COMMUNICATION FEATURES ASSOCIATED WITH CLINICAL PERFORMANCE AND NON-TECHNICAL SKILLS IN HEALTHCARE SETTINGS

by

Yuhao Peng

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science in Industrial Engineering

School of Industrial Engineering

West Lafayette, Indiana

December 2019

THE PUEDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Denny Yu, Chair
School of Industrial Engineering
Dr. Brandon Pitts
School of Industrial Engineering
Dr. Felicia Roberts
School of Communication

Approve by:
Dr. Abhijit Deshmukh

Dedicated to my beloved parents, and my little brother.

ACKNOWLEDGMENTS

Reflecting on the completion of this work, I would like to thank Dr. Felicia Roberts and Dr. Brandon Pitts for their time spent on my work, and invaluable feedback.

I would also like to thank my academic advisor, Dr. Denny Yu, for providing guidance and advice through every stage of this project, always pushing my work forwards with high standards and expectations. His comments at every step motivated me to strive for the highest quality in my research study.

Lastly, I like to thank Jackie, Hamed, and Chuhao who have also supported me for my research and bringing me such great joy in my educational journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACTS	Acute Care Trauma Simulation
CLC	Closed-loop Communication
NTS	Non-technical Skills
VAS	Visual Analogue Scale

ABSTRACT

Author: Peng, Yuhao. MSIE
Institution: Purdue University
Degree Received: December 2019
Title: Communication Features Associated With Clinical Performance and Non-Technical Skills in Healthcare Settings
Committee Chair: Denny Yu

Effective teamwork and communication are critical to patient outcomes, and subjective assessment tools have been developed for measuring team performance using both technical and non-technical skills. However, inherent biases remain with using subjective assessment tools. The objective of this thesis is to investigate the relationships between objective communication measures (e.g., speech duration, ratio, rate, etc.) and healthcare providers' clinical performance and Non-technical Skills (NTS) performance in simulated trauma care team scenarios.

In this study, 3rd-year medical students participated in the Acute Care Trauma Simulation (ACTS). The student performed the role of clinician in a team that included a nurse and a simulated patient. Participants conducted post-operative patient management, patient care diagnoses, and treatment. Audio from all team members was recorded, and speech variables (e.g., speech duration, frequency of interaction, etc.) from student's audio were extracted. For Research Question I, correlation and regression models were used to explore the relationships between vocal features and clinical performance; for Research Question II, additional vocal features were extracted from audio recordings, and these features were used to developed multiple regression models relating vocal features with NTS overall scores and with the communication construct of the NTS score.

Findings showed that a majority (67%) of the communications were initiated by the student. Speech ratio, intensity, and frequency of communications differed when students communicate with the nurse than with the patient (e.g., communication from student to the patient resulted in a higher intensity). The models for Research Question I showed that increasing frequency of checkbacks between student and nurse ($p<0.05$) and speech duration from student to patient

(p=0.001) significantly increased student's clinical performance score. In Research Question II, a positive association ($\rho=0.456$, $p<0.001$) between speech duration from student to patient and overall NTS scores was observed, and this correlation was the strongest amongst all other vocal features with overall NTS score. The forward stepwise regression model predicted overall NT skills scores with adjusted R-squared value of 0.537. Similarly, the forward stepwise regression model predicted communication construct with adjusted R-squared value of 0.54.

Both studies showed significant positive relationships between key vocal features (e.g., speech duration), frequency of communication with respect to performance. Metrics and vocal features derived from audio recordings can be measured in predicting clinical performance and NTS, moreover, it can further contribute to the understanding of communication in the healthcare setting. Most importantly, the potential of providing an objective approach for simulation-based trauma care training.

1. INTRODUCTION

1.1 Communication in Healthcare

Effective teamwork in healthcare is a known contributor to positive patient outcomes (Baker D.P., Gustafson S., Beaubien J.M., Salas E., 2005; Linda T. Kohn, Janet M. Corrigan, 1999; Manser, 2009; Martin, Ummenhofer, Manser, & Spirig, 2010; Tiel Groenestege-Kreb, Van Maarseveen, & Leenen, 2014). Communication contributes to building cohesive healthcare teams and also influences patient care (Ellingson, 2002; Grumbach & Bodenheimer, 2004; Schmutz & Manser, 2013). Studies on healthcare teams show that teams with frequent communication among team members achieve higher productivity than teams with less frequent communication (Reagans & Zuckerman, 2001). According to the Joint Commission, communication failures are one of the most frequently identified root causes (30%) of sentinel events (Commission, 2015; CRICO Strategies., 2015). In six Danish hospitals, 52% of 84 root cause analysis reports cite verbal communication errors (e.g., misinterpreted information between physicians, especially during handoffs), which provides further evidence that effective communication is a major factor contributing to patient safety (Rabøl et al., 2011). Furthermore, when deconstructing communication failures over 22 hours of audio recorded during six high-acuity surgical procedures, researchers found that communication failures occurred once every eight minutes (Hu et al., 2012). Among all communication failures, 22% of communication failures were nurse initiated or anesthesiology-directed communication. The same study found that communication failures also varied depending on the phase of the surgery, with 54% of errors occurring between incision and closure. Moreover, the importance of communication in safety is not just limited to healthcare, but has been frequently shown important in other safety critical domains, such as aviation, military, etc. (Anca, Helmreich, & Kanki, 2010; Helmreich & Foushee, 2010; Leonard, Graham, & Bonacum, 2004; Sexton & Helmreich, 2000).

Communication is a well-established area of study. Generally defined as the behavior of reducing uncertainty from one individual to another, it occurs whenever there is a need for information processing (Buck & VanLear, 2002; N. J. Cooke, Gorman, & Kiekel, 2008; Hackman, 1987). In hospital settings like the emergency department (ED) or clinical ward, communication can appear in many different forms and modalities. Some examples include verbal (e.g., face-to-face, telephone, radio), nonverbal (e.g., gestures), and literal (e.g., writing, reading) (Coiera & Tombs, 1998; Eisenberg et al., 2005). In a study classifying information-sharing activities among 38 clinicians during 159 patient encounters across seven sites, verbal communication (i.e., with patient, staff, or colleague) contributed to 60% of patient encounter time, the remaining time included writing, reading, examination, and other (Tang et al., 1996). Communication patterns also differ whether the target audience is a healthcare professional or patient. Healthcare professionals often use both medical language and everyday language among themselves and patients; however, everyday language is the most frequently-used form when communicating between healthcare professionals and patients (Bourhis, Roth, & MacQueen, 1989). In addition to varying audiences, communication patterns may also change according to circumstances at the moment, including time, safety and sterility, resources, roles, and situation (Lingard, Reznick, Espin, Regehr, & DeVito, 2002).

1.2 Clinical Performance

Understanding and monitoring communication can offer insight into healthcare providers' abilities to deliver safe and effective patient care. Several tools and measurement constructs have been proposed to quantify communication, identify patterns that impact outcomes, and guide team training in healthcare. For example, the Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS) is a tool that has been widely used to improve team performance and has demonstrated positive effects on patient outcomes (King et al., 2008; Mayer et al., 2011; Weaver et al., 2010). TeamSTEPPS is a comprehensive multi-phase teamwork intervention that focuses on leadership, situation monitoring, mutual support, communication, and communication is one of the TeamSTEPPS competencies (Clapper & Kong, 2012). Drawing from TeamSTEPPS communication model, the TeamSTEPPS Teamwork Attitudes Questionnaire (TAQ) has been developed as a self-assessment tool to assess the impact of the training on student learning (Brock et al., 2013; "Teamwork Attitudes Questionnaire (T-TAQ) | Agency for Healthcare Research & Quality," n.d.). The TAQ consists of five Likert-type dimensions: Team Structure, Leadership, Situational Awareness, Mutual Support and Communication. Many other tools are also available for assessing team communications in individual or team-based settings. The Communication Assessment Tool-Team (CAT-T) is designed to assess team communications; however, the tool similarly relies on subjective evaluation (i.e., 5-point response scale, from 1=poor to 5=excellent) (Eisenberg et al., 2005; Fairbanks, Bisantz, & Sunm, 2007; Mercer et al., 2008). Communication assessment is also emphasized in the area of interprofessional non-technical skills (Cha et al., 2019; Flin, O'Connor, & Crichton, 2007; Hull, Arora, Kassab, Kneebone, & Sevdalis, 2011). Similar to the previous tools, non-technical skills assessment tools assess communication with behavioral anchors used by expert raters to reference points. However, like other tools, a key limitation is the dependency on subjective Likert scales (Hull et al., 2012; Yule, Flin, Paterson-Brown, Maran, & Rowley, 2006).

1.3 Non-technical Skills

Social, behavioral, and cognitive centered skills, also known as non-technical (NT) skills have been well identified as critical factors for surgical team performance and patient safety in the operating room (Cha et al., 2019; Yule, Flin, Paterson-Brown, & Maran, 2006). Although it is important to demonstrate advanced knowledge and skills to enhance patient safety, poor non-technical skills are also a major contributor to error in healthcare (Odell, 2011). 82% of 359 preventable anesthetic incidents reported were caused by human errors, with the top three errors including inadequate familiarity with equipment, communication failure, and haste (Cooper, Newbower, Long, & McPeek, 1978). The concept of NT skill development has been widely discussed and incorporated into training in other fields, such as aviation, to reduce human errors (Crichton & Flin, 2004; Flin et al., 2019).

NT skills consists of five constructs, situational awareness (the right information is needed at the right time), decision-making (the process of option selection), communication, teamwork, and leadership (M. R. Endsley, 1995; Flin, O'Connor, & Crichton, 2013; Toner, 2009). Situational awareness is important not only to each individual or as a team, but also one of the frequent causes of error occurrence during real time tasks (M. Endsley, 1995; Gugerty, 1997). In a study in nursing, poor situational awareness can lead to critical risks in patient outcomes, including medication errors and patient misidentification (Fore & Sculli, 2013; Turkelson, Aebersold, & Sculli, 2013). For decision-making in healthcare settings, it is ranked as the most important NT skills and personality trait for a surgical trainee (Cuschieri, Francis, Crosby, & Hanna, 2001; Jacklin, Sevdalis, Darzi, & Vincent, 2008; Morozova, Martindale, & Currie, 2017). Lastly, importance of communication, teamwork, and leadership have also been significantly studied from previous researches (Baker D.P., Gustafson S., Beaubien J.M., Salas E., 2005; Linda T. Kohn, Janet M. Corrigan, 1999; Manser, 2009; Martin et al., 2010).

To measure and capture these behaviors, several NT skills assessment tools have been developed, and the application of these assessment tools vary due to different team roles and specialties. For example, Non-Technical Skills (NOTECHS) originated from aviation industry was now adapted and commonly used in surgery; Non-Technical Skills for Surgeons (NOTSS) concentrates on the individual surgeons' intra-operative non-technical skills; and similarly, Observational Teamwork Assessment for Surgery (OTAS) assesses the surgical team as a whole in the operating room (Flin et al., 2019; Hull et al., 2011; Sharma, Mishra, Aggarwal, & Grantcharov, 2011; Yule, Flin, Paterson-Brown, Maran, et al., 2006). However, these assessment tools are observer based and require trained evaluators, there's need to identify measurable features that can automatically access these NT skills in healthcare settings.

Communication, which has generally been defined as the exchange of information through either verbal or non-verbal means, can potentially be a method to assess NT skills objectively and continuously. Emphasis on communication is found in a number of previously-mentioned assessments. Due to its relationship with all NTS constructs, strong relationships were observed between communication and each construct (Sharma et al., 2011). For example, in a study of effective leadership skills in increasing team performance, Feese et al. observed that considerate leaders spoke with shorter utterances and with more changes in speech loudness. They observed a positive relationship between such features (e.g. fundamental frequency, F0, and speech intensity) and cognitive load (situational awareness) in a military aviation simulator task. Other researchers also found the changes in prosodic features could be affected by cognitive loads in military aviation (Huttunen, Keränen, Väyrynen, Pääkkönen, & Leino, 2011; Tolkmitt & Scherer, 1986; van Reekum et al., 2004). Moorthy et al. measured surgical trainees' communication as the role of surgeon in a simulated operating room. The NT skills of communication-construct concentrated on the measures of trainees and the rest of team members including: 1) politeness, 2) acknowledgements from team, 3) assistance sought from team members, and 4) two-way communication count (Utterance Frequency) between trainee, anesthesiologist, assistant, and scrub nurse. Low scores were frequently observed in patient positioning, waiting for swab count, and informing anesthetist prior to closure (Moorthy, Munz, Adams, Pandey, & Darzi, 2005).

Gittell et al. measured communication and relationships among healthcare providers and found frequency of communication was one of the key factors associated with quality of care improvement, reduced post-operative pain, improved post-operative functioning, and decreased lengths of hospital stay (Gittell et al., 2000). Despite exciting exploration into these objective communication metrics, the application in healthcare is limited due to complex interaction between healthcare team, change of communication styles during high stress environment, difficulties of collecting and monitoring audio of each team member, and the extensive time required to analyze and derive these metrics from audio (Ryan et al., 2019). Furthermore, each individual healthcare provider's communication contains different performance patterns, (e.g. variations on leadership and coping style, team process, etc.) and such variations could result in different (positive and negative) patient outcomes (Howlett et al., 2015; Somech, 2006; Wong, Cummings, & Ducharme, 2013). To overcome these limitations and better understand the relationship between vocal features and clinical skills, we used a controlled training simulations to assess how communication patterns varied with clinical skills among medical students.

1.4 Communication and Hesitation

The importance between communication and clinical performance is well-recognized; however, an individual's performance is also highly dependent on personal knowledge and prior experiences. The challenge is to model the covariate relationship between communication and one's knowledge level. In a simulation study on nursing students' knowledge of advanced cardiac life support, Tawalbeh found that the number of simulation practices can improve student's clinical skills significantly, and most importantly, the number of practices increased student's knowledge background and resulted in a positive effect on students' self-confidence when conducting similar simulations (Tawalbeh & Tubaishat, 2014). This finding was consistent with other studies (Richards, Simpson, Aaltonen, Krebs, & Davis, 2010; Y. K. Scherer, Bruce, & Runkawatt, 2007; Tiffen, Graf, & Corbridge, 2009). To identify one's self-confidence or hesitation during performance assessments objectively, there are limited studies that evaluate these assessments through speech analysis. Abbas et.al, categorized pauses into two types, silent pauses and filled pauses, although they found these pauses do not exist haphazardly in the process of speech production; they depend on the context in which they occur (Abbas, Jawad, & Muhi, 2018). Furthermore, long pauses and fillers often used by speakers tend to indicate their responses are lacking confidence or providing a wrong answer, or indicate the speakers do know the answer but incapable of retrieving it (Brennan & Williams, 1995). Despite all the studies on the relationships between pauses and hesitation, it is still uncertain and an area to be further researched on. Future work may also consider communication metrics such as, speech intensity, pitch level, other than speech disfluency, which can further contribute to the understanding of the relationship between communication and hesitation, and ultimately predicting clinical performance.

1.5 Research Gap

Subjective rating tools are limited by several inherent biases. In medical education, systemic rater errors have biased observer ratings due to halo, severity, central tendency, leniency, logical error, inattention, restriction of the range, etc. (Downing, 2005). For example, subjective evaluation by raters tended to result in a positive/leniency bias, meaning higher scores were given to persons even they performed poorly; low severity on evaluations from expert to novice or vice versa; and serious shortcomings were usually ignored (Albanese, 2000). In addition, common in all the aforementioned tools for assessing communication are limitations in reliability and scalability due to the need for trained experts to perform the evaluations (Albrecht, 1996). Due to the recognized impact of communication on patient care, and limitations of the currently available subjective assessment tools, there is a need to identify objective and scalable approaches to assess and evaluate healthcare providers' communication.

