University of Eastern Finland
Computer Science Department

Evaluating the disturbance of environmental noise on medical students' simulated surgical performance

Fatima Rabia Yapicioglu (306627)
Rhythm Bhatia (308847)
Aleksi Kuikka (258594)

UNIVERSITY OF
EASTERN FINLAND

Contents

1. Abstract

[*This study evaluates the effect of noise on student-surgeons' performance in three microsurgery related tasks. The experiments were conducted in Kuopio University Hospital where 21 students volunteered for the study. The participants performed the three tasks in both a silent environment and a noisy environment where the participants were subject to a 85db noise played via headphones. Their performance was visually recorded and their biometrics as well as eye tracks were collected by specialized machinery [5], [6]. After finishing their tasks, the students self-evaluated how demanding the procedure felt for them and the video footage of their performance was given for two blinded surgeons to evaluate for the quality, efficiency and handling of it. We performed a two-sample t-test on the quality of knot, efficiency and handling and the reported distractions and mental, physical and temporal demands,. We found significant difference (p=0.05) in the performance evaluations in three of those areas, namely: Efficiency, physical demands and distractions. The assessed performance efficiency was reduced by 15% while the reported physical demands increased by 88% and distractions by 156%*]

2. Problem definition and background

The effect of environmental noise on the surgeons' biometric records has been explored in previous study by Cabrera IN, Lee MH [3]. It has been noted that during surgical procedures, the volume of the noise can be as high as 100-120 dB. [4] Such noise disturbance could prove dangerous for the medical staff and the patients [2] and impact the performance of the working surgeons [1]. Concerning that, the surgeons' perceived demands, distractions and performance were evaluated both in noisy and silent environment in three tasks, namely: Mesh alignment, knotting and go-around.

.

2.1 Aim of the Research

This study aims to evaluate if there is a significant effect caused by noise on the participating students' surgical performance.

2.2 Objectives of the Research

1. To compare and analyze the performance and the mental, physical, temporal demands experienced by the surgeon (student) in silent and noisy working environments.
2. To infer by statistical calculations if there is a significant difference between the measured variables in silent and noisy working environments.
3. To visualize the relevant data of the surgeons grouped by the noise conditions.

2.3 Main Question of the Research

What is the effect of noise on surgeon's task performance and perceived demands?

2.4 Sub-questions of the Research

1. What should be the main factors or attributes in the data that we should take into consideration while we are evaluating the effects of noise on the surgeons?

2. Which methods should we use to evaluate the effect of noise, for example, what type of statistical attributes to infer whether there is a significant difference between the two sample groups or not?

3. How can we effectively visualize and represent the difference between the surgeons' data in noisy and silent environments?

3. Methodology

This research mostly focused on quantitative data. The self-reported demands experienced by surgeons were quantified according to a scale. In addition to self-reportage this research utilized scores reported by blinded experts according to participant observation.There were no biasing or assumptions made for data collection. [9] The tasks that were labeled by volunteers are as follows:

1. Knotting: Subject had to tie 3 to 4 knots in each trial.
2. Mesh alignment: Subject had to carry 4 meshes one-by-one and place it on top of each other in a designated area,
3. Go-around: Subject had to pass the thread through a set of needles (4 to 6).

The performance in the tasks were codified and quantified by three scores: Knot quality, efficiency and handling.
There was some human error while performing the experiments which we took into consideration while evaluating the data [10]. It took approximately 2 to 3 days to collect and label the data for each participant.

