



Published in final edited form as:

Hamlyn Symp Med Robot. 2025 June ; 2025: 17–18.

Voice Acoustic Patterns Predict Quality of Interprofessional Team Behavior in Cardiac Surgery

Sanjana Mendu^{1,2}, Shrivatsa Mishra³, Victor Murcia Ruiz⁴, Rafael Fricks⁴, Rayan Harari⁵, Roger D. Dias⁵, Theodora Chaspari³, Marco A. Zenati^{1,2}

¹Medical Robotics & Computer Assisted Surgery (MRCAS) Laboratory, Division of Cardiac Surgery, Veterans Affairs Boston Healthcare System, Boston, MA, USA

²Division of Cardiac Surgery, Brigham and Women's Hospital, Mass General Brigham, Harvard Medical School, Boston, MA, USA

³Massachusetts Veterans Epidemiology Research and Information Center, Boston, MA

⁴Computer Science & Institute of Cognitive Science, University of Colorado Boulder, USA

⁵Department of Emergency Medicine, Mass General Brigham, Harvard Medical School, Boston, MA, USA.

INTRODUCTION

Intraoperative adverse events (IAEs) are a significant unresolved problem in surgery. Over 60% of IAEs are attributed to poor interprofessional team behavior (ITB) [1]. Surgical team members' voice acoustic patterns can serve as indicators of positive and negative ITB. We have demonstrated significant differences in speech patterns across examples of good and poor interprofessional team behavior in a simulated cardiac surgery setting [2]. Identifying vocal behaviors that affect team dynamics in real-time may yield unique insights towards proactively addressing communication challenges, optimizing collaboration, and improving the safety and efficiency of surgical procedures. This paper investigates the relationship between voice acoustic patterns and ratings of ITB using data collected during real-world cardiac surgery procedures.

MATERIALS AND METHODS

Audio recordings were prospectively collected for 28 cardiac surgery procedures conducted at a tertiary academic medical center using a customized portable observation lab (Mangold International GmbH), which included 4 wireless microphones and a wall-mounted camera (Axis Communications) capturing wide fields of view of the operating theatre (OT). Wireless microphones (Sennheiser Electronic GmbH) were placed near or worn by primary team members to capture verbal communications within each subteam.

Subjective ratings of the quality of nontechnical skills exhibited by surgical teams were obtained using the Non-Technical Skills for Surgeons (NOTSS) behavioral rating system

[3]. Three trained raters assessed teams' overall nontechnical skills along 12 behavioral measures grouped in the following 4 categories: situational awareness, decision-making, communication and teamwork, and leadership. Each behavioral measure was rated on a Likert scale from 1 (poor) to 4 (good). Behavioral ratings were averaged within each category to obtain aggregate scores. Given the significant time and resources required to obtain reliable NOTSS ratings, the current analysis focuses specifically on the separation from bypass phase of surgery.

To investigate speaker-specific acoustic patterns, we applied a basic speaker identification model which uses minimally labeled examples of surgeons' speech to identify unlabeled surgeon speaking turns. A stationary spectral gating algorithm [4] was applied to remove noise artifacts. An energy-based voice activity detection algorithm [5] was applied to each audio file to separate segments of human speech from background noise. A member of the research team manually annotated surgeon speaking turns during the first two minutes of the separation from bypass phase of surgery for each case (24.4% of total phase duration on average), resulting in a total of 13.5 minutes of annotations across all cases. Voice activity segments were then cross-referenced with manual labels of surgeon speaking turns to obtain references for instances where the surgeon is speaking and instances when another individual is speaking.

To identify unlabelled instances of surgeon speaking turns, we trained a 2-component Gaussian Mixture Model (GMM) using 13 mel-frequency cepstral coefficients (MFCCs) extracted from each of the labeled examples. The speaker identification model achieved a balanced training accuracy of $83.7 \pm 8.7\%$ on the labelled data. Voice activity segments were classified into surgeon and other speaking turns over 120ms time windows with 50% overlap and majority voting was used to combine speaker labels across overlapping time windows.

