Acoustic analysis of surgeons' voices to assess change in the stress response during surgical in situ simulation

Andrew Hall ⓘ ,[1] Kosuke Kawai,[2] Kelsey Graber,[3] Grant Spencer,[2] Christopher Roussin,[4] Peter Weinstock,[3] Mark S Volk[5]

[1]Department of Otolaryngology, University Hospital of Wales and Noah's Ark Children's Hospital for Wales, Cardiff, UK
[2]Department of Otolaryngology and Communication Enhancement, Boston Children's Hospital, Boston, Massachusetts, USA
[3]Boston Children's Hospital Simulator Program, Boston Children's Hospital, Boston, Massachusetts, USA
[4]Simulator Program, Boston Children's Hospital, Boston, Massachusetts, USA
[5]Otolaryngology, Boston Children's Hospital, Boston, Massachusetts, USA

Correspondence to
Mr Andrew Hall, University Hospital of Wales, Noah's Ark Children's Hospital for Wales, Cardiff, London, UK; andrew.hall2@wales.nhs.uk

Check for updates

ABSTRACT

Introduction Stress may serve as an adjunct (challenge) or hindrance (threat) to the learning process. Determining the effect of an individual's response to situational demands in either a real or simulated situation may enable optimisation of the learning environment. Studies of acoustic analysis suggest that mean fundamental frequency and formant frequencies of voice vary with an individual's response during stressful events. This hypothesis is reviewed within the otolaryngology (ORL) simulation environment to assess whether acoustic analysis could be used as a tool to determine participants' stress response and cognitive load in medical simulation. Such an assessment could lead to optimisation of the learning environment.

Methodology ORL simulation scenarios were performed to teach the participants teamwork and refine clinical skills. Each was performed in an actual operating room (OR) environment (in situ) with a multidisciplinary team consisting of ORL surgeons, OR nurses and anaesthesiologists. Ten of the scenarios were led by an ORL attending and ten were led by an ORL fellow. The vocal communication of each of the 20 individual leaders was analysed using a long-term pitch analysis PRAAT software (autocorrelation method) to obtain mean fundamental frequency (F0) and first four formant frequencies (F1, F2, F3 and F4). In reviewing individual scenarios, each leader's voice was analysed during a non-stressful environment (WHO sign-out procedure) and compared with their voice during a stressful portion of the scenario (responding to deteriorating oxygen saturations in the manikin).

Results The mean unstressed F0 for the male voice was 161.4 Hz and for the female voice was 217.9 Hz. The mean fundamental frequency of speech in the ORL fellow (lead surgeon) group increased by 34.5 Hz between the scenario's baseline and stressful portions. This was significantly different to the mean change of −0.5 Hz noted in the attending group (p=0.01). No changes were seen in F1, F2, F3 or F4.

Conclusions This study demonstrates a method of acoustic analysis of the voices of participants taking part in medical simulations. It suggests acoustic analysis of participants may offer a simple, non-invasive, non-intrusive adjunct in evaluating and titrating the stress response during simulation.

INTRODUCTION

One of the purported advantages of experiential learning, such as that encountered in medical

What is already known on this subject

► The ability of an individual to learn is affected by their response to external demands they receive from their environment. Individuals presented with a set of external demands will respond by entering into a challenge, versus a threat (stress) psychological state. Being in a challenged state is correlated with enhanced educational performance.
► A variety of subjective and objective measurements have been used in an attempt to determine, and therefore titrate, the psychological state (challenge vs threat) of an individual.
► Fundamental frequency of voice has been shown to vary as part of the physiological stress response. However, it has not previously been used to assess individuals undertaking medical simulation.

What this study adds

► We demonstrate a method to obtain and perform acoustic analysis within surgical simulation.
► We demonstrate a distinction in the modulation of fundamental frequency of voice measured in fellows leading emergency paediatric simulation scenarios. This change was not seen in attending surgeons who were leading similar scenarios.
► Given its simplicity and non-invasive nature, acoustic analysis of a participant's voice offers an opportunity to further assess of individuals' psychological state during medical simulation.

simulation, is that it enhances learners' engagement by increasing their arousal. As stated in research derived from the Yerkes-Dodson law, the ability of an individual to learn is generally affected by the amount of situational demands they receive from their environment.[1 2]

As the demands on the learner increases, so does the learner's arousal level. The increase in arousal, in turn, brings about an increase in learning capability until further increases in stimulation results in an optimisation of learning performance.

