augmented training produced lasting differences in Z-torque control after real-time feedback was removed. Among post-training trials, attempts with unsuccessful tube placement or durations exceeding 30 s were excluded from analysis. For each successful trial, z-torque signals

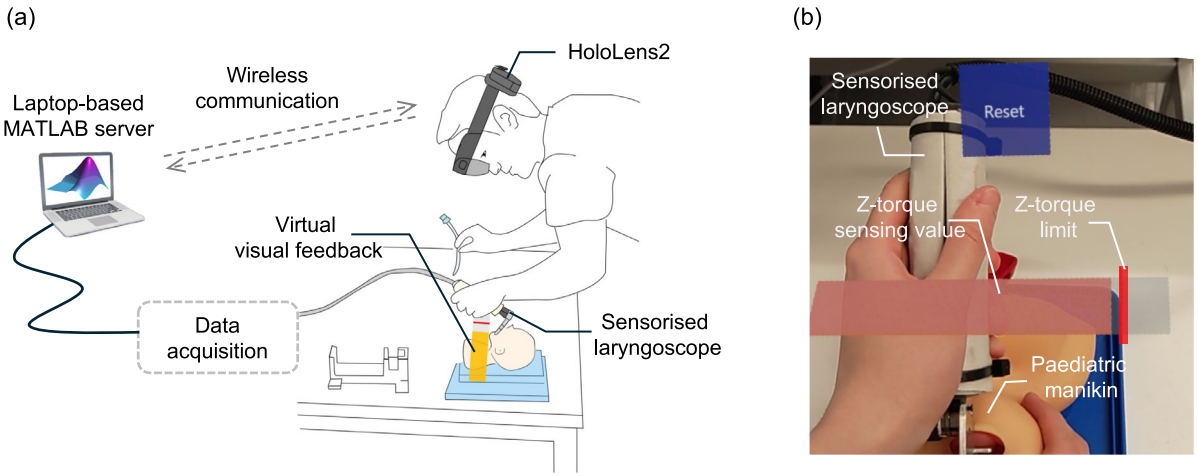

FIGURE 5 | Study 2: MR-SLTS. (a) Physical MR-SLTS setup and data flow. The sensorized laryngoscope transmits torque and position data to a MATLAB-based server, which streams the processed values to the HoloLens 2. The headset generates virtual visual feedback using real-time torque data and overlays it above the manikin within the user's field of view. (b) HoloLens 2 view showing real-time holographic feedback during intubation. The bar-graph feedback indicates the applied torque relative to the defined safe threshold, providing intuitive guidance to the operator.

TABLE 1 | L_{max} and L_{min} of F/T Sensing.

	F_x	F_y	F_z	T_x	T_y	T_z
L_{max}	4.31	1.53	1.16	66.9	79.73	15.53
L_{min}	−3.76	−4.21	−1.38	−28.31	−47.79	−370.65

were time-normalized to 100% trial completion. A sliding window of 1% width with a 0.5% step was then applied to compute smoothed mean and standard deviation curves for visual comparison of Z-torque profiles between groups.

Second, using post-training trials, the percentage errors of force and torque were calculated for each of the six components using the RMSE, following the same formulation as in Study 1 (Equation (1)). In this study, however, L_{max} and L_{min} were redefined as the expert-defined directional limits of force and torque, corresponding to the maximum allowable magnitudes in the positive and negative orientations of each axis ($F_x, F_y, F_z, T_x, T_y, T_z$). These limits were determined by Hou et al. [5] based on the mean performance of three licensed surgeons, each performing twenty intubation sessions. The corresponding directional limits for each force and torque component are summarized in Table 1. The percentage error was obtained as

$$\text{Percentage Error} = \left(\frac{\text{RMSE}}{\text{Max Absolute Range Value}}\right) \times 100\% \quad (3)$$

Third, intubation duration was recorded for all trials and analyzed across three stages: early training (trials 1–4), late training (trials 5–8), and post-training (trials 9–11). The distribution of trial durations across these three stages was analyzed to examine whether feedback-augmented training produced progressive changes in task completion time.

Eye-gaze shifts were processed following the same procedure as in Study 1. Gaze shift distances between consecutive frames were computed for all trials (both training and post-training), and peaks exceeding a 0.03 m threshold were classified as saccades. The total number of saccades per trial was used to compare oculomotor patterns between groups and across the three stages.