Sensing-based approaches may overcome some limitations of current assessment tools by providing objective and potentially automated communication assessments. Several studies have used sensors to measure communication in team-based interpersonal interactions (Onnela, Waber, Pentland, Schnorf, & Lazer, 2014). Audio sensors have been used to infer communication patterns, nurses' personality, and team-workload distribution (Olguin, Gloor, Pentland, & Olguin Olguin, 2009; Yu, Blocker, Hallbeck, Patel, & Pasupathy, 2015; Yu et al., 2016). Previous work has shown that continuous audio streams can estimate social interaction patterns through wearable devices (Onnela et al., 2014), and techniques have been proposed (e.g., Latent Semantic Analysis) to assess team communication by transforming raw speech into textual input (N. J. Cooke et al., 2008; Gorman, Foltz, Kiekel, Martin, & Cooke, 2003). One of the most common uses of audio sensors in healthcare is for content-analysis to complement field observations (Xiao, Seagull, Mackenzie, Ziegert, & Klein, 2003). These studies capture audio to measure the usage of content-based communication events, distinguish between verbal or non-verbal communication, and quantify frequency and duration of communication events. Another area where communication skills can be assessed at a much lower cost is through the application of artificial intelligence (or machine learning), for example, Cogito Corp, a spin-out from Massachusetts Institute of Technology, already developed an application that provides call centers with real-time voice analysis (speech rate, intensity, etc.) and performance feedback, employees working at the customer service are able to receive targeted suggestions on communication skills improvement (Ryan et al., 2019; "Watch your tone | MIT News," n.d.). However, many of these approaches still rely on an analyst to code the data, and few studies have investigated the ability of audio metrics that can be automatically processed from audio recordings (e.g., speech intensity, duration, rate, etc.) to determine performance.

2. RESEARCH QUESTION I: MEASURES OF CLINICAL PERFORMANCE

2.1 Research Framework

The first research question was focusing on identifying more objective and continuous assessments of performance in the clinical environment using audio sensors that can ultimately facilitate targeted performance-enhancing interventions. Towards this goal, this study aimed to:

1. Identify vocal features and patterns that can be derived from audio data-streams and represent healthcare provider's communication behavior and,
2. Assess whether objectively obtained vocal features are associated with clinical performance.

2.2 Methodology

2.2.1 Study Participants

This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at Indiana University (IRB: #1611105172). Informed consent was obtained from each participant. Participants were recruited from one academic medical institution, and each participant provided informed consent to allow for audio and video recording of scenario-based simulations. These Acute Care Team Simulations (ACTS) served as a summative assessment of medical student performance after they had completed their general surgery clerkship rotation.

2.2.2 Study Procedure

Each ACTS scenario (Figure 2.1) took place in a state-of-the-art simulation center located at Fairbanks Hall, Indiana University School of Medicine. This simulation center was designed to educate multidisciplinary healthcare providers through the replication of multiple facets of the patient care environment including a fully functional operating room environment and intensive care unit rooms that can be controlled entirely by a simulation technologist using a computer system. This design allowed healthcare providers to engage in immersive simulated patient care scenarios and learn safe practices for actual patient care.

Figure 2.1 Medical students during their ACTS scenario.

During ACTS, the manikin features pulse and breathing sounds among other physiological features, which can be manipulated according to the details of the simulation scenario. Each ACTS session involved a team consisting of one medical student (i.e., all students were on their surgery clerkship rotation during their 3rd –year of medical school), a nurse confederate, and a simulation technologist controlling a manikin (SimMan 3G, Laerdal Medical, Wappingers Falls, NY). The technologist acted as the participant and his voice was heard through the manikin. Students were randomized into one of six scenarios: 1) Motor vehicle accident shock, 2) Pneumothorax, 3) Hyponatremia, 4) Leg compartment syndrome, 5) Pulmonary embolism, and 6) Heparin-Induced Thrombocytopenia.

These six scenarios captured the range of care management from patient arrival to the ED, post-operative patient management, patient care diagnoses, and treatment. Each scenario averaged approximately 10 minutes in length. Typical tasks performed by the student in every scenario included: completing a patient assessment, determining a diagnosis, and identifying an appropriate treatment. Students were in charge of the patient's care, and they were expected to communicate with both the simulated patient and nurse to deliver correct patient care safely and effectively. Based on the student's actions, a simulation technologist manipulated the patient's health (i.e., improve, remain the same, or worsen) under the guidance of an experienced faculty observer. Students' performances were rated at the completion of the scenario by the same nurse confederate, who had extensive experience evaluating students' performance during ACTS assessments.

2.2.3 Data Collection

Audio from each healthcare team member and patient was recorded throughout the scenario. Voice recorders (Zoom H1, Zoom, Inc, Hauppauge, NY, USA) were placed in the participant's pocket, and a lapel microphone (RØDE smartLav+ Microphone, RØDE Microphones, Silver Water, NSW, Australia) was attached to the scrub or jacket collar of the student, nurse, and patient (Figure 2.2). Although more intrusive than audio recordings from a video recorder, this approach allowed better localization of audio source, noise cancelling, and <1-minute setup time. Video recordings with three room views (i.e., patient view, overhead view, and a view of patient vitals) were collected using the cameras (Panasonic WV-CS574, Panasonic, Kadoma, Osaka Prefecture, Japan) built into the simulation center (B-Line Medical, Washington, DC). These were used as needed to verify and interpret audio observations.

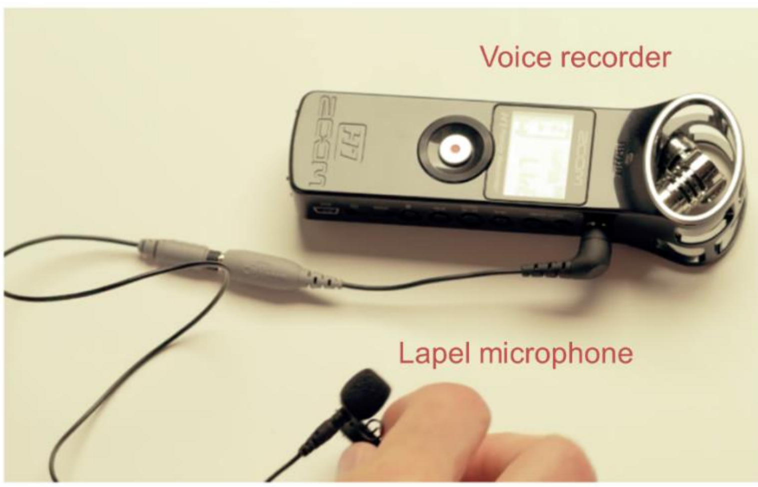


Figure 2.2 Non-intrusive voice recorder (top) with inserted lapel microphone (bottom)

2.2.4 Data Analysis

2.2.4.1 Communication Variables

Communication patterns were analyzed by a trained study researcher listening and annotating the recordings. The assessments focused on "check-backs", part of the closed-loop communication (CLC) strategy according to Härgestam et al. (2013) that consists of three elements: 1) call-out, 2) check-back, and 3) closed-loop. Take Figure 2.3 as an example, 1) person A (conversation initiator) transmits a message as either with a question or statement (call-out): What is one plus one? Then, person B (receiver) acknowledges the message with a response to the initiator (check-back): It's 2. Lastly, the response from person B should be verbally verified by person A to complete CLC: That's correct!

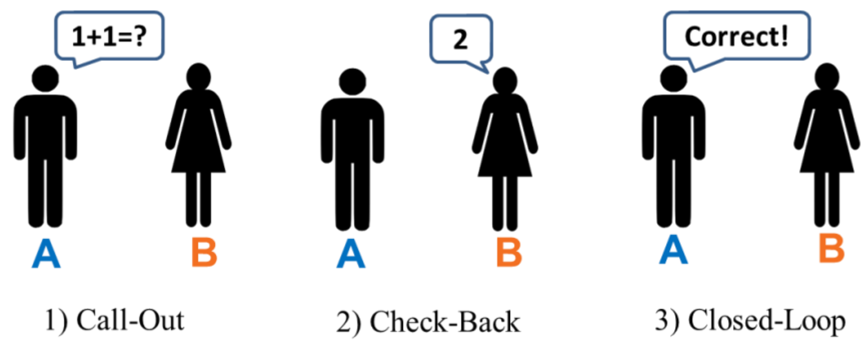


Figure 2.3. Example of a CLC: 1) Call-Out, 2) Check-Back and, 3) Closed-Loop

In this study, closed-loop communication was excluded because the principles of CLC were not part of the standard medical curriculum, hence they were not expected to utilize this form of specialized communication. Accordingly, only call-out and check-back were analyzed. For example, if person A transmitted a message to person B, and person B provided a response to A with a message, then a check-back was recorded. In another example, person A transmitted a message and person B responded with a question as follows:

> Person A: "Can I have …?"
>
> Person B: "Do you mean …?"

In this example, there was a verbal response from person B to A, and at the same time, person B transmitted a new message. Our study considered person B's response a check-back, and a call-out from person B to A. Thus, conversations in the present study were classified using the following categories:

1) Communication initiated with a question followed by a response,
2) Communication initiated with a statement (i.e., non-question) followed by a response,
3) Communication initiated with a question followed by no response, and
4) Communication initiated with a statement followed by no response.

In the ACTS scenarios, communication could originate from three potential sources (i.e., student, nurse, and patient). This resulted in a total of six combinations (i.e., only two-way communications were measured in this study): student to nurse, student to patient, nurse to student, nurse to patient, patient to student, and patient to nurse. Although rare, when there was three-way communication, for example both nurse and patient responding to student's message, it was classified by the content of whom the message initiator was speaking to. The number of check-backs, also known as the frequency of communications, were calculated from all six two-way combinations at the four different categories listed above. This analysis was completed by a research team member using custom Microsoft Excel software to annotate while listening to the audio recordings.

2.2.4.2 Audio Processing

In addition to decomposing the full audio files into dyad communication as mentioned previously, vocal features from the scenarios were also extracted using Praat (Boersma, Paul & Weenink, 2018) software. Since ACTS scenarios were focused on assessing students' clinical performance, this study concentrated on the student's vocal features and communication frequency with the patient and nurse. For this study, a dyad communication was defined by each change in students' communication target (nurse or patient). Figure 2.4 demonstrates an example of a series of three dyad communication from the student to the nurse and patient; the first dyad communication during the first 34 seconds was between the student and the nurse. The background noise of patient/nurse audio have been filtered. In each of the three- dyad communications, the following variables of students' vocal features were calculated: 1) speech duration: the total time the student spoke to nurse or patient, 2) speech ratio: the percentage of time students spoke during a communication event, 3) speech intensity: loudness, and 4) speech rate: speech speed or pace.

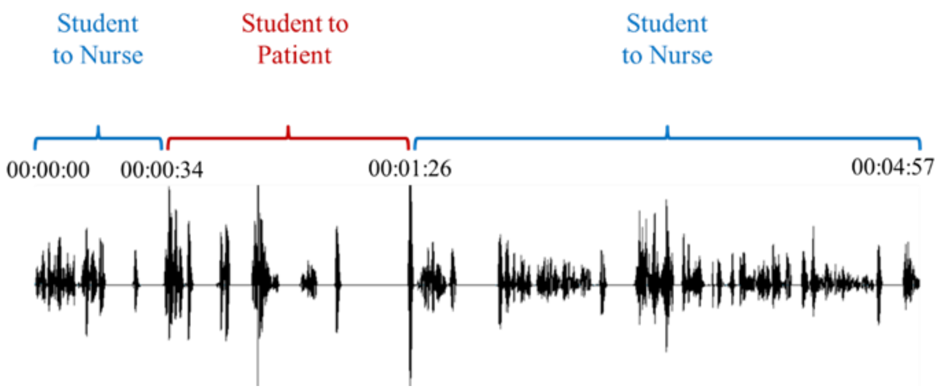
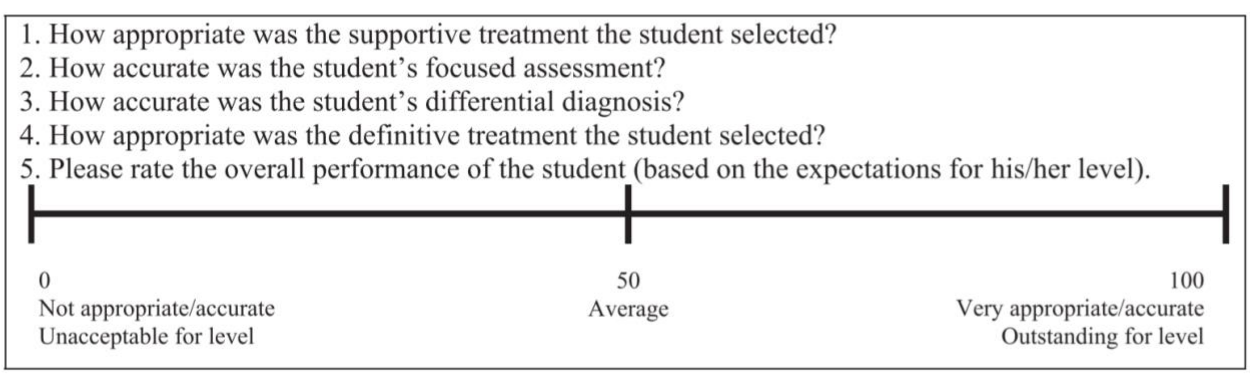


Figure 2.4. An example of student's audio and separated dyad communications

2.2.4.3 Measures of Clinical Performance

Students' ACTS performance was evaluated by the nurse confederate who participated in all the scenarios and sessions, the nurse confederate was an experienced medical educator and was consistent for all participants. Clinical performance (Figure 2.5) was rated using a 100-point Visual Analogue Scale (VAS) from 1= Unacceptable for their level to 100 = Outstanding for their level. Validity evidence for this type of performance assessment in medical education is presented elsewhere (Cha et al., 2019).


1. How appropriate was the supportive treatment the student selected?
2. How accurate was the student's focused assessment?
3. How accurate was the student's differential diagnosis?
4. How appropriate was the definitive treatment the student selected?
5. Please rate the overall performance of the student (based on the expectations for his/her level).

0
Not appropriate/accurate
Unacceptable for level

50
Average

100
Very appropriate/accurate
Outstanding for level

Figure 2.5. Sample questions from clinical performance assessment tool

2.2.4.4 Statistical Analysis

All statistical analyses were performed on (Minitab, 2000). T-tests were performed to identify differences in audio variables between student to nurse and student to patient. Pearson's correlation was used to determine associations between individual audio variables and performance. Regression analysis with forward stepwise variable selection was performed to determine statistically (alpha=0.05) significant audio predictors for clinical performance.

2.3 Results

Forty participants completed the study. The frequency of each scenario and clinical performance score is shown in Table 2.1. Frequency across scenarios was not uniform as participation in the study was voluntary and randomly assigned.

Table 2.1. Average duration and clinical score of each scenario

Scenario	Participated Students	Clinical Score Average	Clinical Score Standard Deviation
Motor vehicle accident shock	8	62.8	28
Pneumothorax	7	65.1	25
Hyponatremia	7	76.4	13.1
Leg compartment syndrome	6	71.3	26.2
Pulmonary embolism	6	72	17.8
Heparin-Induced Thrombocytopenia	6	45	25.9

Average frequency of dyad communication per scenario were as follows: 39 dyad communications initiated by students, 11 dyad communications initiated by nurse, and 5 dyad communications initiated by patient. Figure 2.6 summarizes the conversations with (Figure 2.6.a) and without (Figure 2.6.b) response from initiator to receiver. The nodes represent each individual role in the simulation, and the thickness of the arrows represent the proportion of communications over all scenarios. The direction of the arrowhead represents initiator to receiver relationship. The majority of conversations with response (Figure 2.6.a) were initiated from the student (67%). Conversations initiated by the patient were the least frequent. Five percent of communication from the students did not receive a response from either the nurse or the patient (Figure 2.6.b). Only 1% of the communication resulted in students not responding to the nurse or patient (Figure 2.6.b).