Acoustic measures were extracted using the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) [6], which includes 88 acoustic measures of affect, covering frequency, energy, and spectral parameters. A linear mixed-effects (LME) modeling approach was used to measure differences in voice acoustic patterns with respect to the NOTSS subdimension scores. The LME models accounted for variations in speech patterns between the surgeon and other surgical team members by including the speaker as a random effect. Each NOTSS subdimension was included as fixed effects such that the association between each acoustic measure and each NOTSS subdimension score was computed independently.

RESULTS

Results of the LME analysis revealed significant linear relationships between 69 out of the 88 acoustic measures (78.4%) and NOTSS scores (Fig 1). The mean and standard deviation of scores for each NOTSS category were: situation awareness (3.33 ± 0.30), decision making (3.37 ± 0.31), communication and teamwork (3.37 ± 0.30), leadership (3.36 ± 0.31).

Mean fundamental frequency (F0), which measures vocal pitch, was higher for teams that received higher ratings of leadership ($\beta=4.08$, $p<.01$) as well as communication and

teamwork ($\beta=5.94$, $p<.01$). However, high mean F0 also corresponded to lower decision making scores ($\beta=-7.83$, $p<.01$). Loudness (perceived speech intensity) was higher for teams with higher ratings of situation awareness ($\beta=9.41$, $p<.01$), leadership ($\beta=9.74$, $p<.01$) and communication and teamwork ($\beta=5.86$, $p<.01$), but lower ratings of decision making ($\beta=-14.95$, $p<.01$). Higher levels of shimmer (perceived trembling in speech intensity) corresponded to higher ratings of communication and teamwork ($\beta=2.88$, $p<.01$) but lower ratings of leadership ($\beta=-4.13$, $p<.01$). Similarly, higher levels of jitter (temporal deviation in speech periodicity) corresponded to higher ratings of communication and teamwork ($\beta=3.94$, $p<.01$) but lower ratings of leadership ($\beta=-2.19$, $p=.03$).

Increase in alpha ratio (ratio of the summed speech energy from low frequency of 50–1000 Hz and high frequency of 1–5 kHz) corresponded to higher ratings of decision making ($\beta=6.29$, $p<.01$) and lower ratings of situation awareness ($\beta=-4.46$, $p<.01$) and leadership ($\beta=-2.55$, $p=.01$). Higher levels of spectral flux (temporal difference in spectral content between two consecutive speech frames) corresponded to higher ratings of situation awareness ($\beta=14.16$, $p<.01$), communication and teamwork ($\beta=5.40$, $p<.01$), and leadership ($\beta=7.82$, $p<.01$), but lower ratings of decision making ($\beta=-16.11$, $p<.01$).

DISCUSSION

These findings indicate a strong relationship between voice acoustic patterns and quality of ITB in cardiac surgery. By accounting for variations in speech patterns between surgeons and other surgical team members, these findings highlight the relationship between individual speech patterns and intraoperative team dynamics. Furthermore, the variation in directionality of relationships between acoustic measures and ITB quality suggest that speech patterns can provide a nuanced understanding of different aspects of surgical teamwork. For example, results suggest that emotional arousal, reflected in elevated F0 and loudness, may play a crucial role in sustaining communication, leadership, and teamwork, while potentially hindering effective decision-making. Similarly, variations in frequency and amplitude of speech, reflected in elevated jitter and shimmer, may point to better communication and teamwork but less effective leadership. Further analyses are needed to disaggregate acoustic measures for team members beyond the attending surgeon towards more robust modeling of inter-individual differences in acoustics. Additionally, linguistic analysis of intraoperative communication could provide complementary insights towards a more comprehensive understanding of interprofessional dynamics in cardiac surgery.