5 | Results

5.1 | Study 1 Results

To assess tracking performance, RMSE values were calculated based on deviations between the applied force and the target force range. The first 2 seconds of each trial were excluded to remove initial transients. Figure 6a shows the average RMSE per subject for each type of feedback. The bar graphs produced the lowest average error, with a statistically significant difference observed between the bar and ring conditions ($p < 0.05$, Student's t-test). The difference between bar and line feedback was not statistically significant ($p > 0.05$). The bar graph condition also exhibited the most consistent performance across and within subjects, as indicated by the lowest standard deviation among the three feedback methods.

Eye-gaze behavior was analyzed by counting saccadic shifts, based on displacement thresholds applied to consecutive gaze direction vectors. We applied a peak detection algorithm in Python using a single threshold to identify all peaks with a shift distance greater than 0.03 m, classifying them as saccades. An example of the detected saccades from an experimental trial is illustrated in Figure S5. Figure 6b presents the histogram of the total number of saccades per condition. The first bin (0–1 s) was excluded from analysis due to its notably higher count, reflecting the initial gaze adjustment phase. Trials with ring graphs yielded significantly more saccades than those with line graphs ($p < 0.05$, Mann–Whitney U test), while the difference between bar and line graphs was not significant ($p > 0.3$). Most saccades occurred shortly after changes in the target force values.

To further evaluate force control of subjects, a threshold-based segmentation algorithm was applied to identify periods of force

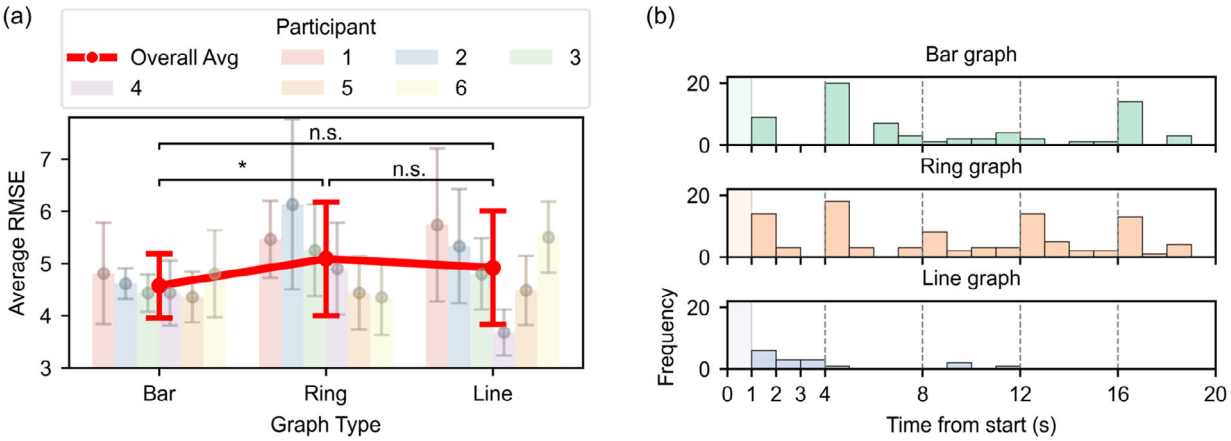

FIGURE 6 | Comparison of target-tracking performance and visual demand across different graph visualization formats in Study 1. (a) Average target tracking RMSE for each graph type across individual subjects in Study 1. The plot includes per-subject results, along with the overall average and standard deviation for each visualization format, highlighting differences in tracking accuracy. (b) Histogram showing the total number of saccades recorded for each visualization type during the target tracking task. This metric reflects the visual demand and eye movement load associated with interpreting each graph format.