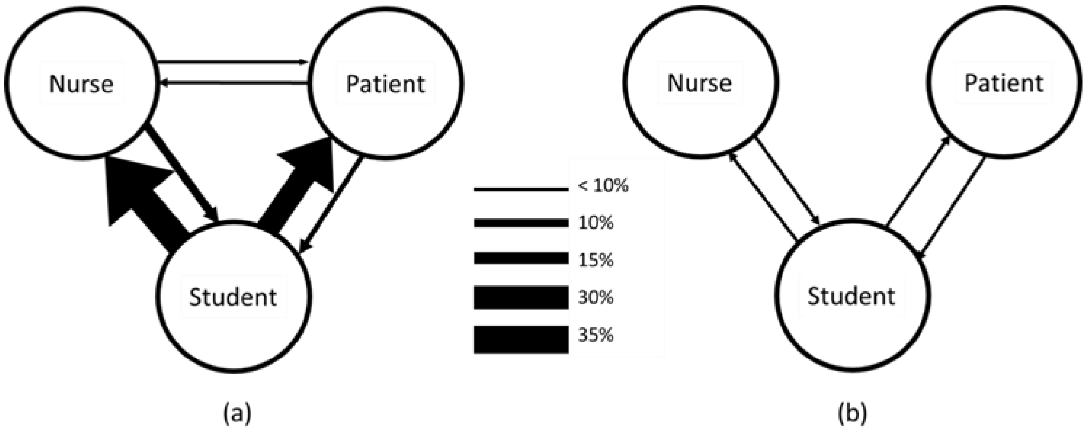


Figure 2.6. Link analyses of proportion of interpersonal communications observed between the team members (link widths are scaled to frequency): (a) with check-back (response), (b) without check-back (response)

Communications were further analyzed by type (question or statement), completion of a check-back, and roles, e.g., student to nurse, student to patient, nurse to patient, etc. (Figure 2.7). The most frequent type of communication was student to nurse questions with check-back (Figure 2.7.a). Out of an average 16 student-to-nurse questions per scenario (Figure 2.7.a and Figure 2.7.c), 99% received a check-back from the nurse. About 2.6% of questions from nurse or patient did not receive a student's check-back.

For statements, students received check-backs from the nurse 91% of the time. However, student statements to the patient received check-backs least frequently at 75% per scenario. Comparing across roles, communication between nurse and patient were the least frequently observed regardless of conversation type (question vs. statement) or occurrence of check-backs.

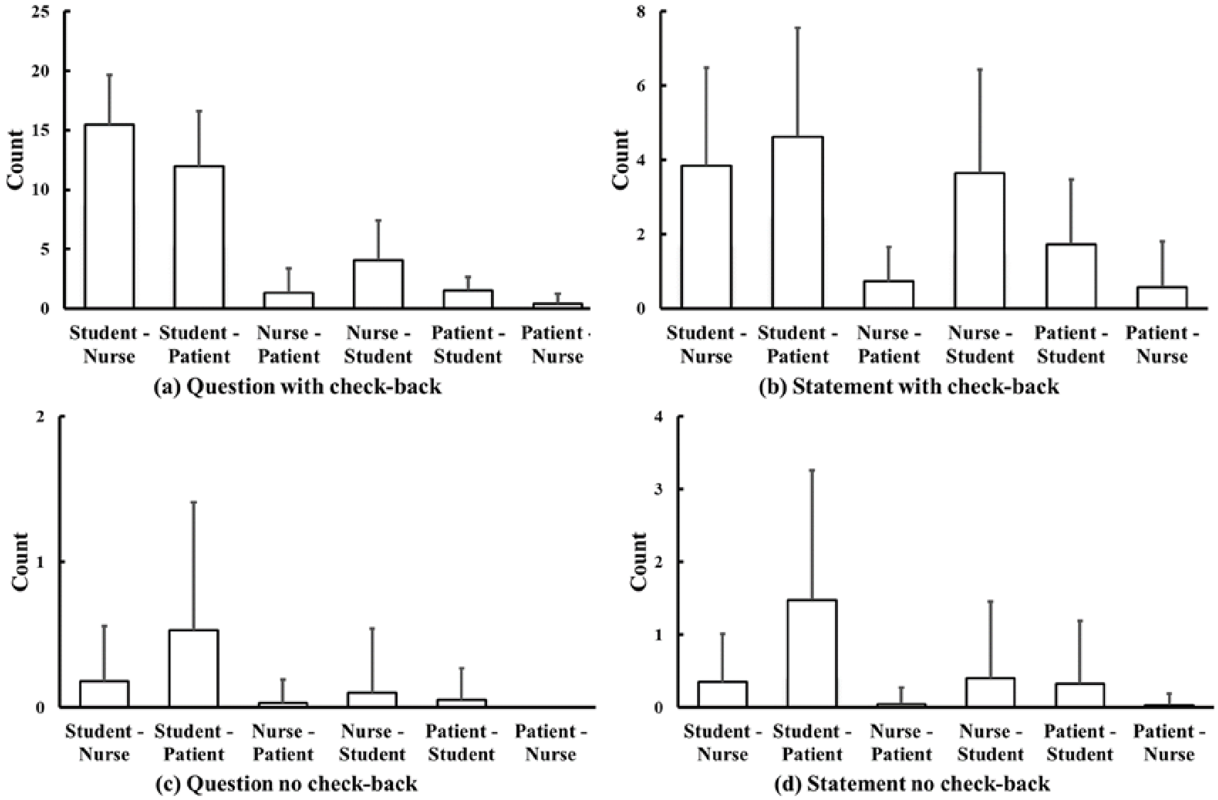


Figure 2.7. Interpersonal communications among student, nurse, and patient of all cases. (a) Overall number of initiated questions with a check-back (response) from all cases. (b) Overall number of initiated statements with a check-back (response) from all cases. (c) Overall number of initiated questions without a check-back (response) from all cases. (d) Overall number of initiated statements without a check-back (response) from all cases.

Findings from the audio processing are shown in Table 2. Comparison of speech features between student to nurse and student to patient showed that student audio features differed depending on whether s/he was speaking to the patient or nurse. On average communication to patient was 3 dB louder ($p<0.05$) than nurse, speech ratio to patient was 5% more ($p<0.05$) and more questions and statements with check-back were directed to nurse ($p<0.05$). Speech duration and speech rate were not statistically different whether the student was communicating with the nurse or patient.

Table 2.2. Comparison between student to nurse and student to patient with significant differences indicated

	To Nurse		To Patient			
Speech Measurables	Mean	SD	Mean	SD	T-Value	*p*-Value
Speech Duration (s)	65.77	20.67	61.92	22.85	0.86	0.40
Speech Ratio	0.21	0.06	0.26	0.07	-5.01	<0.001
Speech Intensity: average (dB)	59.6	3.66	62.56	2.54	-6.7	<0.001
Speech Rate (syllable/second)	2.48	0.6	2.47	0.58	0.14	0.89
Question and Statement with check-backs	19.33	4.5	16.58	5.03	2.62	0.01

Vocal features derived from the audio recordings showed significant correlation to performance scores. Out of all the features, the strongest relationship with respect to performance score was the positive association (ρ= 0.493, p=0.001) between frequency communication initiated by the student (regardless of statement or question or to which team member) as shown in Figure 2.8. Frequency communication between student and nurse (with check-back) were also significantly correlated with performance (ρ= 0.456, p=0.003), while speech duration between student to nurse approached but did not reach significance (ρ= 0.294, p=0.066).

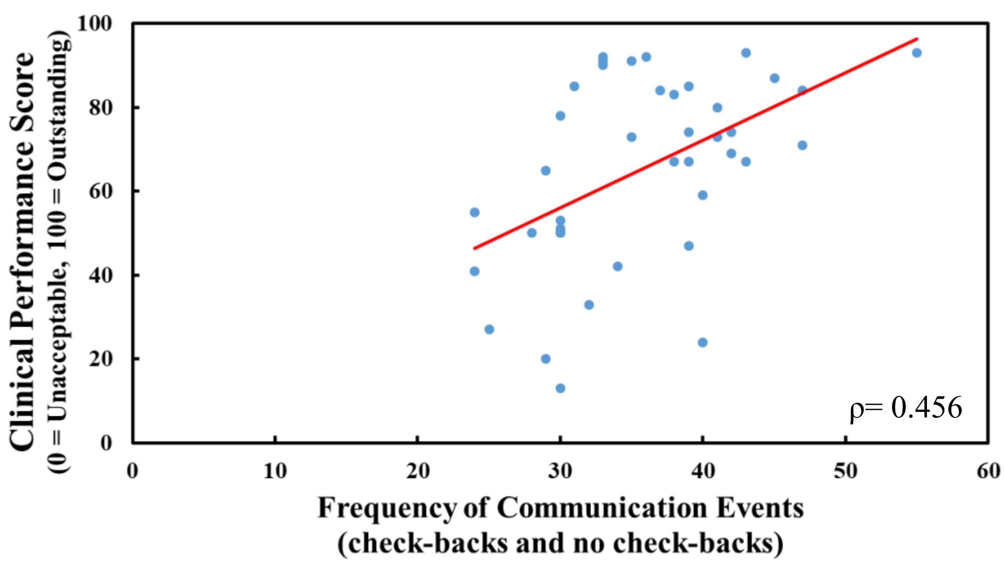


Figure 2.8. Correlation between frequency of communications and clinical performance

Leveraging all communication metrics to determine the relationship between communication and performance, a forward stepwise regression approach identified significant predictors of performance. The stepwise selection helps remove variables that may be collinear, stepwise selection result can be seen in Table 2.3. Resulting model from stepwise regression of speech features to predict clinical performance (Q: communication started with question. NQ: communication started with statement). The final model resulted with an R-squared value of 0.61. Increasing frequency of statements from student to nurse that received check-backs improved performance ($p<0.05$). Similarly, increased frequency of questions from student to nurse that received check-backs improved performance ($p<0.001$). Variables describing conversation without check-backs (i.e., communication that received no response from the other individual) were not predictive of performance. For vocal features, duration of speech directed towards the patient also had positive association ($p<0.001$) with performance. Although not statistically significant, increasing speech intensity to patient ($p=0.07$) and speech ratio ($p=0.10$) were also associated with improved performance.

Table 2.3. Resulting model from stepwise regression of speech features to predict clinical performance (Q: communication started with question. NQ: communication started with statement)

Terms	Coef	P-Value
Constant	77.7	0.266
Student to nurse with check-back (Q)	3.268	**<0.001**
Nurse to student with check-back (Q)	1.23	0.154
Student to nurse with check-back (NQ)	2.52	**0.037**
Nurse to student with check-back (NQ)	-1.39	0.203
Patient to nurse with check-back (NQ)	4.49	0.059
Speech ratio overall	-112.8	0.1
Student to nurse speech Intensity	1.72	0.115
Student to patient speech Duration	0.586	**0.001**
Student to patient speech Intensity	-3.02	0.065

2.4 Limitations

This study contains limitations which are important to note. First, although the sample size of the study (40 participants) was adequate, participant characteristics such as gender, age, or ethnicity were not examined to test if any differences among them exist. Given that these demographic factors may affect speech behaviors, future studies should assess their impact on communication features.

In this simulation, limited communication was observed between nurse and patient, which could be due to the simulation setting that students were the main evaluation subject, being tested by the nurse confederate and the manikin patient controlled by a technologist. Although the participants' actual role showed a difference from the real hospital setting, the diagnostic process, environment, and equipment were identical.

For the six different case scenarios, due to randomized sampling, all case scenarios were not evenly distributed. Furthermore, each diagnostic process was different across all scenarios. Nevertheless, we used this approach to study communication across a variety of clinical case scenarios.

Segmentation of the audio used in this study may not be the most robust way to analyze speech ratio, since ratio of speech is weighted by how long nurses take to response or how long the student waits to initiate another conversation. Furthermore, each scenario was different and thus resulted in different diagnostic process, medical examinations and treatments. Some scenarios may require more hands-on examinations, causing a longer silence in the segment which resulted in a smaller speech ratio. Therefore, future research should develop a more standardized coding on segmentation.

Additionally, for determining the number of interactions, only direct communication (i.e., call-outs and check-backs) were observed and analyzed. Although this showed a detailed number of interactions, this might not directly reflect the actual frequency of communication events. For example, 20 call-outs with check-backs from student to nurse could be one communication event, and two call-outs with check-backs from student to nurse could be another. Future research in full CLC could provide frequency of communication events inclusively.

Due to the context that this was an initial study, communication variables and features were selected based on previous studies. Future research will focus on automating the key features found in the present study.

Finally, the statistical models were generated through stepwise regression, but due to the richness of continuous audio recordings and features, future research with application on machine learning approaches may provide additional insights on the relationship between audio metrics and performance.

3. RESEARCH QUESTION II: MEASURES OF NON-TECHNICAL SKILLS

3.1 Research Framework

The long-term goal of this work is to identify objective sensing-based features to assess NT skills, providing a tool to train and enhance healthcare providers' performance. Toward this goal, this study's primary objective is to identify vocal features that predict communication construct ratings from an observer-based NT assessment tool. Our secondary objective is to identify vocal features that predict ratings on overall NT skills using current NT assessment tools. In this paper we hypothesize that:

1) vocal and communication variables correlate with communication construct under NT skills, and
2) vocal and communication variables correlate with overall NT skills performance.

3.2 Methodology

3.2.1 Study Participants

This work complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at the university (IRB:1611105172). Participants were recruited from one large academic medical institution, and all participants were on their surgery clerkship rotation during their 3rd year of medical school. Each participant had an orientation at their beginning of the surgery clerkship which includes simulated scenarios, however they were significantly different from Acute Care Team Simulations (ACTS), e.g., a patient simulator was used in this study. Therefore, each participant was at the same experience level during ACTS. Each participant provided informed consent for audio and video recording during his/her ACTS.

3.2.2 Study Procedure

All data collection occurred at the state-of-the-art simulation center located at Fairbanks Hall, Indiana University School of Medicine. This simulation center was designed to educate and provide hands-on practice for multidisciplinary healthcare providers through the replication of multiple facets of the patient care environment, including fully-functional care units and operating rooms, that can be regulated and controlled completely by simulation technologists using a computer system. Within the simulation center, healthcare providers can participate in various immersive simulated patient care scenarios and gain practical training for actual patient care.

ACTS were one form of summative evaluation of medical student performance after they had finished their general medical procedure clerkship rotation. There were six scenarios: 1) Motor vehicle accident shock, 2) Pneumothorax, 3) Hyponatremia, 4) Leg compartment syndrome, 5) Pulmonary embolism, and 6) Heparin-Induced Thrombocytopenia. These scenarios captured the range of care management, e.g., patient arriving to the emergency department (ED), post-operative patient management to care diagnoses, and medical treatment, and were further described in previous publications (Cha et al., 2019; Peng et al., 2019).

Each ACTS session involved three individuals: one clinician (role assigned to the medical student participant), a nurse (role performed by the same person confederate for every session), and a patient (role performed by a simulation technologist operating a mannequin (SimMan 3G, Laerdal Medical, Wappingers Falls, NY)). The technologist acted as the patient in each scenario and the technologist's voice can be heard through the mannequin. The mannequin featured functionalities including pulse and breathing sounds among other physiological signals, and these features can be adjusted according to the design of simulation scenario.

Each student performed one scenario, randomly assigned from the six, and each scenario averaged 10 minutes in length. The students were expected to perform tasks including completion of patient assessments, diagnosis determination, and carrying out an appropriate treatment. With training from an experienced faculty observer, the simulation technologist manipulated the patient's health (i.e., improve, remain the same, or worsen) based on student's actions. The role of nurse was also performed by the faculty, to assist students' diagnosis and provide a factual simulation process.

