Automated Assessment System with Cross Reality for Neonatal Endotracheal Intubation Training

Shang Zhao[1] Wei Li[1] Xiaoke Zhang[1] Xiao Xiao[1] Yan Meng[1] John Philbeck[1]
Naji Younes[1] Rehab Alahmadi[1] Lamia Soghier[2] James Hahn[1]

[1]George Washington University [2]National Children's Health Systems *

Abstract

Neonatal endotracheal intubation (ETI) is a resuscitation skill and therefore, requires an effective training regimen with acceptable success rates. However, current training regimen faces some challenges, such as the lack of visualization inside the manikin and quantification of performance, resulting in inaccurate guidance and highly variable manual assessment. We present a Cross Reality (XR) ETI simulation system which registers ETI training tools to their virtual counterparts. Thus, our system can capture all aspects of motions and visualize the entire procedure, offering instructors with sufficient information for assessment. A machine learning approach was developed to automatically evaluate the ETI performance for standardizing assessment protocols by using the performance parameters extracted from the motions and the scores from an expert rater. The classification accuracy of the machine learning algorithm is 83.5%.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Mixed / augmented reality; Machine Learning—Machine Learning algorithms—Feature selection;

1 Introduction

ETI is a time-sensitive resuscitation procedure essential for the ventilation of newborns. Typical neonatal resuscitation training programs mainly rely on practicing on task trainers or simulators under supervision so that trainees can gain some levels of proficiency before clinic exposure [1]. However, both instructors and trainees suffer the lack of situational awareness during training. The small sizes of the intubation space in neonatal models do not allow instructors to fully visualize the events occurring within the simulator to provide feedback and accurate assessment, leading the feedback mainly relies on the ETI outcome. Our ETI simulation system has potential to improve ETI training with a see-through visualization to provide sufficient information for guidance and an automated assessment for ETI performance evaluation to provide interpretable feedback, solving the resource-intensive issue in the ETI training.

2 Methodology

2.1 Cross Reality System

We developed a lightweight XR ETI simulation system that consists of a standard neonatal resuscitation manikin with a laryngoscope and an endotracheal tube, an Ascension trakStar electromagnetic tracking system, and a Hololens device. All the devices were registered into the same coordinate using the EM sensors of the trakStar system. In details, we used the EM sensor to select fiducials on the manikin

†*This work was supported by NIH grant R01HD091179.

‡1email: edwinz, gw_liwei, xkzhang, xxfall2012, mengy, philbeck@gwu.edu, naji@bsc.gwu.edu, raa53@gwu.edu, hahn@gwu.edu

§2email: lsoghier@childrensnational.org

Figure 1: The overview of our XR system setup in the data collection study (left). The top right figure shows a registered sagittal cross-section of the CT-scanned manikin in the Hololens; the bottom right figure shows the instant feedback with performance parameters.

to establish pairwise correspondences between the manikin and its CT scanned counterpart. For Hololens, we attached Aruco markers to the EM sensor to calibrate the initial pose of the Hololens by using the predefined patterns. Compared with optical tracking methods [3], the trakStar system allows 6 degrees of freedom motion tracking at 80Hz without occlusion issues, which is suitable for our intubation procedure. We attached the EM sensors to the manikin and the other components of the simulator to track their corresponding poses during the intubation procedure, and streamed the motion sequences back to a laptop, allowing for real-time and post-trial motion analysis. Moreover, we used UDP protocols to wirelessly transfer the captured data to the Hololens for superimposed visualization.

The traditional 2D videolaryngoscopic [2] used in the intubation training has difficulties in visualizing the internal anatomical structure of the manikin due to the occlusion from the tongue or the lighting conditions, such as the shadow and the specular highlights from the light source. This results in lower efficiency of guidance and training performance for both instructors and trainees. In contrast, in our XR system, both the manikin and the laryngoscope were displayed on the monitor and the Hololens in real-time using direct volume rendering with the cross-sectional view to provide more intuitive 3D information concerning the relative poses between the laryngoscope and the glottis.