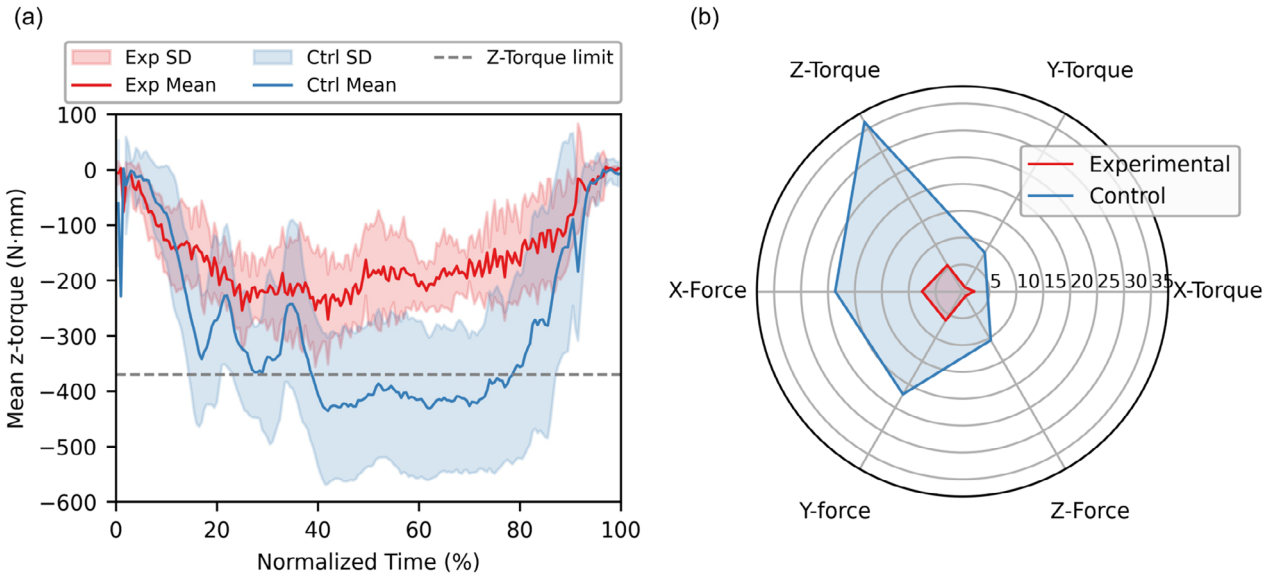

FIGURE 7 | (a) Average *Z*-axis torque measurements plotted over normalized time for all successful post-training trials. The curve represents the typical torque profile applied during ETI, providing insight into force patterns over the course of the procedure. (b) Comparison of average percentage errors in F/T sensing between the experimental group (trained with Z-torque feedback) and the control group (trained without feedback) during post-training trials conducted without real-time feedback for either group. Differences reflect the residual effect of feedback-augmented training on F/T control accuracy.

fluctuation following each target change. An example trial illustrating the segmented fluctuation and stabilized periods of the force sensing signal is shown in Figure S4.

Fluctuation durations were comparable across the three feedback formats. The mean (± median) durations were 2.04 s (1.77 s) for the bar graph, 2.09 s (1.86 s) for the ring graph, and 2.34 s (1.94 s) for the line graph. Although the line-graph feedback yielded slightly longer fluctuation periods than the bar and ring formats, the differences were not statistically significant (Mann–Whitney U tests: bar vs. ring, $p = 0.821$; bar vs. line, $p = 0.088$; ring vs. line, $p = 0.105$). These trends may reflect more cautious motor adjustments in response to the line graph's historical trend information, resulting in smoother and more controlled force modulation.

Based on the findings from Study 1, where bar graph feedback resulted in the lowest tracking error, moderate eye distraction, and faster force calibration in the simplified tracking task, this feedback modality was selected for further evaluation in Study 2.

5.2 | Study 2 Results

In the experiments of Study 2, out of 30 post-training trials (without active feedback) conducted by the 10 subjects, 9 were excluded due to unsuccessful intubation or exceeding the 30 s limit, resulting in 21 valid trials for analysis. Figure 7a shows the normalized z-torque trajectories for the valid post-training trials. The experimental group exhibited consistently lower torque values across the procedure compared to the control group.

Percentage errors for all six force and torque dimensions are summarized in Figure 7b. The experimental group demonstrated lower average percentage errors across all dimensions. The Z-torque error in the control group was the highest among all axes, and an approximate 30% reduction in the experimental group suggests that feedback-augmented training facilitated lasting improvement in rotational force control, persisting after feedback withdrawal.

The duration of intubation was analyzed across all trials (training and post-training), and no statistically significant differences were observed between the experimental and control groups or throughout the training progression ($p > 0.3$, Mann–Whitney U test; see Figure S6). Eye-gaze analysis, also conducted across all trials, revealed that the experimental group exhibited a higher average number of saccadic gaze shifts than the control group ($p < 0.05$, Mann–Whitney U test), a trend that persisted in both the training and posttraining stages.