3.2.3 Data Collection

3.2.3.1 Objective Measures

All participants (student, nurse, and patient) were recorded throughout each scenario. Voice recorders (Zoom H1, Zoom, Inc, Hauppauge, NY, USA) were placed in their pocket, and each recorder was connected to a lapel microphone (RØDE smartLav+ Microphone, RØDE Microphones, Silver Water, NSW, Australia) attached to the scrub or jacket collar of each participant. Although wearing microphones was more intrusive than extracting audio from video, this approach allowed better localization of audio source and noise cancelling to provide more accurate data for this exploratory on an objective audio-based approach to NT skills assessment. The audio setup required <1-minute. Video recordings captured three room views (i.e., patient view, overhead view, and a view of patient vitals) using the cameras (Panasonic WVCS574, Panasonic, Kadoma, Osaka Prefecture, Japan) built into the simulation center (B-Line Medical, Washington, DC). These were used as needed to verify and interpret audio findings.

3.2.3.2 Clinical Performance

Medical students' ACTS performance was evaluated by the faculty nurse confederate who participated in all the scenarios and sessions. Clinical performance was rated using a 100-point Visual Analogue Scale (VAS) from 1= Unacceptable for their level to 100 = Outstanding for their level.

3.2.3.3 Non-technical Skills

The NT skills of the participants was assessed using a modified tool incorporating principles from NOTECHS, NOTSS, and OTAS (Cha et al., 2019). The five NT constructs included in this tool were communication, situation awareness/vigilance, cooperation/team skills, leadership, and decision making/problem-solving. Elements within each construct (e.g., "instructions/questions to team members were distinct" element under the communication construct) were scored between behavioral markers representing very problematic behavior (0) and exemplar behavior (6). The average of the elements was used to obtain a construct score, and the average of the five constructs were used to obtain an overall NT skills score. Each participant's NT skills were evaluated by a total of three raters (two human factors raters and one clinician rater).

3.2.4 Data Analysis

Raw audio data was continuously captured from all team members (Figure 3.1). The full pipeline for deriving variables for statistical analysis from raw audio is illustrated in Figure 2. Specifically, four general steps were performed to process the raw audio:

1) Annotate the "communication type" for each speaking turns
2) Extract speech features
3) Determine the target of the communication, and
4) Summarize each speech feature with statistical descriptors.

From this pipeline, two sets of features/variables were extracted from the raw audio: 1) individual-level variables and 2) team-level variables.

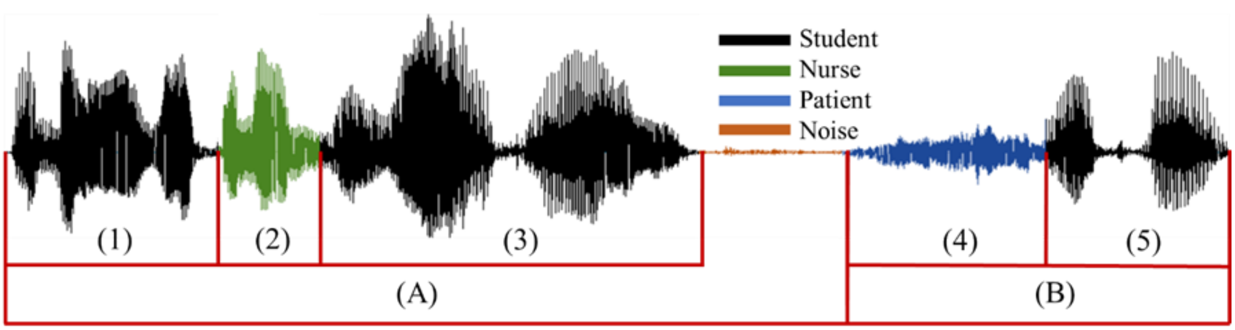


Figure 3.1. Raw audio excerpt illustrating how raw audio was processed into labels for speaker turns (numbers 1 – 5 marks voiced duration of each participant) and communication dyads (letters A and B show each change in medical students' communication target were dyad A was directed at nurse and dyad B was directed at patient)

3.2.4.1 Individual-level Variables

The individual-level variables focused on the audio originating from the speaker (this study focuses on the student speaker). Following the pipeline (Figure 3.2.a), raw audio was annotated and processed as follows to create variables for statistical modeling and hypothesis testing

3.2.4.1.1 Communication Type

Communication type (Figure 3.2.a): The raw audio (Figure 3.1) was first segmented by a study team member using Praat software(Boersma, Paul & Weenink, 2018) into two communication types: 1) each speaker turn and 2) each communication dyad.

- Speaker turn was defined as the voiced duration from a participant, from the first audible word until the end of their speaking turn, i.e., before the start of the next turn. These are illustrated by the numbered segments in Figure 1. For example (Figure 1), speaker turn (1) was student's voiced duration and turn (2) was nurse's voiced duration.
- Communication dyad was defined by each change in medical students' communication target (nurse or patient), communication dyad A and B represents the communication between student to nurse and student to patient accordingly.

3.2.4.1.2 Vocal Features

Vocal features: After segmenting into turns and dyads, Praat software was then used to extract vocal features. Speech duration, intensity, and pitch of the student participant were extracted automatically both at the turn and dyad levels. Note that "articulate speech rate" (defined as number of syllables per turn) was only derived from speaker turns, while "speech rate" (defined as number of syllables per communication dyads) was only derived for communication dyads. Specifically, the key difference between articulate speech rate and speech rate was the latter was influenced by silent sections of audio (Figure 3.1).

In addition to speech duration, intensity, pitch, articulate speech features, two additional time-domain features were calculated from speaker turns: 1) burstiness and 2) successive differences. The burstiness of speaking was a measure of the temporal distribution of time spent speaking, also defined as the coefficient of variation.(Rosen et al., 2018) Higher levels of burstiness of speaking represented speaking that was more clumped together in time with periods of relatively intensity and sparseness, whereas lower burstiness meant a more even distribution of speaking over time (Rosen et al., 2018).

$$Burstiness = \frac{\sigma}{\mu} \times 100\%$$

Burstiness was calculated for each vocal feature, e.g., duration and intensity (Figure 3.2.a). For example, if were to calculate the burstiness of speech duration then here σ and μ would be the standard deviation and mean of speech duration. In Figure 1, if speaker turns numbered (1), (3), and (5) were 5 seconds, 10 seconds, and 4 seconds in duration, respectively, then would result in a σ of 2.62 seconds and μ of 6.33 seconds:

$$Burstiness_{Speech\ Duration} = \frac{2.62}{6.33} \times 100\% = 41\%$$

The successive difference was a measure of the difference between each consecutive speaker turn, following with above example, the successive difference of speaker turns numbered 1, 3, and 5 would be:

$$Successive\ Difference_{turn\ 3\ and\ 1} = 10 - 5 = 5\ seconds$$
$$Successive\ Difference_{turn\ 5\ and\ 3} = 4 - 10 = -6\ seconds$$

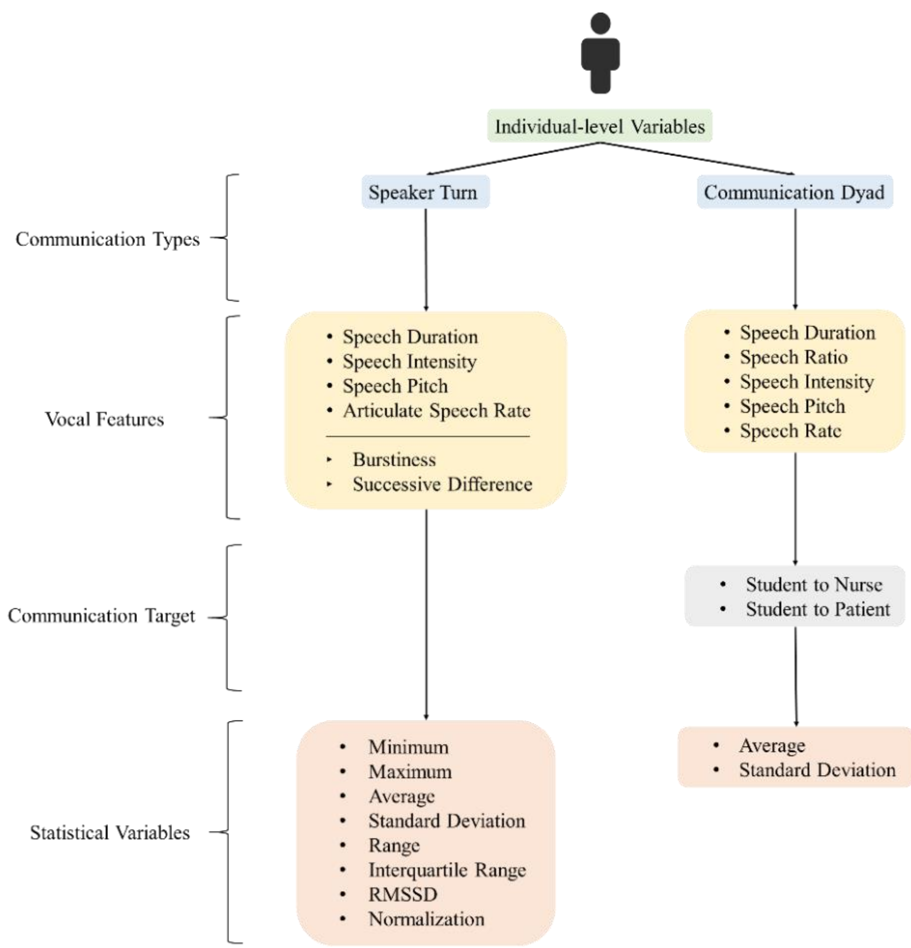
Similarly, successive difference was also calculated for each vocal feature. A negative value of successive difference between two speaker turns represented a decrease of the vocal feature, and positive represented an increase.

3.2.4.1.3 Communication Target

Communication target: This step was only applicable to dyad communication type. A communication target was the message recipient, for example, if student initiated a message to nurse, and nurse was defined as the communication target. Although rare, when both nurse and patient responding to student's turn, the communication target was identified by the content of whom the message initiator was speaking to.

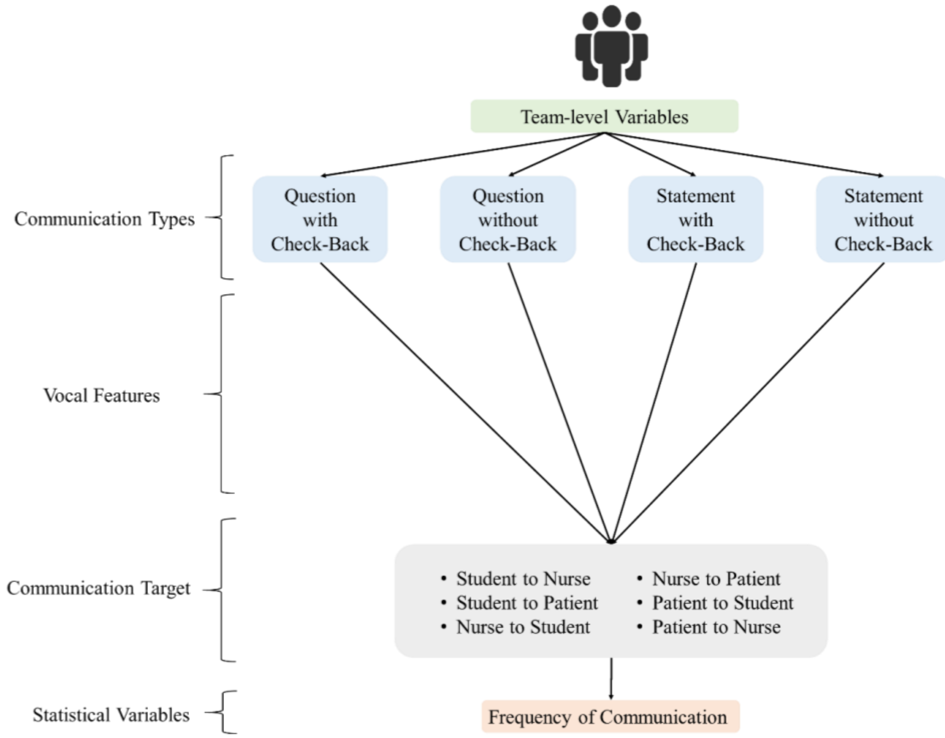
3.2.4.1.4 Statistical Variables

Statistical variables: This last step of the pipeline summarized the features into statistical descriptors over the scenario. Example statistical descriptors for this study included: minimum, maximum, average, standard deviation, range, inter-quartile range, burstiness, successive turn difference, and Root Mean Square of the Successive Differences (RMSSD) (Figure 3.2.a). For communication dyad, only average and standard deviation were calculated.



(a)

Figure 3.2 continued



(b)

Figure 3.2. Vocal features characterized by (a) individual and (b) team-level variables

3.2.4.2 Team-level Variables

The team-level variables were annotated using video annotation software published previously (Yu et al., 2014). Software enabled study team members to annotate communication type and targets while audio/video of the simulation was played simultaneously. All annotations were time-stamped and saved in an excel file. The pipeline for analysis of team-level variables is described below (Figure 3.2.b).

3.2.4.2.1 Communication Type

Communication type: In the team-level analysis, four communication types were annotated based on 1) whether the speaker initiated a question o statement, and 2) whether that initial message received a check-back or not. The resulted four types are shown in Figure 3.2.b. These categories were annotated based on findings from published studies (Härgestam, Lindkvist, Brulin, Jacobsson, & Hultin, 2013; Peng et al., 2019).

3.2.4.2.2 Vocal Features

Vocal features: This step in the pipeline was not applicable in team-level analysis.

3.2.4.2.3 Communication Target

Communication target. Since the simulation design involved only three team-members, only three pairs of dyads were possible, i.e., student and nurse, student and patient, and nurse and patient. All three pairs were annotated for analysis. Each initiated message was catogorized into four communication types.

3.2.4.2.4 Statistical Analysis

Statistical Analysis: All hypotheses were investigated using regression analysis with forward stepwise variable selection. Inputs for the model were Statistical Variables (Figure 3.2) and the outputs were observed rated communication NT score (hypothesis 1) and overall NT score (hypothesis 2). All statistical analyses were performed on Minitab (Minitab, 2000). Pearson's correlation was used to determine associations between individual audio variables and NTS scores.

3.3 Results

Fifty-six participants consented to the study, and audio recordings for their simulation sessions were analyzed. Participants consisted of 43% females and 57% males with overall average age (± standard deviation) of 25 ± 1.8 years (three participants' age were unidentified). Average and standard deviation of participants on six scenarios were 9 ± 1.2.

For individual-level variables, Table 3.1.b summarizes key vocal features based on speaker turn and communication dyad. For speaker turn, the average and standard deviation of speech duration, intensity, pitch, and articulate rate were 1.9±0.7, 62.6±3.0, 182.5±49.1, and 182.5±49.1, respectively. For communication dyad, the average and standard deviation of speech duration, intensity, pitch, ratio, and rate were 128.5±31.8, 61.2±2.7, 186.3±50.8, 0.2±0.1, and 2.7±1.6, respectively. As shown in the table, average and standard deviation of speech intensity and pitch between speaker turn and communication dyad are close to equivalent. Average and standard deviation of burstiness and successive difference of speech duration, intensity, pitch, and articulate rate are also shown in Table 3.1.