During the training process, we extracted several performance parameters to provide users situational awareness (See Sec. 2.3). The parameters were showed in place overlaying on the manikin to avoid interrupting the training procedures from the monitor. This can provide instant feedback to the trainees, and avoid critical mistakes such as the over-penetration of the laryngoscope to the upper gum. An overall score will be evaluated by our automatic assessment system and shown to users after finishing the entire procedure. Users can watch the 3D playback with the reported performance parameters to improve their movements and gain proficiency.

2.2 Study Design

We collected the motions of 40 subjects including both attending neonatologists and residents by using our system. The diversity of subjects made the dataset contain different levels of performance and different strategies of neonatal ETI. The subjects performed

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.00218

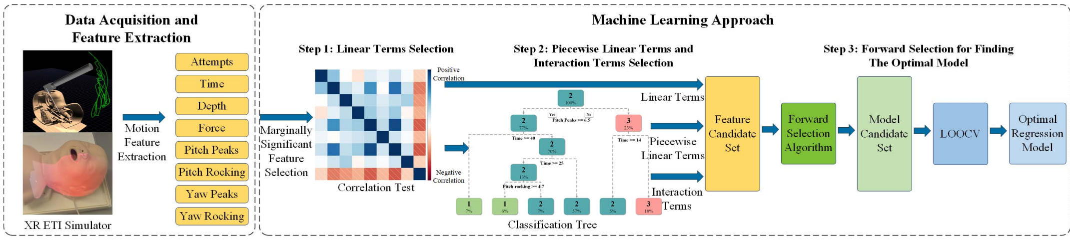

Figure 2: Pipeline of our automated assessment framework.

on our system which was set up on an infant warmer in the NICU (see Fig. 1). The study was approved by the Institutional Review Board of the Children's National Health Systems. Each participant performed 5 ETI trials on the system. Only the instructor was able to see the XR contexts during the procedure.

2.3 Performance Evaluation: A Machine Learning Approach

In this section, we describe our automated assessment system for evaluating ETI performance in details (see Fig. 2). We aim to select important performance parameters (features) based on which we can build an interpretable and predictive model for the overall score of ETI performances. Our system extracted 8 primary performance parameters for training automated assessment model and quantifying the performance from raw motions. These parameters were chosen *apriori* based on the qualities that the expert instructor deemed as important features so that the correlation between parameters and training rubric can be preserved. The process of finding the optimal model is composed of 3 steps:

Step 1: We select performance features which have statistically significant marginal associations (P-value < 0.05) with the overall score by fitting the multinomial regression model for the score, an ordinal response y. Each multinomial regression model was fitted by the multinomial generalized estimating equation method (GEE).

Step 2: We construct a classification tree for the overall performance score on the selected features from Step 1. The classification tree can identify predictive features hierarchically - the closer to the root, the more predictive - and their optimal splits. The feature candidate set for finding the optimal model comprises of linear terms (the marginally significant features which did not appear in the classification tree), piecewise linear terms (the predictive features in the classification tree), and their two-way interaction terms.

Step 3: Multinomial regression was applied in the forward selection algorithm to find the candidate models for predicting the overall score. In detail, we first initialized the model candidate set $M = \phi$ which is the candidate set for choosing the optimal model. The reference model m_i is defined for the feature selection in each iteration i, which was initialized to a model with intercept terms only. The loop for feature selection was executed after initialization. The number of loops was determined by the size of the feature candidate set. We did the following substeps in each iteration i : 1) We constructed the model c_i by adding one additional candidate feature term that not included in m_i. 2) We evaluated P-value by the Wald test between c_i and m_i. 3) We determined the optimal model c_i^* in the current step, which has the minimum P-value of the Wald test among all possible c_i models. 4) If the P-value of the optimal model was larger than 0.05 or the model introduced numerical errors, then we broke this loop procedure. Otherwise, we set the new reference model m_{i+1} as the optimal model c_i^* of current step and added c_i^* to the selected model set M. Note that we selected the optimal model c_i^* with the minimal P-value in the Wald test as the new reference model. Eventually, we found the optimal model that has the highest classification accuracy in the model candidate set M from Leave-One-Subject-Out-Cross-Validation (LOOCV).