6 | Discussion

This study investigated whether real-time holographic feedback delivered via a head-mounted display can support torque regulation during simulated pediatric ETI training.

In a preliminary sensorimotor task (Study 1), the bar graph feedback produced the lowest tracking error and comparable saccade behavior to the line graph, whereas the ring graph elicited significantly more saccades, consistent with greater visual demand [20, 21]. Fluctuation durations were shortest for the bar format, though differences across formats were not statistically significant. These findings informed its implementation in the MR-SLTS system. In the subsequent training study (Study 2), post-training assessment (without active feedback) showed that subjects who had trained with bar graph torque feedback applied lower z-torque, remaining further from the expert-defined safety threshold, with no statistically significant difference in procedure

duration between groups. Lower percentage errors were also observed across other force and torque dimensions, suggesting that the training effect generalized beyond the single feedback axis. Eye-tracking data, analyzed across all trials, indicated that the feedback group exhibited more frequent saccadic shifts than the control group, consistent with active engagement with the holographic display during the training stages [20, 21]. Taken together, the findings provide preliminary evidence that MR-based feedback support improved torque regulation that persisted after feedback withdrawal in simulation-based training settings. That said, the retention interval was short (approximately 5 minutes), and the number of unaided trials was limited; longitudinal studies incorporating delayed retention and transfer tests are needed to characterize the durability of this effect.

Beyond the retention question, several methodological limitations should be acknowledged. The sample sizes in both studies were small ($n = 6$ in Study 1; $n = 5$ per group in Study 2), which constrains statistical power and introduces susceptibility to individual variability. The present results are therefore best interpreted as pilot evidence that motivates larger-scale validation.

Study 1 employed a simplified single-axis force-tracking task with a 3D-printed prototype, which captures one dimension of the intubation task but not anatomical variability, multi-axis force demands, or concurrent procedural tasks. Further evaluation across tasks of increasing fidelity is warranted.

The MR-SLTS system focused exclusively on z-torque feedback, and the present findings therefore pertain specifically to MR-based torque regulation. Although effective, this single-axis feedback may not capture the full spectrum of trainee performance. Moreover, the expert-defined torque threshold was derived from three surgeons in prior work [5], and future work should establish benchmarks from a larger and more diverse clinician panel.

The eye-tracking analysis relied solely on saccade frequency as a behavioral proxy for visual demand. Recent work has shown that neurophysiological indices can detect workload changes that oculomotor metrics miss or detect only with delay [20, 21], and future studies should incorporate complementary measures such as subjective workload scales or EEG-based indices.

More broadly, expanding to multidimensional feedback, including x- and y-forces and integrating haptic or animated cues, could provide more comprehensive guidance. Machine learning algorithms could also enable adaptive and personalized feedback strategies. Future work should explore the combination of electromyography, motion tracking, and expert modeling to allow performance evaluations. Larger clinical trials will be essential for validating the system's readiness for real-world deployment and for benchmarking its educational impact across different learner populations.

7 | Conclusion

This study introduced and evaluated the MR-SLTS, a training platform that delivers real-time quantitative feedback through HoloLens 2. The system integrates graphic visualization, low-latency wireless communication, and eye tracking to support sensorimotor skill development during pediatric ETI training.

In Study 1 ($n = 6$), bar-graph feedback yielded the lowest tracking error and comparable saccade behavior to the line graph, while the ring graph elicited significantly more saccades. In Study 2 ($n = 10$), novices receiving real-time z-torque feedback via the selected bar-graph format maintained torque values closer to expert-defined limits without increasing procedure duration. As a pilot study with limited sample sizes and no delayed retention test, these findings provide initial evidence that head-mounted torque feedback is feasible for simulated intubation training and can influence force regulation during active guidance. Larger-scale validation with longitudinal retention protocols is needed to establish whether these effects reflect durable skill acquisition.

Author Contributions

Jiaqi Li: conceptualization (lead), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), software (lead), validation (lead), visualization (lead), writing – original draft (lead). **Ryman Hashem**: formal analysis (equal), validation (equal), visualization (equal), writing – review and editing (equal). **Ningzhe Hou**: data curation (supporting), methodology (supporting). **Perla Maiolino**: methodology (supporting), supervision (supporting). **Louis Halamek**: conceptualization (equal), investigation (equal), methodology (equal), supervision (supporting), writing – review and editing (supporting). **Liang He**: conceptualization (lead), funding acquisition (lead), investigation (lead), methodology (lead), project administration (lead), resources (lead), supervision (lead), writing – review and editing (lead).