BMJ

Additional intensification of the demands after that ideal level becomes deleterious and results in a stress response and a deterioration of learning ability. The arousal response to external situational demands differs among individuals. It is based on such factors as those individuals' knowledge, experience and general comfort level in a particular environment. The reaction to increasing demands with improvement and then deterioration of performance is known as the challenge–threat response outlined as follows:

> Assuming task engagement, evaluations of personal resources, and situational demands determine the degree to which individuals experience psychological states of challenge vs threat. Whereas the challenge occurs when individuals evaluate high personal resources relative to situational demands, threat occurs when individuals evaluate low resources relative to demands.[3]

The negative impact of the threat response, or acute stress, has been well documented in the surgical literature highlighting the impact on individual task performance (technical skills), team performance (non-technical skills) and ultimately patient care.[4–6] More specifically concerning patient safety, one study detailed 40% of surgeons have directly witnessed an operative complication directly related to high stress.[7] A key goal of involving surgeons within medical simulation is to provide an experience to prepare for these complex scenarios through contextual stress. This allows practice of 'performance under pressure' without compromising patient safety and may improve memory consolidation.[8]

Since a surgical simulation scenario is a controlled environment, a simulation could be varied to optimise the level of participants' arousal. Variations in the simulation environment could take the form of varying the cognitive load by changing the complexity of the clinical problem, the 'patient's' comorbidities or the acuity (time pressure) of any clinical changes.[9] Other external factors that could be changed include surgeon's role (primary vs assistant surgeon), the familiarity of the clinical environment, the expertise/training level of the other participants, visitors' presence or even the noise level in the room.[5] The purpose of making these adjustments would be to enable the learners to enter an arousal level that is consistent with a challenge state rather than a threat state. Placing them in a challenge state will then hopefully optimise the participants' learning experience. Using objective and subjective measurements attempts have been made to determine the arousal state in either simulated or actual clinical situations. Unfortunately, no one has been able to easily ascertain the arousal state of the involved healthcare providers.[10]

One possible and as yet underused method to determine arousal level in simulation participants is to study their speech. How a person speaks changes with their emotions. Slavich *et al* have shown elements of language such as word use, syntax and cadence all differ in different emotional states. Changes in emotional states also alter the non-verbal components of communication, such as the volume of speech, pitch, jitter, energy, speaking rate, length and number of pauses.[11] Another element of speech production that changes with alterations in emotional state is voice's fundamental frequency. Of all the previously mentioned non-verbal alterations in voice, fundamental frequency is the one that is most readily associated with an emotional state by other people.[12] Is fundamental frequency a means to study the emotional state of simulation participants? To answer this question, this investigation will measure the fundamental frequency of participants during in situ simulations when participants are experiencing low and high situational demands during a simulated clinical event.

Acoustic analysis and stress

During ordinary conversation, people often intuitively detect the changes in vocal frequency associated with specific emotions such as stress. Objective analysis techniques have, for many years, been used to assess vocal parameters of speech to determine the emotional state of the speaker. The most well-known of these is vocal analysis as part of lie-detector testing to identify deception.[13] Similarly, attempts to assess an individual's underlying stress of an individual within experimental or real-world scenarios have long been subject to acoustic analysis.[14 15]

The most commonly assessed parameters of voice include the previously mentioned fundamental frequency and the first four formants.[16] Fundamental frequency (F_0) corresponds to the frequency of vibration and the opening-closing of vocal folds per second and is measured in Hertz (Hz). The first four formants (F_1, F_2, F_3, F_4) correspond to the natural resonance of the vocal tract. These baseline values vary with age, smoking and a variety of other factors.[16 17]

Protopapas and Liberman demonstrated perceptual variations in F_0 on changes in emotional stress in helicopter pilots in difficult situations.[12] Sondhi *et al* further developed this theme in 2015 assessing mean fundamental frequency (F_0) and formant frequencies (F_1, F_2, F_3, F_4) within stressed and non-stressed environments.[18] Both groups demonstrated that with stress F_0 increases, and in the latter, F_1 and F_2 were shown to decrease.

We hypothesise that the surgeons' response to a stressful event within a simulated paediatric emergency will lead to a concomitant increase in the fundamental frequency and a decrease in the first two formants.

Methodology

Paediatric otolaryngology (ORL) simulation emergency scenarios were created and used to focus on interdisciplinary teamwork and the refinement of clinical skills. They occur within an actual operating room environment ('in situ') with a multidisciplinary team consisting of ORL surgeons, OR nurses and anaesthesiologist.[19] These team training simulations take place approximately four to five times per year and have been occurring from 2008 until the present.

Before each simulation session, all participants signed consent for video/audio recording of the scenarios and prospective data analysis of the anonymised recordings. Video/audio files from each scenario were recorded and stored on the CAE Learning-Space platform (Sarasota, Florida, USA).