Table 3.1. Descriptive statistics (Average ± Standard Deviation) of key vocal features and communication variables of a) individual-level variables categorized by communication type, and b) team-level variables categorized by communication type (n=56 participants)

(a) individual-level variables categorized by communication type

Vocal Features	Speaker Turn	Communication Dyad
Speech Duration (s)	1.9±0.7	128.5±31.8
Speech Intensity (dB)	62.6±3.0	61.2±2.7
Speech Pitch (Hz)	182.5±49.1	186.3±50.8
Speech Ratio	-	0.2±0.1
Speech Rate (Syllable/second)	-	2.7±1.6
Articulate Speech Rate (Syllable/second)	3.7±0.3	-
Burstiness of Speech Duration	0.9±0.1	-
Burstiness of Speech Intensity	0.1 ±0.09	-
Burstiness of Speech Pitch	0.2 ±0.05	-
Burstiness of Articulate Speech Rate	0.3 ±0.04	-
Successive Difference of Speech Duration	0±0.08	-
Successive Difference of Speech Intensity	-0.06±0.1	-
Successive Difference of Speech Pitch	-0.08±0.8	-
Successive Difference of Articulate Speech Rate	-0.01±0.03	-

(b) Team-level variables categorized by communication type

Message Initiator	Message Receiver	Question Initiated Message		Statement Initiated Message	
		with Check-back	without Check-back	with Check-back	without Check-back
Student					
	Nurse	29.1%±6.9%	0.3%±0.8%	7.5%±4.9%	0.8%±1.5%
	Patient	23.6%±8.3%	0.8%±1.4%	9.0%±5.2%	2.8%±3.1%
Nurse					
	Student	8.2%±5.9%	0.2%±0.6%	6.3%±4.1%	0.4%±1.1%
	Patient	1.7%±3.3%	0.02%±0.2%	1.3%±1.5%	0.2%±0.5%
Patient					
	Student	3.1%±2.3%	0	2.8%±3.1%	0.5%±1.1%
	Nurse	0.6%±1.5%	0	0.6%±1.5%	0.06%±0.3%

The ICC analysis for labeling type and check-back resulted in 0.859 and 0.879 for agreement and consistency, respectively. These coefficients fell in into the range of 0.81 to 1.00, which indicate significant reliability (McHugh, 2012). For team-level variables (Table 3.1.b), average frequency of communication per scenario shows that 74% (±9%) of message was initiated by the student, 18% (±7%) by the nurse, and 8% (±5%) by the patient; 79% (±9%) of the question initiated message was made by the student, nurse and patient contributed to 15% (±8%) and 6% (±4%) each; 63% (±17%) of the statement initiated message was made by the student, 26% (±14%) by the nurse, and 12% (±11%) by the patient. For question initiated messages, 55% (±12%) was directed from student to nurse, and 45% (±12%) was directed from student to patient; for statement initiated messages, 43% (±22%) was directed from student to nurse, and 57% (±22%) was directed from student to patient. Table 3.1.b also summarizes communication transmitted message with and without check-back. Majority of the messages followed with response were initiated by the student (69%), and the patient was the least frequent message initiator (7%). Only 5% of messages initiated by the student were not followed by a response.

NT skills scores were summarized in Table 3.2. Overall NT score was 2.9±0.4, which represented neither problematic nor model behavior. Among the six subscales that compose the overall NT score, communication and leadership had the highest average score, while leadership was the lowest.

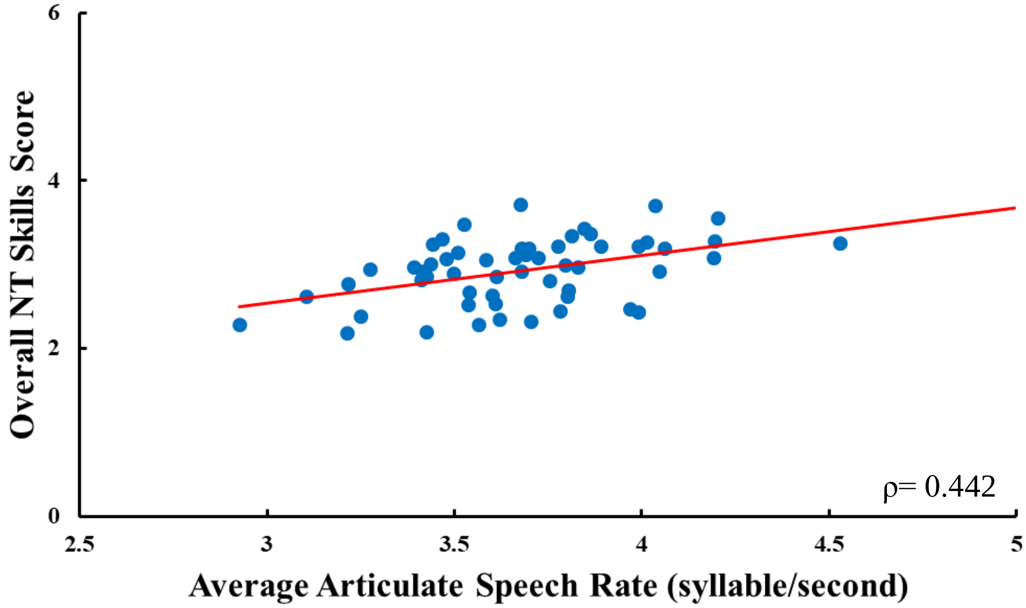
Table 3.2. Descriptive statistics of NT skills scores across all scenarios

Construct	Score (mean±SD)	Minimum Score	Maximum Score
Communication	3.1±0.3	2.5	3.9
Situational Awareness	2.9±0.6	1.8	4.3
Teamwork	3.0±0.2	2.5	3.8
Leadership	2.8±0.6	1.0	4.0
Decision Making	2.9±0.4	1.7	3.7
Overall	2.9±0.4	2.2	3.7

Score of 0 represents problematic performance and 6 represents model behavior.

Figure 3.3 shows the significant correlations between objective predictors (i.e., vocal features and speech variables) and both NT skill and communication construct. Two individual-level variables: 1) “average articulate speech rate” (speaker turn based variable), 2) “average speech duration from student to patient” (communication dyad based variable); and one team-level variables: “question initiated message from student to patient with check-back” had the strongest correlation with overall NT skills scores ($\rho=0.442$, $p=0.001$; $\rho=0.456$, $p<0.001$; and $\rho=0.429$, $p=0.001$, respectively) among all other speech features (Figure 3.2). For the communication construct, speaker turn based variable “average articulate speech rate” ($\rho=0.303$, $p=0.023$) and dyad variable “Average speech duration from student to patient” ($\rho=0.376$, $p=0.004$) under individual-level variables had the strongest correlation. In addition, communication construct was also strongly correlated with the team-level variable of frequency of communication (number of questions-initiated messages from student to patient followed with check-back) ($\rho=0.394$, $p=0.003$).

As shown in Figure 3, variables related to articulation speech rate, speech duration from student to patient and the interaction between student to patient correlated with all NT skill constructs and overall NT skills.



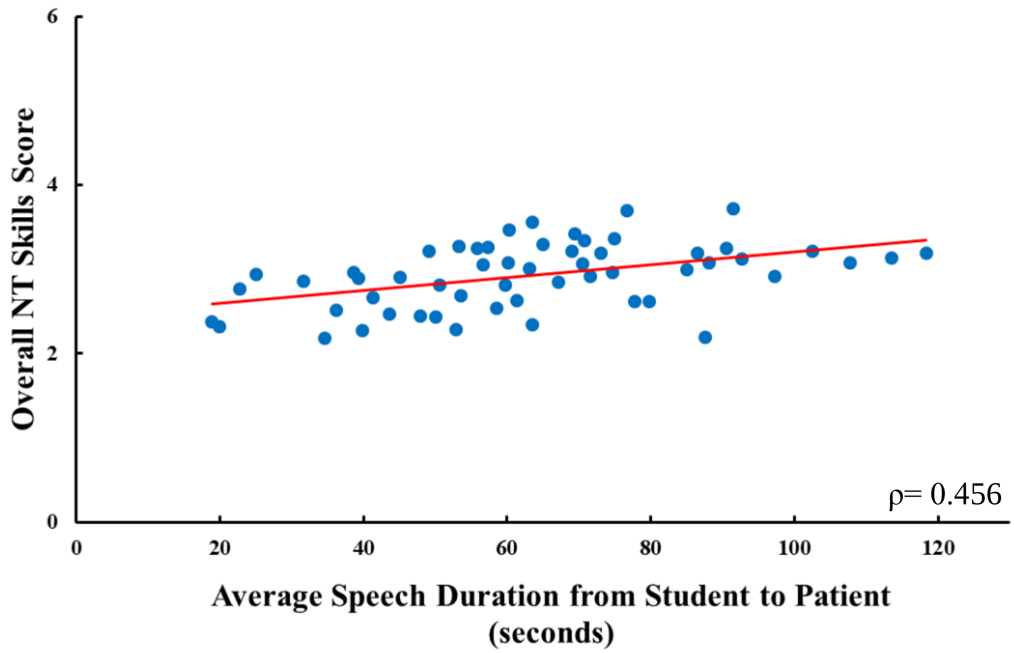
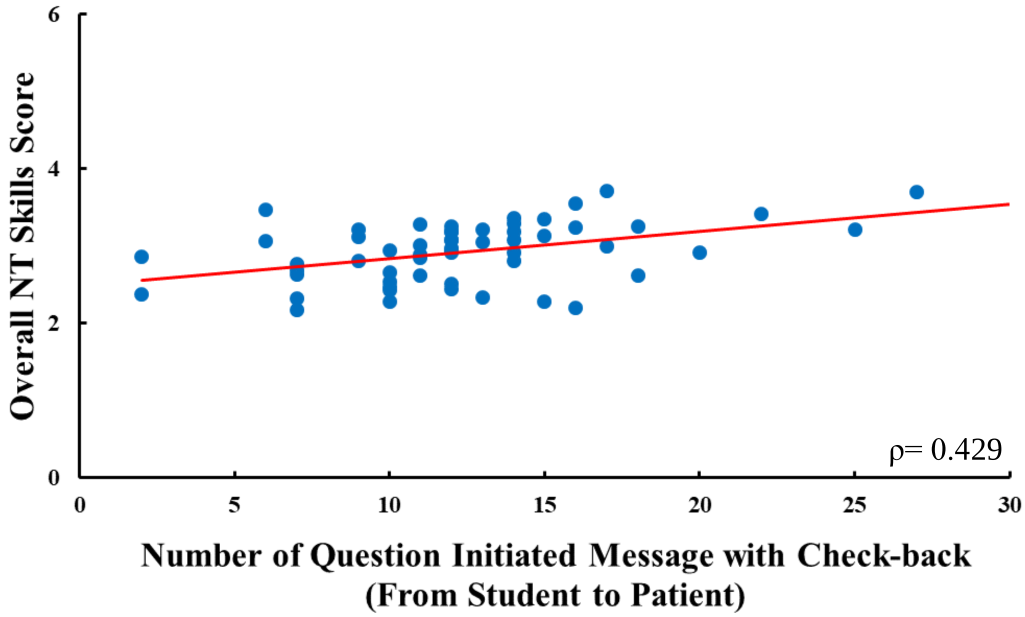


Figure 3.3 continued



(a)

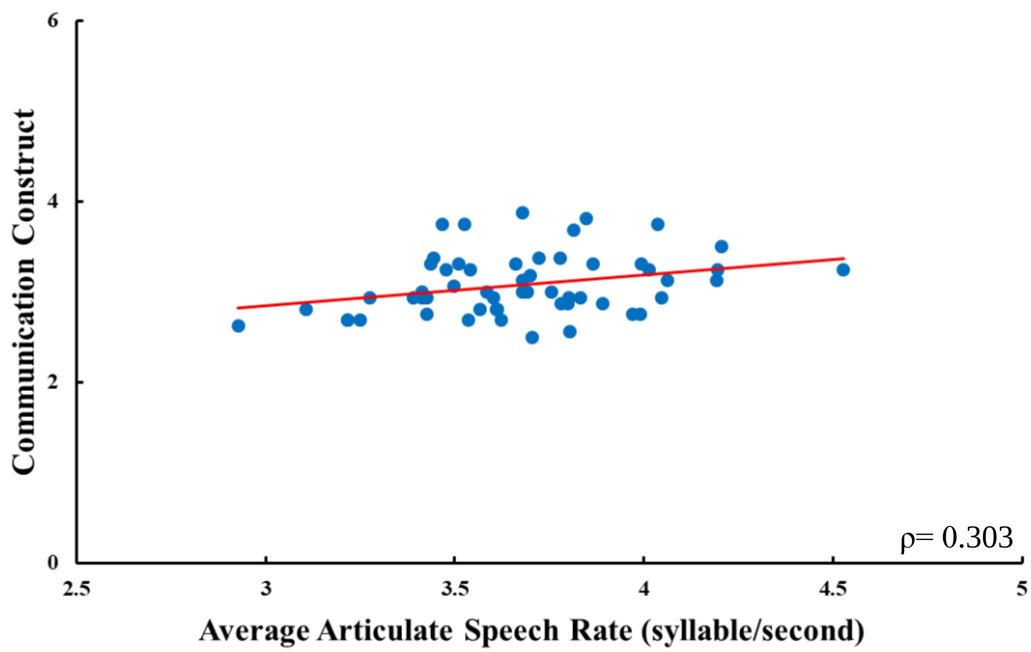
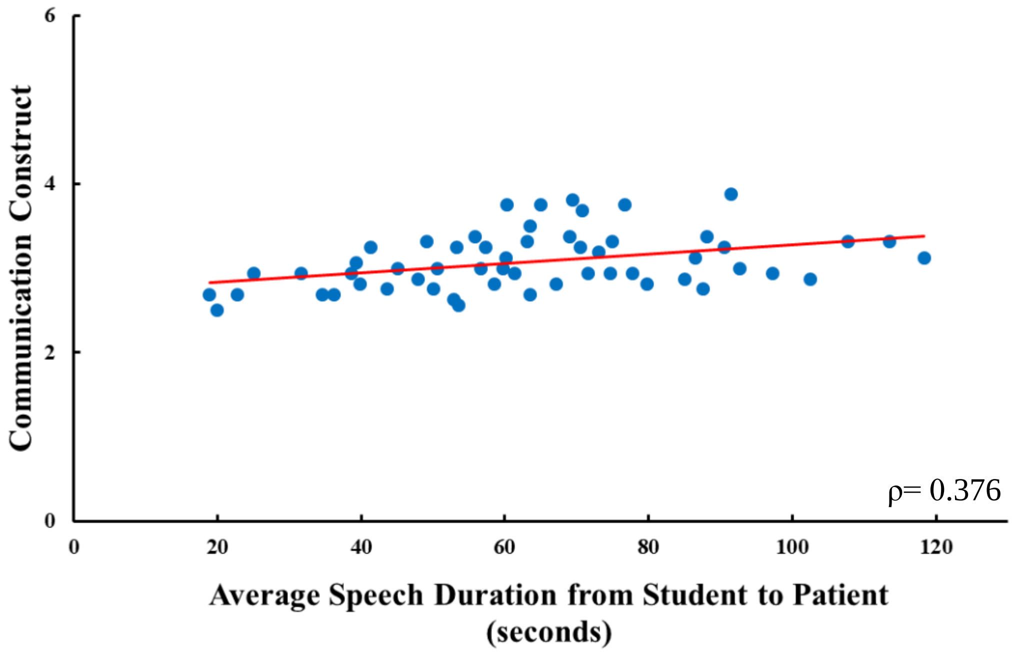
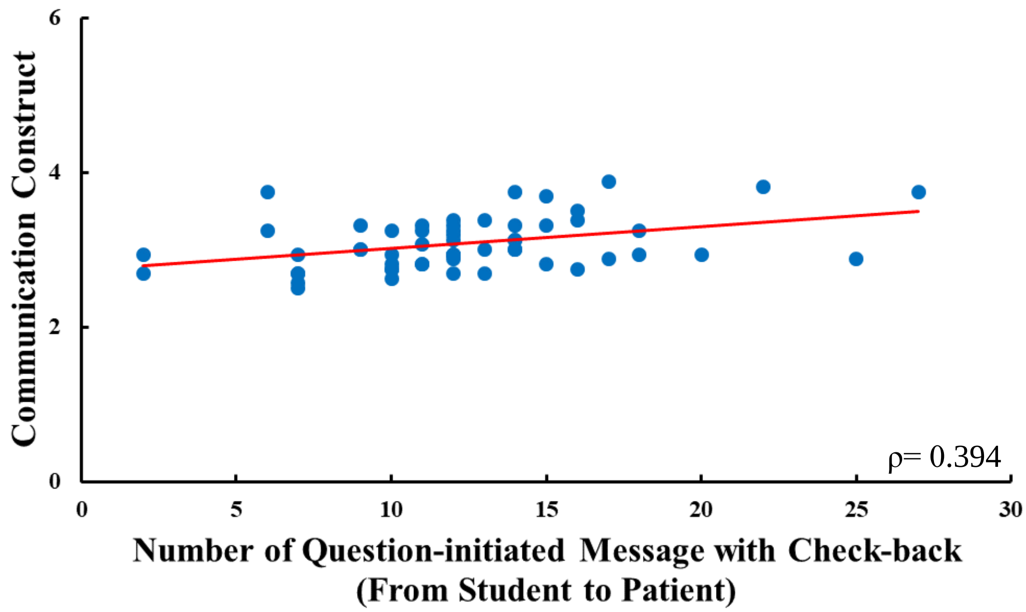


Figure 3.3 continued





(b)

Figure 3.3. Relationship vocal features and speech variables with NTS score and communication construct: (a) significant correlation between objective predictors and NTS score, (b) significant correlation between objective predictors and communication construct

Results of the forward stepwise regression that enabled the modeling of all features, i.e., all statistical descriptors of each vocal feature (Figure 3.2), is show in Table 3.3. Diagnostic testing of the resulting regression model indicated showed no violation of regression assumptions of normal distribution, linearity, and homoscedasticity. In the overall NT score model, significant factors include four speaker turn-based vocal features (minimum speech pitch, average articulate speech rate, burstiness of speech intensity, and maximum successive difference of articulate speech rate), two communication dyad-based vocal features (speech duration from student to patient, and normalized speech intensity), and one team-level variable (interaction frequency between student to nurse statements with check-backs). The resulting model had an adjusted R-squared value of 0.537 and both standardized and unstandardized coefficients of the model is shown in Table 3.3.