Acknowledgments

This work was funded by Royal Society Research grant No. RGS/R2/242066. Dr. Louis P. Halamek's time was supported by the Endowment for the Center for Advanced Pediatric and Perinatal Education at Packard Children's Hospital at Stanford.

Funding

This work was funded by Royal Society Research grant No. RGS/R2/242066.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. T. Rahman, S. Chandran, D. Kluger, et al., "Tracking Manikin Tracheal Intubation Using Motion Analysis," *Pediatric Emergency Care* 27, no. 8 (2011): 701–705, https://journals.lww.com/pec-online/fulltext/2011/08000/tracking_manikin_tracheal_intubation_using_motion.4.aspx.

2. J. N. Carlson, S. Das, F. D. L. Torre, C. W. Callaway, P. E. Phrampus, and J. Hodgins, "Motion Capture Measures Variability in Laryngoscopic Movement during Endotracheal Intubation: A Preliminary Report," *Simulation in Healthcare* 7 (2012): 255, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3412923/.

3. A. C. Alvarado and P. Panakos, *Endotracheal Tube Intubation Techniques*. (StatPearls, 2023), https://doi.org/7;https://www.ncbi.nlm.nih.gov/books/NBK560730/.

4. A. G. Miller, "Endotracheal Intubation Training and Skill Maintenance for Respiratory Therapists," *Respiratory Care* 62, no. 2 (2017): 156–162, https://rc.rcjournal.com/content/62/2/156-https://rc.rcjournal.com/content/62/2/156.abstract.

5. N. Hou, L. He, A. Albini, L. Halamek, and P. Maiolino, "The Design of a Sensorized Laryngoscope Training System for Pediatric Intubation," in 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Press, 2024), 2414–2420, https://ieeexplore.ieee.org/document/10801567/.

6. H. Zhou, S. Yang, L. Halamek, and T. Nanayakkara, "A Method to Use Haptic Feedback of Laryngoscope Force Vector for Endotracheal Intubation Training," in Proceedings - IEEE International Conference on Robotics and Automation 2023 (IEEE Press, 2023), 6810–6816.

7. A. Covelli, S. Bardelli, R. T. Scaramuzzo, et al., "Effectiveness of a New Sensorized Videolaryngoscope for Retraining on Neonatal Intubation in Simulation Environment," *Italian Journal of Pediatrics* 46, no, 1 (2020): 13, https://ijponline.biomedcentral.com/articles/10.1186/s13052-020-0774-z.

8. Y. Noh, C. Wang, M. Tokumoto, et al., "Development of the airway Management Training System WKA-5: Improvement of mechanical designs for high-fidelity patient simulation," in IEEE International Conference on Robotics and Biomimetics, ROBIO, (2012), 1224–1229,

9. P. Maiolino, M. Maggiali, G. Cannata, G. Metta, and L. Natale, "A Flexible and Robust Large Scale Capacitive Tactile System for Robots," *IEEE Sensors Journal* 13 (2013): 3910–3917.

10. K. Cleary and T. M. Peters, "Image-Guided Interventions: Technology Review and Clinical Applications," *Annual Review of Biomedical Engineering* 12 (2010): 119–142, https://www.annualreviews.org/content/journals/10.1146/annurev-bioeng-070909-105249.

11. Y. Zou, Y. Chen, M. Gao, et al., "Coronary Heart Disease Preoperative Gesture Interactive Diagnostic System Based on Augmented Reality," *Journal of Medical Systems* 41, no. 8 (2017): 126.

12. S. Eom, S. Kim, S. Rahimpour, and M. Gorlatova, "AR-Assisted Surgical Guidance System for Ventriculostomy," in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), (IEEE Computer Society Conference Publishing Services (CPS), 2022), 402–405.

13. M. Herrlich, P. Tavakol, D. Black, et al., "Instrument-Mounted Displays for Reducing Cognitive Load during Surgical Navigation," *International Journal of Computer Assisted Radiology and Surgery* 12, no. 9 (2017): 1599–1605.