A total of 20 ORL simulations from the years 2011–2008 were analysed. An ORL attending led 10 of these simulation sessions, and 10 were led by an ORL fellow. Acoustic analysis was performed on voice samples obtained from the lead surgeon during two distinct events within the recorded scenarios. Voice samples obtained were a minimum of 10 s of continuous, uninterrupted speech by the lead surgeon in both baseline and test conditions.

For the purpose of this study, we assumed that the start of the scenario is the least stressful. It is also when the standard, WHO Surgical Timeout (STO), is performed and the time when the lead surgeon explains the operative plan and organises the team. Because it is a standard procedure, is performed at the beginning of each of the scenarios and involves speech by the lead surgeon, we chose this voice sample as the baseline 'non-stress' sample for acoustic analysis.

Later in the scenario the 'patient' develops an episode of acute clinical deterioration shown on the clinical monitoring. The lead surgeon reacts verbally to this deterioration in vital signs. This
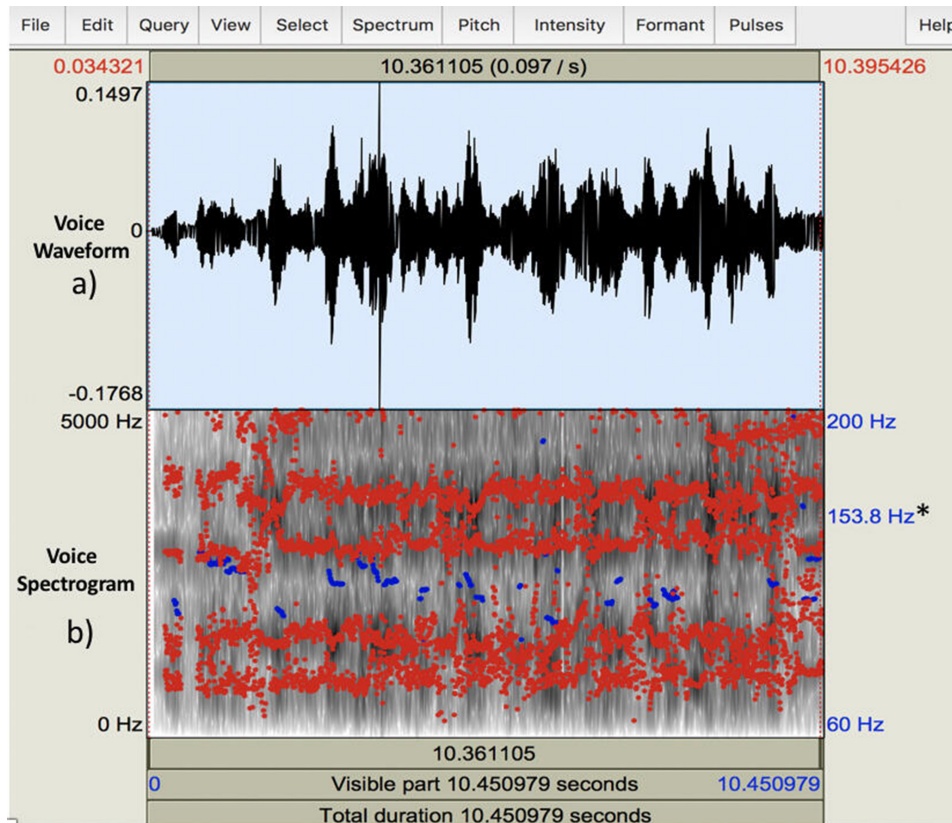
Figure 1 PRAAT software interface demonstrating (A) voice waveform displaying changes in amplitude and (B) 'heat map' spectrogram visually representing the overall spectrum of frequencies. Fundamental frequency (F0) shown (*) calculated over 10.45 s (total duration).

second voice sample is designated as the 'stress phase' sample and used for acoustic analysis. Acute time pressure has been demonstrated to be the greatest source of subjective stress in surgical task performance in previous work.[20 21]

Because the lead surgeon was the only participant over the 20 scenarios who consistently spoke during the WHO checklist and immediately after the 'patient's' condition deteriorated, we used only their voices for the analysis.

The two voice sample audio files from each individual were initially 'cleaned' using Audiocity software (V.2.03) to minimise any background interference to participant's voice.[22] The data were imported into PRAAT (V.5.356, Amsterdam, Netherlands), a software package with a specific emphasis on speech analysis and the underlying physical properties of speech (phonetics).

Long-term pitch analysis PRAAT script was used with an autocorrelation method to obtain a waveform and spectrogram. From this, the mean fundamental frequency (F0) over the recorded 10 s vocal sample was obtained for the two samples from each participant (figure 1). The first four formant frequencies (F1, F2, F3 and F4) for each subject were also calculated for each sample.