Table 3.3. Coefficient of communication and speech variables with overall NT skill scores.

Terms	**Unstandardized Coefficient**	**Standardized Coefficient**
Average Speech Duration (Communication Dyad based: Student to Patient)	0.01	0.23
Average Normalized Speech Intensity (Communication Dyad based: Student to Patient)	1.11	0.14
Minimum Speech Pitch (Speaker Turn based)	(-0.004)	(-0.14)
Average Articulation Speech Rate (Speaker Turn based)	0.53	0.17
Burstiness of Speech Intensity (Speaker Turn based)	1.23	0.11
Maximum Successive Difference of Articulation Speech Rate (Speaker Turn based)	0.16	0.12
Interaction from student to nurse statement with check-back (Team-level variable)	(-0.03)	(-0.07)
R-squared (adjusted)	**53.7%**	**53.7%**

Bolded terms indicated statistically significant variable (p< 0.05).

For communication construct model (Table 4), significant factors include six turn based vocal features (interquartile range of speech intensity, burstiness of speech intensity, range of articulation speech rate, burstiness of articulation speech rate, maximum successive difference of articulation speech rate, and speech rate, maximum successive difference of articulation speech intensity), and three team-level based variables (interaction from student to nurse question with check-back, interaction from student to patient question with check-back, and interaction from patient to student statement without check-back). Resulting with an adjusted R-squared value of 0.54.

Table 3.4. Coefficient of communication and speech variables with communication construct.

Terms	**Unstandardized Coefficient**	**Standardized Coefficient**
Average Speech Intensity (Dyad based: Student to both Nurse and Patient combined)	(-0.02)	(-0.06)
Interquartile Range Speech Intensity (Turn based)	0.15	0.19
Burstiness Speech Intensity (Turn based)	1.82	0.16
Range Articulation Speech Rate (Turn based)	(-0.14)	(-0.08)
Burstiness Articulation Speech Rate (Turn based)	(-3.35)	(-0.13)
Maximum Successive Difference of Articulation Speech Rate (Turn based)	0.10	0.07
Maximum Successive Difference of Speech Intensity (Turn based)	(-0.03)	(-0.12)
Frequency of Communication from Student to Nurse Question with Check-back (Team-level variable)	(-0.02)	(-0.08)
Frequency of Communication from Student to Patient Question with Check-back (Team-level variable)	0.03	0.14
Frequency of Communication from Patient to Student Statement without Check-back (Team-level variable)	(-0.13)	(-0.10)
R-squared (adjusted)	**54%**	**54%**

3.4 Limitations

Several limitations exist in this study and are important to note. Given an extensive sample size (56 participants), their demographic characteristics such as gender, age, or ethnicity were not tested in the study. Future study needs to include such factors to perform a more comprehensive understanding of impact of vocal features on NT skills, for example, pitch level was found to be one of the predictors in the statistical model, however, differences exist between male and female's speech pitch where female produces a higher pitch level than male. Moreover, although automated programs existed for automated audio segmentation by speaker, algorithms tested by the study team had limitations in accuracy. Thus, this initial exploration relating vocal features to trainee skill was segmented manually by a study team member to generate a more precise dataset for the predictive modeling.

Current study did not take into account of the speech baseline, for example, the differences between person speaking in a high level of tension environment and in an unstrained environment. Although this study considered the change of speech patterns, for example, the successive changes of speech intensity throughout the simulation, it is also important to provide an overall difference with respect to the baseline speech patterns.

Due to the simulation setting and scenarios, students were randomly assigned with one of the six scenarios, and each consisted of different diagnostic process and difficulty levels may vary among them. Some scenario may require more communication, and some may result in more examinations, which caused the differences in the frequency of communication and speech duration. However, several vocal features across different scenarios yielded a significant relation with overall NT skills and each construct.

Lastly, according to the simulation purpose and setup, the Acute Care Trauma Simulation (ACTS) was part of the course requirements for the medical students, and the nurse was also the instructor for this course. Therefore, the students know the nurse before took part in the simulation. Such relationships may have an effect on the speech measurables, for example, the change of speech patterns between speaking to a stranger and a known person. However, since this simulation setup was identical to real world hospital settings, doctors may also have relationships with fellow nurses and assistants, future studies should still look into how relationships can affect speech patterns. Furthermore, in a close to real world hospital setting simulation as mentioned previously, it is also necessary to identify any changes between formal and informal speech patterns.

4. PILOT TRANSFERABILITY INTO OPERATING ROOM STUDY

4.1 Specific Aims

Because of the significant findings from simulated study in ACTS (RQ I and RQ II), and to further test the functionality of this methodology, the framework from previous ACTS studies was implemented into actual operating room (OR) environment. The purposes of this pilot study were to:

1. derived vocal features and frequency of communication in surgical settings, and
2. understand surgical team's communication patterns.

4.2 Methodology

4.2.1 Study Participants

This pilot study took place during robotic-assisted surgery (RAS) where surgeons operated using a robot. The robotic system includes a surgeon console with controls (e.g., foot pedals, master controls, and controls to adjust positioning) and tele-surgical robotic arms. A total of 13 RAS cases were collected, each surgical case consisted of one head surgeon, one assistant, one technician, one anesthetist, one to two circulating nurses, and others (observers and medical students).

4.2.2 Data Collection and Surgical Procedure

Both audio and video was collected for each RAS case from the start (when the surgeon getting access to patient's body) to the end (end of suturing) of the case. These videos were later segmented into five phases:

1. Robotic arm docking into patient's body (starting with head surgeon's verbal indicator ("Ready to dock", or similar, ending with head surgeon seated in front of console),
2. 10 minutes before robotic arm un-docking (removing) from patient's body,
3. Critical phase during the surgery (e.g., tumor excision),
4. 5 minutes before critical phase, and
5. 5 minutes after critical phase.

The critical phase of each RAS was defined by a surgeon. Out of the 13 RAS cases, 5 cases' critical phase segment were performed by one surgeon, 5 were performed by two surgeons simultaneously, and two had operator switched during the segment.

4.3 Data Analysis

4.3.1 Audio Processing

Similar to RQ 2, vocal features for individual level variables were also extracted here in the RAS setting, and furthermore, the complete CLC technique was applied in the team level variables. For each call-out and check-back, they are categorized as one of the follows:

1. Request: direction, instructing, or requesting someone to do or report something
2. Confirmation: assuring that a request was acted upon
3. Question: asking about a value, state, or an action
4. Goal sharing or status: create expectation of a desired future state
5. Case irrelevant: communication that does not relate to current situation

Communication variables were centered around the attending surgeon, in other words, only surgeon's audio and interactions were analyzed.

4.3.2 Statistical Analysis

T-tests were performed to identify differences in audio variables between pre-critical and post critical phase, and descriptive statistics for individual and team level variables were also analyzed.

4.4 Results

Due to the complexity and changeable of each surgical case, the docking and critical phase duration varied, Table 4.1 shows average ± standard deviation, and descriptive result of above durations. In comparison of 5 minutes before and 5 minutes after critical phase, the interquartile range of articulate speech resulted in the indication of the most significant difference. The remaining variables (individual level and team level variables) showed no difference of change with respect to the pre and post critical phase.

Table 4.1. Descriptive results of docking duration and critical phase duration of all surgical cases

Duration Type		Duration	Minimum	Max
Overall Duration (min)		194±77	90	339
	Docking (min)	9±3	4	14
	Critical Phase (min)	43±25	9	92

For frequency of communication, table shows the average ± standard deviation of each segmentation category.

Table 4.2. Mean frequency and SD of CLC with respect to segmentation categories

Segmentation Categories	Call-out	Check-back	Close-loop
Docking	31±12	25±14	4±2
10 minutes before un-docking	17±9	14±12	0.8±0.6
5 minutes before critical phase	10±6	13±10	0.5±1
Critical phase	61±34	32±21	6±5
5 minutes after critical phase	9±5	13±14	0.3±0.5

5. DISCUSSION

5.1 RQ I

Communication is widely known to impact clinical performance, and the findings of this study explored quantitative communication metrics that predict clinical performance in a team-based healthcare simulation. The results of the study support our first hypothesis that vocal features can be derived from audio data to further understand healthcare providers' communication behavior. The first approach distinguished communication by roles that initiated the dialogue, intent of the communication (i.e., statement or question), and by whether check-backs were observed was based on previous work in the clinical literature (Capella et al., 2010). Our results suggest that students tended to direct more questions to the nurse than the patient, which primarily included questions such as asking the nurse for the patient's current status and requesting equipment. However, the main communications from student to patient were statements, which primarily focused on stating the causes of the patient's symptoms and communicating the care plan. In addition to interprofessional communication, patient communication is another area of major focus for medical education, training, and practice. Studies have shown that it is important for patients to receive unambiguous statements from healthcare providers during the diagnostic process, appropriate treatments can then be conveyed precisely (Kripalani & Weiss, 2006). According to closed-loop communication, it ensures a clear understanding regarding clinical information and is shared among healthcare providers before conducting appropriate patient care (Miller, Riley, & Davis, 2009). Results showed conversations with no response from student to patient were commonly detected, which was likely the result of the variety of symptoms and conditions presented in the different scenarios. For example, repeated examination by the student may be causing severe pain to the patient, and thus resulted in no response from the patient.

In seeking additional insight into communication, our results from students' speech features showed that several audio metrics yielded a significant outcome in comparison between students to nurse and students to patient. Although the audio recordings revealed that the students spent less time talking to the patient than to the nurse, the difference in speech duration of two were not significant statically. The increase in student-nurse communication was attributed to the student gathering information and ordering laboratory examinations and additional tests from the nurse. Furthermore, speech ratio was calculated based on the speech duration of the student during each conversation segment, where higher speech ratio from student to patient could indicate less action and thinking time, and a smaller speech ratio from student to nurse could indicate that there were more hands-on examinations of the patient and time spent making clinical decisions. Previous studies found that speech rate differs by sex, age, dialect region, and speaker relationship; however, Schachter and colleagues found that when communicating with different person under the same topic, speech rate remains the same (Jiahong, Liberman, & Cieri, 2006; Schachter, Christenfeld, Ravina, & Bilous, 1991). Similarly, in this study, students' rates of speech remained consistent while communicating to a nurse or patient. Sex, age, and dialect region were not investigated since our study aimed to derive metrics directly from the audio, without the consideration of demographic factors. Another significant phenomenon that was observed was that students spoke with a higher volume to the patient than to nurse. In a study of patient satisfaction, a greater satisfaction was reported when physician was speaking in a higher volume to the patients (Mast, Hall, Klöckner, & Choi, 2008). It is unclear why this phenomenon was observed; however, we hypothesize that students were trying to communicate clearly with a patient who may be in physiological or psychological distress, or who they are not familiar with vs. a team member who they have worked with throughout their training. It is also possible that the lack of facial expressions of the manikin may have also influenced this student behavior.

Lastly, statistical methods were performed to identify the relationships between measured variables and performance scores. According to the indication of students' performance, all measured variables had a positive relationship with their scores. The frequency of communication with nurse and patient yielded a relatively high correlation with performance score (Figure 4) compared with other communication variables: e.g., speech duration, communication from student to nurse with check-backs. Students with a higher communication frequency tended to have higher performance scores because students asked questions to gather more information about the patient, to further identify the patient's symptoms with learned knowledge. Accordingly, statement-initiated conversations were later made by students to conduct corresponding treatments, which therefore resulted in a positive patient outcome. In addition to correlation, a predictive model that links these variables with performance was developed. The result from the stepwise regression model shows that there were nine relatively significant factors, resulting a R-squared value of 0.61. However, communications from student to nurse with check-backs weighted the highest in the regression model, which also verifies a strong correlation between frequency communication and performance score. From the regression model, it is shown that both question and statement events between student and nurse accounted for the largest portion compared with student and patient. Furthermore, all significant predictors of performance contained check backs, and thus communication with check backs was more predictive of performance. Future research should focus on the development of automated speech processing based on identifying speech signatures between different people and identify questions vs. statements.

Though clinical performance is primarily dependent on personal knowledge, communication is clearly a significant factor that contributes to clinical performance. Indeed, students' performance differed according to measured audio variables. Thus, clinical training and assessment can be supplemented by identifying objective audio variables. These variables may be useful in evaluating healthcare providers' communication in team-based simulations and the relationship of communication variables and healthcare providers' performance in the clinical setting.

5.2 RQ II

Communication is one of the nonnegligible factors in impacting clinical performance, and also acts as one of the pivotal roles in accessing NT skills. This study explored the novel features from a team-based healthcare simulation to predict NT skills performance, such quantitative communication metrics provided an objective and continuous method with comparison of current assessment tools. These metrics were divided into individual and team-based communication, and were based on previous work in the language processing and clinical literature (Capella et al., 2010; Gittell et al., 2000; Kao & Lee, n.d.). As advanced in information processing and speech recognition, such vocal features individual based communication can play a significant role in uncovering social interactions and emotion. In a study of automated depression diagnosis, Mantri et al. discovered that individual speech features such as speech intensity and pitch resulted in an accuracy of 78% in predicting depression performance (Shamla T. Mantri, Dipti D. Patil, Pankaj Agrawal, 2019). Our study categorized vocal features into various aspects from individual and team, and demonstrated such features that can be derived from raw audio recordings

Communication metrics can significantly predict NT skills and communication construct in a simulated performance in ACTS. It was found that significant metrics from both individual-level and team-level showed a positive related with overall NT skills and communication construct; average articulate speech rate (from speaker turn based variable under individual-level), average speech duration from student to patient (from communication dyad based variable under individual-level), and frequency of question initiated message with check-back from student to patient (from team-level variable) had the strongest correlation with both. In the findings from Gervits et al. (Gervits, Eberhard, & Scheutz, 2016), low articulate speech rate could result in team ineffectiveness, which was not focused on in this study, however, the construct of teamwork was consisted in the measure of overall NT skills. Furthermore, in Scherer et al.'s study, leaders and experts tend to have higher articulate speech rate compared with non-leaders and non-experts (S. Scherer, Weibel, Oviatt, & Morency, 2012). As for the correlation between articulated speech rate and communication construct, a higher rate could be explained by the students' well-preparation of the content and knowledge. However, this study contained a limited timeframe for the simulation, and an increase of articulate speech rate could also be affected by the time pressure, this was similar to Gervits et al.'s findings.