Acknowledgment:

This work was supported by the National Institute of Health (NHLBI - R01HL126896 and R01HL157457) and the National Science Foundation (NSF - IIS2310187).

REFERENCES

- [1]. Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, and Sundt TM, "Disruptions in surgical flow and their relationship to surgical errors: An exploratory investigation," *Surgery*, vol. 142, no. 5, pp. 658–665, Nov. 2007, doi: 10.1016/j.surg.2007.07.034. [PubMed: 17981185]

- [2]. Mishra S, Dias RD, Zenati MA, and Chaspari T, "Acoustic Patterns of Interprofessional Communication and Quality of Teamwork in the Cardiac Operating Theatre," *Hamlyn Symp Med Robot*, vol. 16, pp. 103–104, Jun. 2024. [PubMed: 39076410]
- [3]. Yule S, Flin R, Maran N, Rowley D, Youngson G, and Paterson-Brown S, "Surgeons' Non-technical Skills in the Operating Room: Reliability Testing of the NOTSS Behavior Rating System," *World j. surg*, vol. 32, no. 4, pp. 548–556, Apr. 2008, doi: 10.1007/s00268-007-9320-z. [PubMed: 18259809]
- [4]. Sainburg T, timsainb/noisereducer: v1.0. (Jun. 11, 2019). Zenodo. doi: 10.5281/zenodo.3243139.
- [5]. vad: VAD - Simple Voice Activity Detection. Python. Accessed: Apr. 01, 2025. [OS Independent]. Available: <https://github.com/MorenoLaQuatra/vad>
- [6]. Eyben F. et al. , "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," *IEEE Trans. Affective Comput*, vol. 7, no. 2, pp. 190–202, Apr. 2016, doi: 10.1109/TAFFC.2015.2457417.

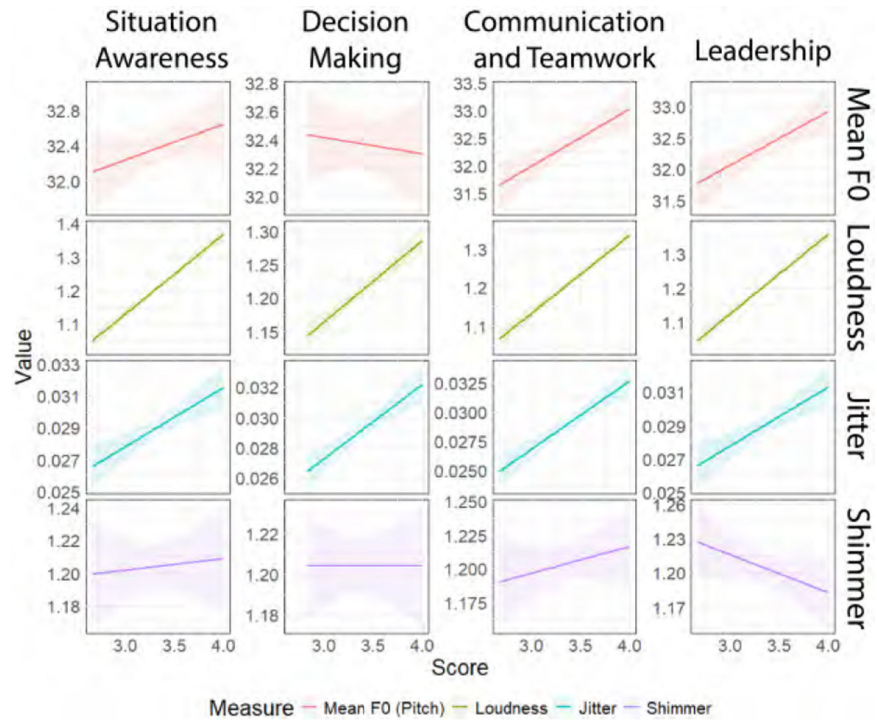


Fig. 1. Relationship between acoustic measures and NOTSS sub-dimension scores.