The fundamental frequency and first four formants of voice were extracted from both the voice sample during the WHO checklist and during the stress phase portion of each individual's scenario. Means of these differences were calculated for both the fellow and attending group and compared. A t-test determined the statistical significance of the difference of the means. Data distribution is displayed through the inclusion of box–whisker plots.

RESULTS
Baseline vocal characteristics of participants (gender)

Due to differences in laryngeal development and anatomy, the absolute fundamental frequency values between males and females classically varies.[23] As expected, there was a disparity
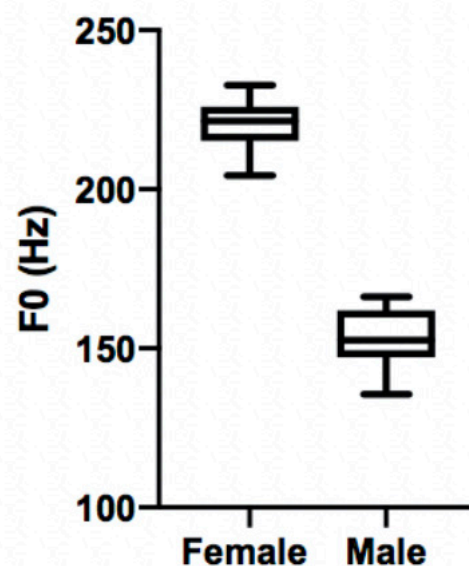
Figure 2 Box-and-whisker diagram showing acoustic analysis of non-stress voice samples showing a baseline difference in fundamental frequency between the female and male voice.

Change in fundamental frequency (F0)
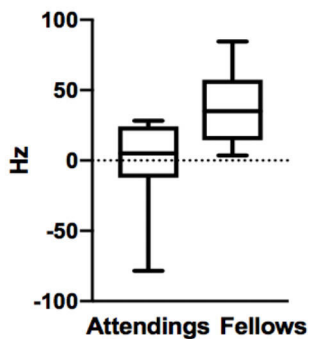
Figure 3 Box-and-whisker diagram demonstrating the difference in change of fundamental frequency of voice between non-stressed and stressed scenario by attendings and fellows (p=0.0106).

between the 10 male and 10 female voices included within this study (figure 2).

During the initial WHO checklist voice sample, the mean F0 for male voice was 153.1 Hz while the mean F0 for female voice was 220.4 Hz. Mean differences in F0 between the male voices and female voices was 67.3 Hz when carrying out the WHO checklist (p≤0.0001). In a recent Australian study, females displayed a baseline F0 that is 76 Hz higher than males.[24] Our higher baseline F0 in the female group is therefore directly comparable with known demographic data.[25] This provided support for the overall validity of the methodology and waveform analysis undertaken within our study.
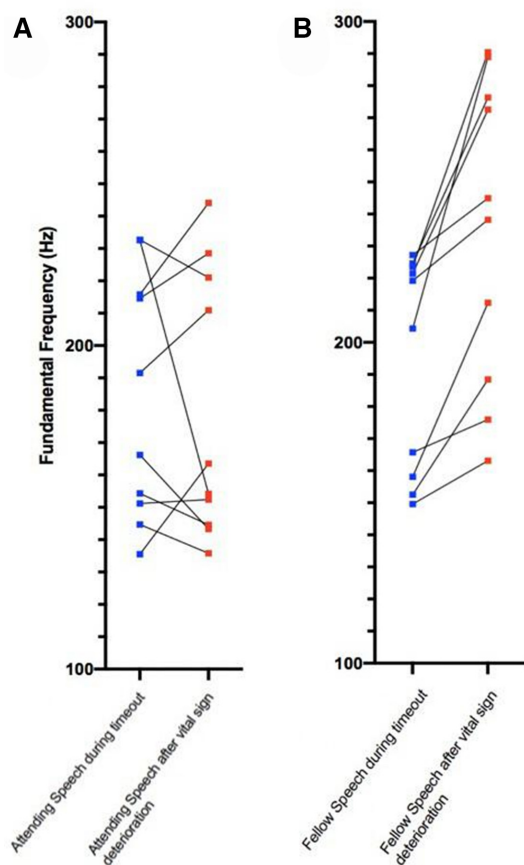
Change in fundamental frequency (F0)

In assessing the change in fundamental frequency (F0) between stressed and unstressed speech, the results were distinct between the fellow and attending groups (figures 3 and 4). The fellows showed a mean increase in their vocal frequency between the baseline voice sample and the stress sample of 34.5 Hz. This is in contrast to the attending group, which demonstrated a mean decrease of their F0 of −0.5 Hz. The difference in the means between the fellow and attending groups was significant (−0.5 vs 34.5 Hz; p=0.01). This result is consistent with an increase in voice modulation during the stress portion by the less experienced fellows group. This change in voice modulation was not identified within the attending group.