For communication dyad type communication and team-level variable, interactions between student and patient both took part in the most significant correlators as mentioned above: average speech duration from student to patient, and frequency of question-initiated message with check-back from student to patient. The more the student speaks to the patient, and the more questions directed to the patient with patient's response tend to result in a better overall NT skills and communication construct. This finding mirrored work in a social interaction setting by Yu et al. (Yu et al., 2016), that more frequent interactions and longer speaking duration could indicate a low team workload environment. Most importantly, higher frequency of communication showed an increased 'knowledge sharing' among the team (Bleakley, Allard, & Hobbs, 2013). The identified significant factors here could be used to provide better personalized feedback to students with low overall NT skills score and communication construct, such as more engaging in doctor-patient interaction, and more patient-centeredness generally associated with increased in patient satisfaction (Williams, 1998).

In addition to correlation, a predictive model that related these variables with overall NT skills and communication construct was developed respectively. There was a total of seven significant variables contributed to the overall NT skills model and resulted an adjusted R-squared value of 53.7% (Table 3). All predictors in the model were significant ($p<0.05$) except for the team-level variable of Frequency of communication from student to nurse statement with check-back, of the six significant individual-level variables, four were speaker-turn based variables and two were communication dyad variables. Lastly, a total of 10 significant variables contributed to the communication construct model resulting an adjusted R-squared of 54%. Majority of the significant predictors were speaker turn based (individual-level). In other words, both models in predicting overall NT skills and communication construct were individual-level variables oriented. And interestingly, vocal features categorized by communication target (nurse or patient) did not seem to be significant contributors. Nonetheless, this quantitative and predictive result provides a representation of which specific vocal features to be further implemented in clinical training, and improvements on NT skills performance in ACTS trainings.

Lastly, in the current study, where the NTS consists of communication, situational awareness, leadership, decision making, and teamwork. Although communication was found to be related to each construct (Sharma et al., 2011), however, for example, using communication to assess overall NTS in situations when leadership or teamwork isn’t present may be impartial. Future studies should link communication measurables with each individual construct, derive specific speech metrics (e.g., speech intensity, pitch, rate, etc.) to correlate with each individual construct, and further identify metrics that affect these constructs significantly. To further create an adjusted model based on selected constructs in predicting overall NTS to avoid inequitable results.

6. CONCLUSION AND FUTURE RESEARCH

Communication (verbal and non-verbal) is used from day to day, it is a well-established field and is generally defined as the behavior in the sense of reducing uncertainty from one to another (Buck, 1991). Communication is mostly seen in team-based settings, Hackman called this Input-Process-Output framework, where input represents individual knowledge among the team, processes usually include decision making, level of communication, and coordination, and output yields team performance. Nancy Cooke's study showed that improvements in team performance are accompanied by improvements in team process behaviors such as communication and coordination (Nancy J. Cooke, Gorman, & Kiekel, 2017). Speakers often use acoustic cues and prosodic features in information exchanging, including identities of speech sounds, phonemes, syllables, phrase boundaries, and other emotional states of the speaker (Santen, Mishra, & Klabbers, 2008). In the healthcare setting, more than 50% of the communication was achieved through verbal (including face-to-face, telephone, and paging) communication (Coiera, 2006; Eisenberg et al., 2005). Therefore, understanding and access communication through objective measures can better comprehend intrapersonal interactions, facilitate healthcare providers' communication skills, improve team performance and clinical training, and provide patient safety.

Overall, the findings from this study emphasized the use of objective communication metrics in understanding performances in ACTS. In the first study (RQ I), by analyzing interpersonal communication (question/statement-initiated message with/without check-back) among medical student, nurse, and patient during a simulated clinical scenario using audio recordings. Further, communication's vocal features (speech duration, intensity, ratio, and rate) were derived through audio processing. This study demonstrated the feasibility of quantifying communication features and patterns among participants. Importantly, the results showed a difference during communication from student to nurse and student to patient, and also the frequency of student communication with the nurse and patient predicted their clinical performance during the scenarios.

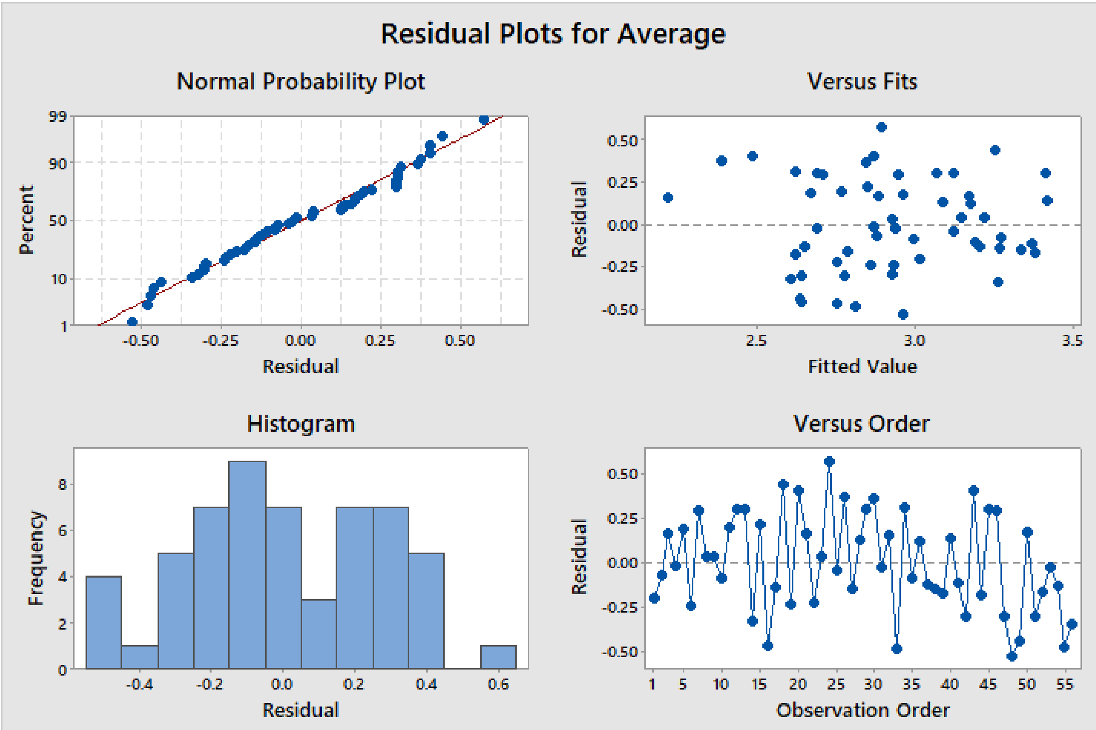
Findings from the second study (RQ II) provided evidence that audio vocal features and communication variables can be used to access NT skills (overall NT skills and communication construct). RQ II categorized variables into individual-level and team-level, and communication type was further defined as speaker turn based and communication dyad based on individual-level. with extensive vocal features, our second study was able to provide more insights into discover the relationship between communication and NT skills, and predicting the latter. The results from this study showed that average articulate speech rate (individual-level: speaker turn based variable), average speech duration from student to patient (individual-level: communication dyad based variable), and number of questions initiated messages from student to patient (team-level: frequency of communication) were all significantly and positively correlated with overall NT skills and communication construct. The majority of the significant factors in the model predicting overall NT skills and communication construct were speaker turn based variables, resulting in an R-squared value of 53.7% and 54%, respectively. Therefore, objective communication metrics can be used to assess healthcare providers' NT skills, and provide insights for further improvements in clinical training.

Studies from RQ I and RQ II showed the capabilities of relating one's speech prosodic features and performance skills (technical and non-technical). With results listed above, there's potential of providing an improved clinical education and training from the aspect of communication. The first study's (RQ I) approach was to explore fundamental speech prosodic features and deliver preliminary results linking with clinical performance; given the promising outcome from the first study, our second study (RQ II) included a more in depth exploration on the speech features on the basis of the first study, using such to relate speech features with NTS. Our last pilot study discussed above, took a step further to analyze speech features in a real clinical setting (OR), identifying the communication patterns of the surgical team as a whole. With the research originating from a simulated environment to actual hospital setting, the goal is to provide an objective viewpoint of evaluating future and current healthcare providers' performance.

Lastly, the findings from this research provided some fundamental communication measures and metrics for future research in developing automation in accessing communication in a healthcare setting. Given the current analysis methods that were time-consuming, with automated technology such vocal features and communication metrics can be accessed in real-time, and provide detail feedbacks to individual healthcare providers, or even provide a comparison of performance (Ryan et al., 2019). Nevertheless, future research should also explore speech disfluency (e.g., number of pauses, duration of pauses), speech interjection, and how to measure hesitancy through speech. Further, this research was conducted in a simulated clinical setting, and future researchers should explore the transferability from simulated to an actual clinical setting.

APPENDIX A. REGRESSION MODEL ASSUMPTIONS

Below figure shows the data normality, constant variance, skewness, and randomness of the stepwise regression model.



LIST OF REFERENCES

Abbas, N. F., Jawad, R. T., & Muhi, M. T. (2018). Pauses and Hesitations in Drama Texts. *International Journal of English Linguistics*, *8*(4), 106. https://doi.org/10.5539/ijel.v8n4p106

Albanese, M. A. (2000). Challenges in using rater judgements in medical education. *Journal of Evaluation in Clinical Practice*. https://doi.org/10.1046/j.1365-2753.2000.00253.x

Albrecht, G. (1996). Using Subjective Health Assessments in Practice and Policy-making. In *CARE ANALYSIS* (Vol. 4). Retrieved from https://link.springer.com/content/pdf/10.1007%2FBF02249315.pdf

Anca, J., Helmreich, R. L., & Kanki, B. G. (2010). Crew Resource Management. In *Intensiv*. https://doi.org/10.1055/s-0029-1241139

Baker D.P., Gustafson S., Beaubien J.M., Salas E., B. P. (2005). Medical Teamwork and Patient Safety: The Evidence-based Relation. Literature Review. *Agency for Healthcare Research and Quality Publication No. 05-0053*.

Bleakley, A., Allard, J., & Hobbs, A. (2013). 'Achieving ensemble': communication in orthopaedic surgical teams and the development of situation awareness—an observational study using live videotaped examples. *Advances in Health Sciences Education*, *18*(1), 33–56. https://doi.org/10.1007/s10459-012-9351-6

Boersma, Paul & Weenink, D. (2018). *Praat: doing phonetics by computer [Computer program]*. Retrieved from http://www.praat.org/

Bourhis, R. Y., Roth, S., & MacQueen, G. (1989). Communication in the hospital setting: A survey of medical and everyday language use amongst patients, nurses and doctors. *Social Science and Medicine*. https://doi.org/10.1016/0277-9536(89)90035-X

Brennan, S. E., & Williams, M. (1995). The Feeling of Another′s Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language*, *34*(3), 383–398. https://doi.org/10.1006/jmla.1995.1017

Buck, R. (1991). Temperament, Social Skills, and the Communication of Emotion. In *Personality, Social Skills, and Psychopathology* (pp. 85–105). https://doi.org/10.1007/978-1-4899-0635-9_4

Buck, R., & VanLear, C. A. (2002). Verbal and nonverbal communication: Distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. *Journal of Communication*. https://doi.org/10.1093/joc/52.3.522

Capella, J., Smith, S., Philp, A., Putnam, T., Gilbert, C., Fry, W., … Remine, S. (2010). Teamwork training improves the clinical care of trauma patients. *Journal of Surgical Education*. https://doi.org/10.1016/j.jsurg.2010.06.006

Cha, J., Anton, N., Mizota, T., Hennings, J., Rendina, M., Stanton-Maxie, K., Ritter, H., Stefanidis, D., and Yu, D. (n.d.). Use of Non-Technical Skills Can Predict Medical Student Performance in Acute Care Simulated Scenarios. *American Journal of Surgery*.

Cha, J. S., Anton, N. E., Mizota, T., Hennings, J. M., Rendina, M. A., Stanton-Maxey, K., … Yu, D. (2019). Use of non-technical skills can predict medical student performance in acute care simulated scenarios. *American Journal of Surgery*. https://doi.org/10.1016/j.amjsurg.2018.09.028

Clapper, T. C., & Kong, M. (2012). TeamSTEPPS ®: The Patient Safety Tool That Needs to Be Implemented. *Clinical Simulation in Nursing*. https://doi.org/10.1016/j.ecns.2011.03.002

Coiera, E. (2006). Communication systems in healthcare. *The Clinical Biochemist. Reviews / Australian Association of Clinical Biochemists*.

Coiera, E., & Tombs, V. (1998). Communication behaviors in a hospital setting: an observational study. *BMJ British Medical Journal*. https://doi.org/10.1136/bmj.38638.497477.68

Commission, J. (2015). Sentinel Event Statistics Released Through Second Quarter 2015. *Joint Commission Online*.

Cooke, N. J., Gorman, J. C., & Kiekel, P. A. (2008). Communication as Team-Level Cognitive Processing. In *Macrocognition in Teams: Theories and Methodologies* (pp. 51–64). Retrieved from https://content.taylorfrancis.com/books/e/download?dac=C2015-0-90514-8&isbn=9781315593166&doi=10.1201/9781315593166-4&format=pdf

Cooke, Nancy J., Gorman, J. C., & Kiekel, P. A. (2017). Communication as team-level cognitive processing. In *Macrocognition in Teams: Theories and Methodologies*. https://doi.org/10.1201/9781315593166

Cooper, J. B., Newbower, R. S., Long, C. D., & McPeek, B. (1978). Preventable anesthesia mishaps: A study of human factors. *Anesthesiology*. https://doi.org/10.1097/00000542-197812000-00004

Crichton, M. ., & Flin, R. (2004). Identifying and training non-technical skills of nuclear emergency response teams. *Annals of Nuclear Energy*, *31*(12), 1317–1330. https://doi.org/10.1016/J.ANUCENE.2004.03.011

CRICO Strategies. (2015). Malpractice Risks in Communication Failures 2015 Annual Benchmarking Report.

Cuschieri, A., Francis, N., Crosby, J., & Hanna, G. B. (2001). What do master surgeons think of surgical competence and revalidation? *American Journal of Surgery*. https://doi.org/10.1016/S0002-9610(01)00667-5

Downing, S. M. (2005). Threats to the validity of clinical teaching assessments: What about rater error? *Medical Education*. https://doi.org/10.1111/j.1365-2929.2005.02138.x

Eisenberg, E. M., Murphy, A. G., Sutcliffe, K., Wears, R., Schenkel, S., Perry, S., & Vanaerhoef, M. (2005). Communication in emergency medicine: Implications for patient safety. *Communication Monographs*, *72*(4), 390–413. https://doi.org/10.1080/03637750500322602

Ellingson, L. L. (2002). Communication , Collaboration , and Teamwork among Health Care Professionals. *Communication Research Trends*.