14. F. Paas, A. Tricot, K. Ouwehand, F. Lespiau, G. Mugisha, and A. Arguel, "Procedural Learning in Mixed Reality: Assessing Cognitive Load and Performance," *Education Sciences* 15 (2025): 339, https://www.mdpi.com/2227-7102/15/3/339.

15. V. Ronca, F. Babiloni, and P. Aricò, "A Novel Mutual Information-Based Approach for Neurophysiological Characterization of Sense of Presence in Virtual Reality," *IEEE Transactions on Biomedical Engineering* 72 (2025): 2313–2320, https://ieeexplore.ieee.org/abstract/document/10884049.

16. N. Bhavsar, S. Sriram, S. Balasubramanian, et al., "AIRWAY-XR: Augmented Instruction to Refine Wayfinding and Yielding Skills in Emergency Medicine Residents for Intubation Using Mixed Reality Technology,"Annals of Emergency Medicine 82 (2025): S113-S114. http://medrxiv.org/lookup/doi/10.1101/2025.01.06.24319788.

17. F. C. S. Junior, D. Litchfield, J. Sandars, and D. Cecilio-Fernandes, "Use of Eye Tracking in Medical Education," *Medical Teacher* 46, no. 11 (2024): 1502–1509, https://doi.org/10.1523/ENEURO.0196-18.2019.

18. O. Nachmani, J. Coutinho, A. Z. Khan, P. Lefèvre, and G. Blohm, "Predicted Position Error Triggers Catch-Up Saccades during Sustained Smooth Pursuit," *ENeuro* 7 (2020), https://www.eneuro.org/content/7/1/ENEURO.0196-18.2019.

19. A. Coudiere and F. R. Danion, "Eye-Hand Coordination All the Way: From Discrete to Continuous Hand Movements," *Journal of Neurophysiology* 131, no. 4 (2024): 652–667, https://doi.org/10.1152/jn.00314.2023.

20. A. Giorgi, G. Borghini, F. Colaiuda, et al., "Driving Fatigue Onset and Visual Attention: An Electroencephalography-Driven Analysis of Ocular Behavior in a Driving Simulation Task," *Behavioral Sciences* 14 (2024): 1090, https://www.mdpi.com/2076-328X/14/11/1090.

21. F. Dello Iacono, L. Guinti, M. Cecchetti, et al., "Analysis of Neurophysiological Correlates of Mental Fatigue in Both Monotonous and Demanding Driving Conditions," *Brain Sciences* 15, no. 9, 2025): 1001, https://www.mdpi.com/2076-3425/15/9/1001.

22. Y. B. Eisma, C. Borst, R. van Paassen, and J. de Winter, "Augmented Visual Feedback: Cure or Distraction?," *Human Factors* 63 (2021): 1156–1168, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8521352/.

23. V. Puttemans, S. Vangheluwe, N. Wenderoth, and S. P. Swinnen, "Bimanual Directional Interference: The Effect of Normal versus Augmented Visual Information Feedback on Learning and Transfer," *Motor Control* 8, no. 1 (2004): 33–50, https://journals.humankinetics.com/view/journals/mcj/8/1/article-p33.xml.

24. P. McWilliam, L. P. Halamek, B. King, et al., "Sensor-Integrated Laryngoscope: A Key Step Toward Safe Neonatal Intubation," *Sensor-Integrated Laryngoscope. Journal of Clinical Engineering* 45, no. 1 (2020): 39–44, https://journals.lww.com/10.1097/JCE.0000000000000386.

25. F. von Haxthausen, R. Moreta-Martinez, APD. de la Lastra, J. Pascau, and F. Ernst, "UltrARsound: In Situ Visualization of Live Ultrasound Images Using HoloLens 2," *International Journal of Computer Assisted Radiology and Surgery* 17, no. 11 (2022): 2081–2091.

26. T. Khan, T. S. Zhu, T. Downes, et al., "Understanding Effects of Visual Feedback Delay in AR on Fine Motor Surgical Tasks," *IEEE Transactions on Visualization and Computer Graphics* 29, no. 11 (2023): 4697–4707.

27. T. Honda, N. Hagura, T. Yoshioka, and H. Imamizu, "Imposed Visual Feedback Delay of an Action Changes Mass Perception Based on the Sensory Prediction Error," *Frontiers in Psychology* 4 (2013): 760, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3805955//pmc/articles/PMC3805955//pmc/articles/PMC3805955/?report=abstract.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.