Change in first four formants

The first formant values measured (F1) did not demonstrate a statistically significant reduction when comparing either attendings and fellows or stress/non-stress scenarios (p=0.72). Similarly, for the other three formants, statistically significant differences in stress and non-stress scenarios between fellows and attendings did not emerge within our work.

DISCUSSION

As previously mentioned, the Yerkes-Dodson law states, the response to an increasing level of stimulation results in a parabolic curve with initial performance improvement in learning followed by decreased performance as the amount of stimulation continues to escalate (figure 5).[1 2] The physiological changes that occur from the increased arousal on the left side of the curve's peak are considered a 'challenge' response and positively effect performance and learning. Those physiological changes that occur as the arousal increases to the right of the

Figure 4 Graphical depiction of change of fundamental frequency of voice from the non-stressed and stressed scenario shown by attending group (A) and fellows group (B).

peak are categorised as a 'threat' response and negatively effect learning.[26 27]

The physiological changes associated with the threat response refers to what is normally thought of as 'stress'. The adverse effects of the threat response (or stress) are closely related to the challenge response's positive effects. The same situational demands that lead to what is commonly known as stress, when applied to a lesser degree or to an individual with greater
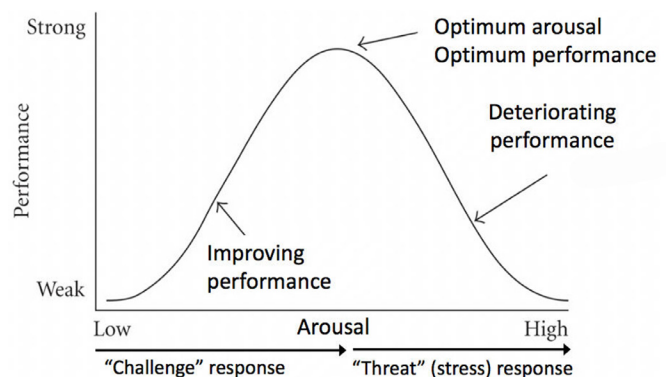
Figure 5 Yerkes-Dodson curve. Initially, as the amount of arousal increases, the individual will show the physiological signs of a challenge response with improved performance. As the arousal continues to increase, the physiological response changes to a threat response (better known as stress) and performance declines. Adapted from Teigen and Dodson.[2]

resources, can lead to a positive outcome, namely the challenge response. If the challenge–threat theory is correct, then it can be difficult to determine the role of stress in education and in clinical practice. This is because a given set of situational demands may have a deleterious effect on one person's performance, by causing them to react with a threat response, while another person for whatever reason (more experience, more knowledge of the problem, higher comfort level in the environment, etc) may be subjected to the same situational demands and may develop a challenge response with a resulting enhanced result. Knowing the challenge–threat state of an individual could have a significant impact on further educational or even clinical interventions. If a simulation participant is in a challenge state during most of a simulation, then the simulation is likely optimised for that participant and possibly for others at their level of training. Conversely, determining that a clinician is in a threat state during a clinical experience may enable an intervention which will allow him or her to react more confidently and appropriately during a subsequent clinical encounter.

Medical simulation scenarios are controlled clinical events. As such, the degree of arousal can, in theory, be adjusted to optimise the learning experience of the participants. This adjustment may take the form of alterations in scenario design or changes in how acute a clinical situation would deteriorate within a scenario. However, to ascertain the effect of these changes, how does one determine a participant's level of arousal? How does one know if a simulated situation has placed the participant(s) in a better (challenge) or worse (threat) mental state for learning?

By measuring an individual's response to the situational task before them, one can, in theory, determine if they are in a challenge or threat state. Responses can be measured either subjectively or objectively. Subjective responses may be obtained by several different tools. Examples are the State-Trait Anxiety Index (STAI)[28] and the NASA Task Load Index (TLX).[29] Both are questionnaires. The STAI is used to determine the emotional state of a participant after an experience or activity. The NASA TLX evaluates the degree of burden that a task presents to an individual. The inference of this being that an individual's perceived increase in task load may represent an increase in threat response or stress. These questionnaires, by necessity, are administered soon after the activity in question.