Endsley, M. (1995). Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32-64. In *Human Factors: The Journal of the Human Factors and Ergonomics Society*. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*. https://doi.org/10.1518/001872095779049543

Fairbanks, R. J., Bisantz, A. M., & Sunm, M. (2007). Emergency Department Communication Links and Patterns. *Annals of Emergency Medicine*. https://doi.org/10.1016/j.annemergmed.2007.03.005

Flin, R., Martin, L., Goeters, K.-M., Hörmann, H.-J., Amalberti, R., Valot, C., & Nijhuis, H. (2019). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. In *Human Factors and Aerospace Safety*. https://doi.org/10.4324/9781315194035-1

Flin, R., O'Connor, P., & Crichton, M. (2007). Safety at the Sharp End. In *Safety at the Sharp End*. https://doi.org/39/1/1 [pii]

Flin, R., O'Connor, P., & Crichton, M. (2013). Safety at the sharp end: A guide to non-technical skills. In *Safety at the Sharp End: A Guide to Non-Technical Skills*.

Fore, A. M., & Sculli, G. L. (2013). A concept analysis of situational awareness in nursing. *Journal of Advanced Nursing*. https://doi.org/10.1111/jan.12130

Gervits, F., Eberhard, K., & Scheutz, M. (2016). Team Communication as a Collaborative Process. *Frontiers in Robotics and AI*, *3*, 62. https://doi.org/10.3389/frobt.2016.00062

Gittell, J. H., Fairfield, K. M., Bierbaum, B., Head, W., Jackson, R., Kelly, M., … Zuckerman, J. (2000). Impact of Relational Coordination on Quality of Care, Postoperative Pain and Functioning, and Length of Stay: A Nine-Hospital Study of Surgical Patients. In *Care* (Vol. 38). Retrieved from http://ipls.dk/pdf-filer/gittell_2000.pdf

Gorman, J. C., Foltz, P. W., Kiekel, P. a., Martin, M. J., & Cooke, N. J. (2003). Evaluation of Latent Semantic Analysis-Based Measures of Team Communications Content. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/10.1177/154193120304700337

Grumbach, K., & Bodenheimer, T. (2004). Can health care teams improve primary care practice? *JAMA: The Journal of the American Medical Association*. https://doi.org/10.1001/jama.291.10.1246

Gugerty, L. J. (1997). Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/1076-898X.3.1.42

Hackman, J. R. (1987). The design of work teams. *Handbook of Organizational Behavior*. https://doi.org/10.1016/0090-2616(78)90031-1

Härgestam, M., Lindkvist, M., Brulin, C., Jacobsson, M., & Hultin, M. (2013). Communication in interdisciplinary teams: Exploring closed-loop communication during in situ trauma team training. *BMJ Open*, *3*(10). https://doi.org/10.1136/bmjopen-2013-003525

Helmreich, R. L., & Foushee, H. C. (2010). Why CRM? Empirical and Theoretical Bases of Human Factors Training. In *Crew Resource Management*. https://doi.org/10.1016/B978-0-12-374946-8.10001-9

Howlett, M., Doody, K., Murray, J., LeBlanc-Duchin, D., Fraser, J., & Atkinson, P. R. (2015). Burnout in emergency department healthcare professionals is associated with coping style: A cross-sectional survey. *Emergency Medicine Journal*. https://doi.org/10.1136/emermed-2014-203750

Hull, L., Arora, S., Aggarwal, R., Darzi, A., Vincent, C., & Sevdalis, N. (2012). The impact of nontechnical skills on technical performance in surgery: A systematic review. *Journal of the American College of Surgeons*. https://doi.org/10.1016/j.jamcollsurg.2011.10.016

Hull, L., Arora, S., Kassab, E., Kneebone, R., & Sevdalis, N. (2011). Observational teamwork assessment for surgery: Content validation and tool refinement. *Journal of the American College of Surgeons*. https://doi.org/10.1016/j.jamcollsurg.2010.11.001

Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., & Leino, T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Applied Ergonomics*. https://doi.org/10.1016/j.apergo.2010.08.005

Jacklin, R., Sevdalis, N., Darzi, A., & Vincent, C. (2008). Mapping surgical practice decision making: an interview study to evaluate decisions in surgical care. *American Journal of Surgery*. https://doi.org/10.1016/j.amjsurg.2007.02.016

Jiahong, Y., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*.

Kao, Y.-H., & Lee, L.-S. (n.d.). *Feature Analysis for Emotion Recognition from Mandarin Speech Considering the Special Characteristics of Chinese Language*. Retrieved from https://www.isca-speech.org/archive/archive_papers/interspeech_2006/i06_1504.pdf

King, H. B., Battles, J. B., Baker, D. P., Alonso, A., Salas, E., Webster, J., … Grady, M. L. (2008). TeamSTEPPS: Team Strategies and Tools to Enhance Performance and Patient Safety Advances. *Patient Safety: New Directions and Alternative Approaches (Vol. 3: Performance and Tools)*. https://doi.org/NBK43686 [bookaccession]

Kripalani, S., & Weiss, B. D. (2006). Teaching about health literacy and clear communication. *Journal of General Internal Medicine*, Vol. 21, pp. 888–890. https://doi.org/10.1111/j.1525-1497.2006.00543.x

Leonard, M., Graham, S., & Bonacum, D. (2004). The human factor: The critical importance of effective teamwork and communication in providing safe care. *Quality and Safety in Health Care*, Vol. 13. https://doi.org/10.1136/qshc.2004.010033

Linda T. Kohn, Janet M. Corrigan, and M. S. D. (1999). To Err Is Human. Building a Safer Health System, Volume 6. In *International Journal of Public Health*. https://doi.org/10.17226/9728

Lingard, L., Reznick, R., Espin, S., Regehr, G., & DeVito, I. (2002). Team communications in the operating room: Talk patterns, sites of tension, and implications for novices. *Academic Medicine*. https://doi.org/10.1097/00001888-200203000-00013

Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: A review of the literature. *Acta Anaesthesiologica Scandinavica*. https://doi.org/10.1111/j.1399-6576.2008.01717.x

Martin, J. S., Ummenhofer, W., Manser, T., & Spirig, R. (2010). Interprofessional collaboration among nurses and physicians: making a difference in patient outcome. *Swiss Medical Weekly*. https://doi.org/smw-12648

Mast, M. S., Hall, J. A., Klöckner, C., & Choi, E. (2008). Physician gender affects how physician nonverbal behavior is related to patient satisfaction. *Medical Care*. https://doi.org/10.1097/MLR.0b013e31817e1877

Mayer, C. M., Cluff, L., Lin, W.-T., Willis, T. S., Stafford, R. E., Williams, C., … Amoozegar, J. B. (2011). Evaluating Efforts to Optimize TeamSTEPPS Implementation in Surgical and Pediatric Intensive Care Units. *The Joint Commission Journal on Quality and Patient Safety*. https://doi.org/10.1016/S1553-7250(11)37047-X

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*. https://doi.org/10.11613/BM.2012.031

Mercer, L. M., Tanabe, P., Pang, P. S., Gisondi, M. A., Courtney, D. M., Engel, K. G., … Makoul, G. (2008). Patient perspectives on communication with the medical team: Pilot study using the communication assessment tool-team (CAT-T). *Patient Education and Counseling*. https://doi.org/10.1016/j.pec.2008.07.003

Miller, K., Riley, W., & Davis, S. (2009). Identifying key nursing and team behaviours to achieve high reliability. *Journal of Nursing Management*. https://doi.org/10.1111/j.1365-2834.2009.00978.x

Minitab, I. (2000). MINITAB statistical software. *Minitab Release*.

Moorthy, K., Munz, Y., Adams, S., Pandey, V., & Darzi, A. (2005). A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. *Annals of Surgery*. https://doi.org/10.1097/01.sla.0000186298.79308.a8

Morozova, G., Martindale, A., & Currie, I. (2017, April 10). *Identifying the cognitive demands on experts' decision making in liver transplantation* (pp. 241–246). pp. 241–246. Retrieved from https://www.research.ed.ac.uk/portal/en/publications/identifying-the-cognitive-demands-on-experts-decision-making-in-liver-transplantation(42ec5677-3f5f-4cea-a34a-4ae46d2c26ca).html

Odell, M. M. (2011). Human factors and patient safety: Changing roles in critical care. *Australian Critical Care*, *24*, 215–217. https://doi.org/10.1016/j.aucc.2011.02.001

Olguin, O., Gloor, P. A., Pentland, A., & Olguin Olguin, D. (2009). *Wearable sensors for pervasive healthcare management*. https://doi.org/10.4108/ICST.PERVASIVEHEALTH2009.6033

Onnela, J. P., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to quantify social interaction patterns. *Scientific Reports*. https://doi.org/10.1038/srep05604

Peng, Y., Anton, N. E., Cha, J., Mizota, T., Hennings, J. M., Stambro, R., … Yu, D. (2019). Objective Measures of Communication Behavior Predict Clinical Performance. *Journal of Surgical Education*. https://doi.org/10.1016/J.JSURG.2019.03.017

Rabøl, L. I., Andersen, M. L., Østergaard, D., Bjørn, B., Lilja, B., & Mogensen, T. (2011). Republished error management:Descriptions of verbal communication errors between staff. an analysis of 84 root cause analysis-reports from danish hospitals. *Postgraduate Medical Journal*. https://doi.org/10.1136/pgmj.2010.040238rep

Reagans, R., & Zuckerman, E. W. (2001). Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. *Organization Science*. https://doi.org/10.1287/orsc.12.4.502.10637

RICHARDS, E. (LIBBY), SIMPSON, V., AALTONEN, P., KREBS, L., & DAVIS, L. (2010). Public Health Nursing Student Home Visit Preparation. *Home Healthcare Nurse: The Journal for the Home Care and Hospice Professional*, *28*(10), 631–638. https://doi.org/10.1097/NHH.0b013e3181f85e10

Rosen, M. A., Dietz, A. S., Lee, N., Jeng Wang, I., Markowitz, J., Wyskiel, R. M., … Pronovost, P. J. (2018). Sensor-based measurement of critical care nursing workload: Unobtrusive measures of nursing activity complement traditional task and patient level indicators of workload to predict perceived exertion. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0204819

Ryan, P., Luz, S., Albert, P., Vogel, C., Normand, C., & Elwyn, G. (2019). Using artificial intelligence to assess clinicians' communication skills. *BMJ (Online)*. https://doi.org/10.1136/bmj.l161

Santen, J. van, Mishra, T., & Klabbers, E. (2008). Prosodic Processing. In *Springer Handbook of Speech Processing* (pp. 471–488). https://doi.org/10.1007/978-3-540-49127-9_23

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.60.3.362

Scherer, S., Weibel, N., Oviatt, S., & Morency, L.-P. (2012). *Multimodal prediction of expertise and leadership in learning groups*. Retrieved from http://www.anoto.com

Scherer, Y. K., Bruce, S. A., & Runkawatt, V. (2007). A comparison of clinical simulation and case study presentation on nurse practitioner students' knowledge and confidence in managing a cardiac event. *International Journal of Nursing Education Scholarship, 4*(1). https://doi.org/10.2202/1548-923X.1502

Schmutz, J., & Manser, T. (2013). Do team processes really have an effect on clinical performance? A systematic literature review. *British Journal of Anaesthesia*. https://doi.org/10.1093/bja/aes513

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing Cockpit Communications: The Links Between Language, Performance, Error, and Workload. *Journal of Human Performance in Extreme Environments*. https://doi.org/10.7771/2327-2937.1007

Shamla T. Mantri, Dipti D. Patil, Pankaj Agrawal, V. M. W. (2019). Real Time Multimodal Depression Analysis. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(9). Retrieved from https://www.researchgate.net/profile/Dipti_Patil5/publication/334645363_Real_time_multimodal_depression_analysis_IJITEE/links/5d37f3d8299bf1995b453ce0/Real-time-multimodal-depression-analysis-IJITEE.pdf

Sharma, B., Mishra, A., Aggarwal, R., & Grantcharov, T. P. (2011). Non-technical skills assessment in surgery. *Surgical Oncology*. https://doi.org/10.1016/j.suronc.2010.10.001

Somech, A. (2006). The effects of leadership style and team process on performance and innovation in functionally heterogeneous teams. *Journal of Management*. https://doi.org/10.1177/0149206305277799

Tang, P. C., Jaworski, M. A., Fellencer, C. A., Kreider, N., LaRosa, M. P., & Marquardt, W. C. (1996). Clinician information activities in diverse ambulatory care practices. *Proceedings : A Conference of the American Medical Informatics Association / ... AMIA Annual Fall Symposium. AMIA Fall Symposium*.

Tawalbeh, L. I., & Tubaishat, A. (2014). Effect of simulation on knowledge of advanced cardiac life support, knowledge retention, and confidence of nursing students in Jordan. *Journal of Nursing Education*, *53*(1), 38–44. https://doi.org/10.3928/01484834-20131218-01

Tiel Groenestege-Kreb, D., Van Maarseveen, O., & Leenen, L. (2014). Trauma team. *British Journal of Anaesthesia*. https://doi.org/10.1093/bja/aeu236

Tiffen, J., Graf, N., & Corbridge, S. (2009). Effectiveness of a Low-fidelity Simulation Experience in Building Confidence among Advanced Practice Nursing Graduate Students. *Clinical Simulation in Nursing*, *5*(3), e113–e117. https://doi.org/10.1016/j.ecns.2009.01.009

Tolkmitt, F. J., & Scherer, K. R. (1986). Effect of Experimentally Induced Stress on Vocal Parameters. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.12.3.302

Toner, E. (2009). Creating situational awareness: A systems approach. *Workshop on Medical Surge Capacity, Institute of …*.

Turkelson, C., Aebersold, M., & Sculli, G. (2013). Board 262 - Program Innovations Abstract Improving Nursing Communication Skills In An Intensive Care Unit Using Simulation and Nursing Crew Resource Management (NCRM) Strategies (Submission #1200). *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*. https://doi.org/10.1097/01.sih.0000441527.86805.b9

van Reekum, C. M., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. R. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition and Emotion*. https://doi.org/10.1080/02699930341000167

Watch your tone | MIT News. (n.d.).

Weaver, S. J., Rosen, M. A., Diazgranados, D., Lazzara, E. H., Lyons, R., Salas, E., … King, H. B. (2010). Does teamwork improve performance in the operating room? A multilevel evaluation. *Joint Commission Journal on Quality and Patient Safety*. https://doi.org/10.1016/S1553-7250(10)36022-3

Williams, S. (1998). Doctor-patient communication and patient satisfaction: a review. *Family Practice*. https://doi.org/10.1093/fampra/15.5.480

Wong, C. A., Cummings, G. G., & Ducharme, L. (2013). The relationship between nursing leadership and patient outcomes: A systematic review update. *Journal of Nursing Management*. https://doi.org/10.1111/jonm.12116

Xiao, Y., Seagull, F. J., Mackenzie, C., Ziegert, J., & Klein, K. J. (2003). Team Communication Patterns as Measures of Team Processes: Exploring the Effects of Task Urgency and Shared Team Experience. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/10.1177/154193120304701228

Yu, D., Blocker, R. C., Hallbeck, S., Patel, A. M., & Pasupathy, K. S. (2015). Wearable sociometers: Calibration in Chaotic Simulated Environments. *Journal of Medical Devices*, *9*(2), 020946. https://doi.org/10.1115/1.4030150

Yu, D., Blocker, R. C., Sir, M. Y., Hallbeck, M. S., Hellmich, T. R., Cohen, T., … Pasupathy, K. S. (2016). Intelligent Emergency Department: Validation of Sociometers to Study Workload. *Journal of Medical Systems*. https://doi.org/10.1007/s10916-015-0405-1

Yu, D., Minter, R. M., Armstrong, T. J., Frischknecht, A. C., Green, C., & Kasten, S. J. (2014). Identification of technique variations among microvascular surgeons and cases using hierarchical task analysis. *Ergonomics*. https://doi.org/10.1080/00140139.2014.884244

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*. https://doi.org/10.1016/j.surg.2005.06.017

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006). Development of a rating system for surgeons' non-technical skills. *Medical Education*. https://doi.org/10.1111/j.1365-2929.2006.02610.x