3 Result

Based on the coefficient estimates obtained from Eq. (1), the probability $P(OverallScore = y)$, $j = 1,2,3$, for a new trial can be derived from the fitted values of y_j, and the predicted overall score corresponds to the one with the largest probability. The classification accuracy of our automated scoring system is 83.5%. The optimal multinomial regression model obtained from the forward selection algorithm in Step 3 is given as follows: For $j = 1,2$,

$$
\begin{aligned}
y_j = {} & \log\left\{\frac{P(OverallScore \le j)}{1 - P(OverallScore \le j)}\right\} \quad (1) \\
= {} & \beta_{0j} + \beta_1 \cdot F_a + \beta_2 \cdot F_d + \beta_3 \cdot F_{pp} \cdot I(F_{pp} \ge 6.5) \\
& + \beta_4 \cdot F_{pr} \cdot I(F_{pr} \ge 4.7) + \beta_5 \cdot F_t \cdot I(F_t < 14.0) \\
& + \beta_6 \cdot F_t \cdot I(14.0 \le F_t < 40.0) + \beta_7 \cdot F_t \cdot I(F_t \ge 40.0) \\
& + \beta_8 \cdot F_{pp} \cdot I(F_{pp} \ge 6.5) \cdot F_t \cdot I(F_t \ge 40.0) \\
& + \beta_9 \cdot F_{pp} \cdot I(F_{pp} \ge 6.5) \cdot F_t \cdot I(14.0 \le F_t < 40.0) \\
& + \beta_{10} \cdot F_{pr} \cdot I(F_{pr} \ge 4.7) \cdot F_d,
\end{aligned}
$$

Table 1: The coefficients of the optimal feature set that can predict the overall performance score.

Coefficient	β_{01}	β_{02}	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Estimate	-18.366	-12.203	1.417	0.128	0.224	0.256	0.165	0.253	0.088	0.013	-0.007	-0.003
Standard Deviation	5.129	4.567	0.696	0.054	0.115	0.444	0.186	0.097	0.047	0.009	0.004	0.007
***P*-Value**	<0.001	0.008	0.042	0.018	0.050	0.560	0.374	0.009	0.060	0.165	0.114	0.682

where F_a, F_d, F_{pp} and F_{pr} represent the feature "Attempts", "Depth", "Pitch peaks", and "Pitch rocking", respectively. $I(\cdot)$ is the indicator function needed to incorporate piecewise linear terms, β_{0j} ($j = 1,2$) are the category-specific intercepts and $\beta_k, j = 1,\ldots,10$ are coefficients. Note that $P(OverallScore \le j) = e^{y_j}/(1+e^{y_j})$ is an increasing function of y_j, so a larger fitted y_j indicates a higher probability of achieving a lower overall score. The coefficient estimates are given in Table 1. With considering the significant terms, we can conclude that 1) The overall score is generally negatively associated with a subject's attempts and depth; 2) Subjects' pitch peaks feature is also generally negatively associated with the overall score if they are beyond 6.5; 3) The overall performance score is in general better with a shorter time, but it is significantly negatively associated with the overall score when the time is between 14.0 seconds and 40.0 seconds. This coincides with our assumption that the selected performance features in the optimal assessment model are correlated with the motion characteristics that instructors consider important in training assessment.

4 Conclusion

We proposed an XR ETI system that provides complete visualization and automated scoring for training pediatric trainees. In the future, the machine learning approach will learn from a panel of expert instructors and larger dataset to yield a more robust model.

References

[1] J. J. Klotz and *et al.* Pediatric and neonatal intubation training gap analysis: instruction, assessment, and technology. *Simulation in Healthcare*, 9(6):377–383, 2014.

[2] M. Qian and *et al.* Augmented reality (ar) assisted laryngoscopy for endotracheal intubation training. In *International Conference on Human-Computer Interaction*, pp. 355–371. Springer, 2019.

[3] J. Rolland and *et al.* Development of a training tool for endotracheal intubation: Distributed augmented reality. 2003.