The physiological stress-response involves the interplay between the autonomic nervous system (ANS) and the hypothalamic-pituitary-adrenal (HPA) axis. Numerous objective physiological measurements may be used to measure the activity of these two systems. For the ANS, these include the galvanic skin response, cardiac output (CO), heart rate (HR), ventricular contractility (VC), total peripheral resistance (TPR) and heart rate variability (HRV). The HPA axis activity is commonly measured by salivary cortisol and amylase levels.[30]

While all of these represent measurements used to assess the ANS and HPA axis activity, their interplay, and therefore their interpretation, during stress is complex. In the challenge state, the ANS changes result in the CO, VC and HR are all increased while the TPR is decreased. The ANS also responds in the threat state with an increase in VC and HR. However, in this instance the TPR is increased and the CO remains relatively unchanged. These ANS responses persist for a more extended time after the stress-producing episode has ended in the threat state than in the challenge state. In observing the cortisol levels, which measure HPA axis activity, the cortisol reaches a higher level for a significantly prolonged period in the threat state compared with the challenge state.[31] These states are also associated with a difference in their subjective emotions. The individuals who are

undergoing a challenge state often report feelings of confidence and pride. This is in contrast to those in a threat state who experience anxiety and shame.

None of the current subjective or objective measurements is ideal. The subjective techniques all depend on participants providing candid, accurate responses. This may not always be the case. The responses are recorded shortly after the activity. Also, these techniques evaluate the activity, whether it is simulated or actual as a whole. They cannot easily evaluate a specific occurrence during a string of events within a scenario or clinical occurrence.

Objective measurements such as CO, VC, HR and TPR have the advantage of providing continuous recording and can be correlated with specific events within a scenario. However, they require special equipment that needs to be placed on the participant beforehand. HRV requires a prolonged period of recording and cannot be linked to specific events.[32] Salivary cortisol and amylase have an approximately 15 min delay between an event and a subsequent change in salivary levels. They also are influenced by the time of day, previous stressors and recent food or drink ingestion.[33]

The main drawback of acoustic analysis is that it cannot gauge a participant's state if the individual is not speaking. However, it has many potential advantages. First, since the sound source can be taken from a video, the sound sample can easily be correlated with what is happening during the scenario. Likewise, since voice samples may be extracted from video, the acoustic measurements can be derived from archived scenarios done in the past. This archiving allows a comparison of responses many years apart.

This investigation is the first pilot research assessing acoustic analysis in medical simulation scenarios. We demonstrated a distinction in the modulation of fundamental frequency between attendings and fellows undertaking emergency simulation scenarios. However, within our dataset, we did not find an associated decrease in the first two formants (F1 and F2). This may, however, be secondary to the variability of measurements of the resonance of the vocal tract. Within the voice literature, there is a current focus on F0 for its reliability in comparison with other measures.[25]

Our work was undertaken using acoustic analysis software that could be replicated in other areas of medical simulation. Our initial findings suggest that focusing on vocal fundamental frequency, as determined by acoustic analysis, may provide an objective measure of participants' physiological response to the external stimulation from their environment.

In the future, information from vocal acoustic analysis of recorded simulations may help design simulation scenarios that provide the proper degree of arousal to optimise learning for clinicians with varying degrees of clinical experience.[34] This may become more feasible as recent technologies have enabled acoustic analysis of speech in various environments using smartphones, either as simple recording devices or by using self-contained apps.[35] Similarly, there is the possibility that acoustic analysis of participants' voices at the time of a simulation may allow titration of stressful stimuli during an on-going scenario in real time.

Another application of vocal acoustic analysis would be to determine if fundamental frequency is perceptible by clinical team members and whether it serves as a signalling mechanism to them. Excessive external displays of stress are known to have an impact on perceptions of leadership and regularly, within simulation debriefing meetings, discussions related to measured and calm approaches to the encountered emergency are raised

as positive traits. Despite the expressed desire to remain composed, often participants report during the post-scenario debriefings that *"I didn't feel calm"* or *"I don't know why, but I felt anxious."* Participants frequently expressed these types of comments at various levels of hierarchy within the clinical team. Given the immersive, high-fidelity level of modern simulation, this sentiment was expressed by even the most experienced clinicians, namely the attendings. Yet there was little change in the vocal modulation of the attendings between the non-stress and stressed voice samples in our pilot study. Because of their experience and seniority, did they truly not feel stressed by the situation, or did they modulate their voice to hide their emotional state? There has been a correlation between emotions such as sorrow, anger and fear with vocal acoustic changes. But in this instance, rather than a true reflection of emotional state, is fundamental frequency in some individuals a result of conscious or unconscious vocal modulation to conceal their anxiety and project a calmer demeanour? It could very well be an adaptive means by which someone in a leadership position can, by the non-language aspects of their voice, project a message that "all is well."

One outlier from the attending group demonstrated a significant fall in fundamental frequency during the stressful encounter. This was of a similar order of magnitude (78 Hz) to the rise of F0 shown in the fellow group. Could this reduction in vocal frequency be in the form of a learnt modulation or 'vocal camouflage' to a stressful encounter? The concept of whether attendings' voices were better able to disguise their performance stress than the fellows' group is one that is difficult to answer presently. Voice modulation may indeed be an acquired trait through experience and assessing its sensitivity would require further work. Assessment of voice modulation (be it unconscious or conscious) could potentially be studied to determine if such modulation impacts the team members' opinion of surgical leadership. In this study, we chose to perform an analysis of the voice of solely lead surgeons as they reliably made specific utterances in both of the specified non-stressed and stressed situations within the scenarios. In future studies, as more experience is gained in acoustic analysis in the healthcare setting and pursue some of the questions suggested previously, the voices of other team members could and should also be analysed.

During this study, no additional biophysical monitoring, such as those mentioned previously, was used to measure the physiological response correlating with the change in fundamental frequency. To make the correlation, measurement of some combination of physiological parameters such as respiratory rate, HR, HRV, galvanic skin resistance, as well as catecholamine and glucocorticosteroid responses would be necessary.[4] The current means to measure an individual's response to stress is complicated and unwieldy. While performing acoustic analysis is a relatively easy technique, until the correlations between acoustic analysis and the other subjective and objective measurements are made, it is difficult to speculate how helpful voice analysis will be in determining the psychological state of healthcare providers. However, contrary to other measurements, changes in emotions have been shown to effect changes in vocal properties.[36 37] This ability of the voice to reflect emotional states may in be useful in readily determining the challenge–threat state and enhancing medical education and clinical environments.

CONCLUSIONS

The fundamental frequency of the human voice changes with alterations in emotional state and appears to be associated with the physiological stress response. Medical simulation can produce a stressful environment so participants in simulation frequently develop a stress response during such activities. This study demonstrates a successful method of performing acoustic analysis of the participants voices taking part in medical simulations.

Using recordings obtained during previous simulations, the lead ORL surgeons' voices during simulated paediatric surgical emergencies were analysed. Voice analysis was performed on 10 of the sessions that were led by attending surgeons and 10 similar sessions led by paediatric ORL fellows. Within each scenario, the lead surgeon's voice was analysed and the change in fundamental frequencies obtained during a specific period of low stress was compared with the fundamental frequency of a high-stress event. In comparing these two groups, we observed that the fellows displayed an increase in the fundamental frequency of their voices between the low-stress and high-stress portions of the scenarios. No mean change in fundamental frequency was seen in the attending group. The difference of the means between these two groups was statistically significant.

Given the relative simplicity and non-invasive nature of the measurement, our pilot acoustic analysis suggests participants' voice offers an opportunity for further assessment of participants' stress response during simulation. The ability to more accurately assess stress experienced by those undergoing simulation may allow appropriate calibration of scenarios to optimise their educational value. Further validation of acoustic analysis perhaps combined with other objective biophysical measurements and subjective participant experience offers an opportunity to better prepare for, and benefit from, the stressful scenarios inevitably encountered in medicine.

ORCID iD
Andrew Hall http://orcid.org/0000-0002-4134-661X

REFERENCES

1 Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. *J Comp Neurol Psychol* 1908;18:459–82.
2 Teigen KH, Dodson Y. 'A law for all seasons'. *Theory & Psychology* 1994;4:525–47.
3 Saltsman TL, Seery MD, Ward DE, *et al*. Is satisficing really satisfying? Satisficers exhibit greater threat than maximizers during choice overload. *Psychophysiology* 2021;58:e13705.
4 Jones KI, Amawi F, Bhalla A, *et al*. Assessing surgeon stress when operating using heart rate variability and the state trait anxiety inventory: will surgery be the death of us? *Colorectal Dis* 2015;17:335–41.
5 Wetzel CM, Kneebone RL, Woloshynowych M, *et al*. The effects of stress on surgical performance. *The American Journal of Surgery* 2006;191:5–10.
6 Arora S, Sevdalis N, Nestel D, *et al*. The impact of stress on surgical performance: a systematic review of the literature. *Surgery* 2010;147:318–30.
7 Anton NE, Montero PN, Howley LD, *et al*. What stress coping strategies are surgeons relying upon during surgery? *The American Journal of Surgery* 2015;210:846–51.

8   LeBlanc VR. The effects of acute stress on performance: implications for health professions education. *Academic Medicine* 2009;84:S25–33.

9   Fraser KL, Ayres P, Sweller J. Cognitive load theory for the design of medical simulations. *Simulation in Healthcare* 2015;10:295–307.

10 Georgiou K, Larentzakis A, Papavassiliou AG. Surgeons' and surgical trainees' acute stress in real operations or simulation: a systematic review. *Surgeon* 2017;15:355–65.

11 Slavich GM, Taylor S, Picard RW. Stress measurement using speech: recent advancements, validation issues, and ethical and privacy considerations. *Stress* 2019;22:408–13.

12 Protopapas A, Lieberman P. Fundamental frequency of phonation and perceived emotional stress. *J Acoust Soc Am* 1997;101:2267–77.

13 Streeter LA, Krauss RM, Geller V, *et al*. Pitch changes during attempted deception. *J Pers Soc Psychol* 1977;35:345–50.

14 Williams CE, Stevens KN. On determining the emotional state of pilots during flight: an exploratory study. *Aerospace Medicine* 1969;40:1369–72.

15 Tolkmitt FJ, Scherer KR. Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology* 1986;12:302–13.

16 Pegoraro Krook MI. Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. *Folia Phoniatr Logop* 1988;40:82–90.

17 Traunmüller H, Eriksson A. The frequency range of the voice fundamental in the speech of male and female adults (Department of Linguistics, University of Stockholm), 1994. Available: https://pdfs.semanticscholar.org/aa8b/acb5e7843740 fbea24742c3046fbcc009a49.pdf

18 Sondhi S, Khan M, Vijay R, *et al*. Acoustic analysis of speech under stress. *Int J Bioinform Res Appl* 2015;11:417–32.

19 Volk MS, Ward J, Irias N, *et al*. Using medical simulation to teach crisis resource management and decision-making skills to otolaryngology housestaff. *Otolaryngology–Head and Neck Surgery* 2011;145:35–42.

20 Ng R, Chahine S, Lanting B, *et al*. Unpacking the literature on stress and resiliency: a narrative review focused on learners in the operating room. *J Surg Educ* 2019;76:343–53.

21 Poolton JM, Wilson MR, Malhotra N, *et al*. A comparison of evaluation, time pressure, and multitasking as stressors of psychomotor operative performance. *Surgery* 2011;149:776–82.

22 Mazzoni D. Audacity(R): a free, cross-platform digital audio editor, (version 2.0.3), 2013. Available: http://www: audacity.sourceforge.net/ [Accessed 14 Nov 2020].

23 Titze IR. Physiologic and acoustic differences between male and female voices. *J Acoust Soc Am* 1989;85:1699–707.

24 Leung Y, Oates J, Papp V. Speaking fundamental frequencies of adults speakers of Australian English and effects of sex, age and geographic locations. *J Voice* 2020;27:S0892-1997:30202–2.

25 Leong K, Hawkshaw MJ, Dentchev D, *et al*. Reliability of objective voice measures of normal speaking voices. *Journal of Voice* 2013;27:170–6.

26 Diamond DM, Campbell AM, Park CR, *et al*. The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plast* 2007;2007:1–33.

27 Tomaka J, Blascovich J, Kelsey RM. Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology* 1993;18:616–24.

28 Harvey A, Nathens AB, Bandiera G, *et al*. Threat and challenge: cognitive appraisal and stress responses in simulated trauma resuscitations. *Med Educ* 2010;44:587–94.

29 Zheng B, Jiang X, Tien G, *et al*. Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surg Endosc* 2012;26:2746–50.

30 Marques AH, Silverman MN, Sternberg EM. Evaluation of stress systems by applying noninvasive methodologies: measurements of neuroimmune biomarkers in the sweat. *Heart Rate Variability and Salivary Cortisol Neuroimmunomodulation* 2010;17:205–8.

31 Mendes WB, Major B, McCoy S, *et al*. How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *J Pers Soc Psychol* 2008;94:278–91.

32 Valentin B, Grottke O, Skorning M, *et al*. Cortisol and alpha-amylase as stress response indicators during pre-hospital emergency medicine training with repetitive high-fidelity simulation and scenarios with standardized patients. *Scand J Trauma Resusc Emerg Med* 2015;23:31.

33 Clarke S, Horeczko T, Cotton D, *et al*. Heart rate, anxiety and performance of residents during a simulated critical clinical encounter: a pilot study. *BMC Med Educ* 2014;14:153.

34 Fraser KL, Ayres P, Sweller J. Cognitive load theory for the design of medical simulations. *Sim Healthcare* 2015;10:295–307.

35 Kojima T, Fujimura S, Hori R, *et al*. An Innovative Voice Analyzer "VA" Smart Phone Program for Quantitative Analysis of Voice Quality. *Journal of Voice* 2019;33:642–8.

36 Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *J Acoust Soc Am* 1972;52:1238–50.

37 Cowen AS, Elfenbein HA, Laukka P, *et al*. Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist* 2019;74:698–712.