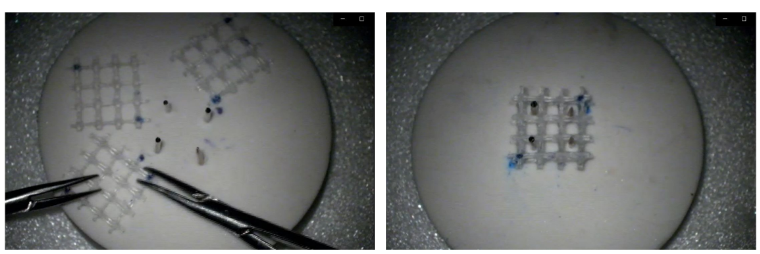

Figure 1.0. Mesh Alignment Task

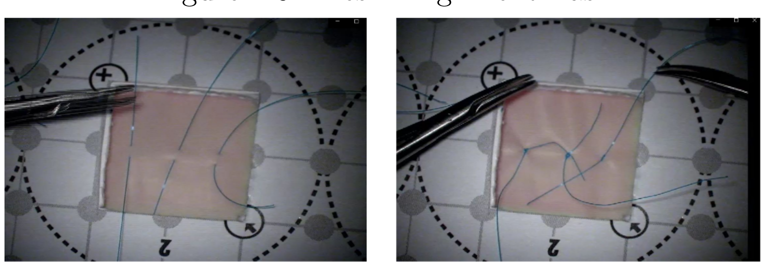

Figure 1.1. Knotting Task

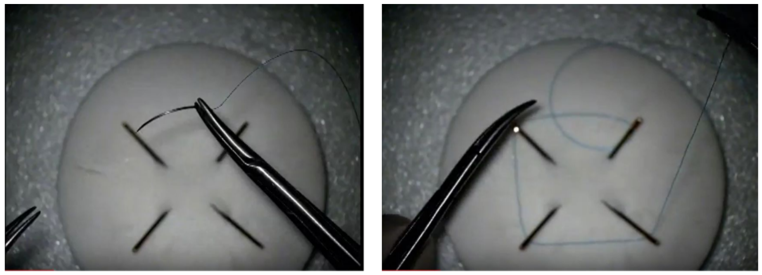

Figure 1.2. Go-Around Task

3.1 Experiment Setup

The experiment took place in a quiet, evenly-lit lab in the university hospital. A standardized microsurgical workstation and layout was setup identically for each participant. Each participant was provided with a table-mountable Zeiss Pico surgical microscope, high-quality straight microsurgical forceps, needle holder and mesh-grids that were used in the tests. Before the experiment started, each participant was asked to close their eyes, rest their hands on a table and relax for one minute.

3.2 Data Collection and Assessment

Each surgeon was filmed individually performing the tasks. The filming process started after receiving the consent of the surgeon. The field of view was the workstation and surgeons' hands.

Video clips with muted sound were assessed. Each video showed a surgeons' hands in the noisy or silent environment performing the task. The assessors were blinded to the surgeons' identity and to whether they were performing the task in noisy or silent environment. Main References for the data collection were [9] and [11].

3.3 Post-Task Assessment

We used the following scheme for the post-task evaluations:

Please evaluate the procedure by marking 'X' on each of the six scales at the point which best fits your experience.

Mental Demands

How mentally fatiguing was the procedure?

___|___|___|___|___|___|___|___|___|___|___

Very Low Very High

Physical Demands

How physically fatiguing was the procedure?

___|___|___|___|___|___|___|___|___|___|___

Very Low Very High

Temporal Demands

How hurried or rushed was the pace of the procedure?

___|___|___|___|___|___|___|___|___|___|___

Very Low Very High

Task Complexity

How complex was the procedure?

___|___|___|___|___|___|___|___|___|___|___

Very Low Very High

Situational Stress

How anxious did you feel while performing the procedure?

___|___|___|___|___|___|___|___|___|___|___

Very Low Very High

Distractions

How distracting was the operating environment?

___|___|___|___|___|___|___|___|___|___|___

Very Low Very High

Figure 1.3. Post-Task Assessment Method

In the above Figure 1.3. , each box is 2 points, the scale being between 0-20.
Each participants's performance was assessed by two blinded surgeons and the participants evaluated their selves by filling the post-task assessment. Finally, when this part of the data collection had completed there was a summarized *.csv file which shows the quantified demands reported by the participants.
The final experiment outcomes are as in the following Table 1.0.

Table 1.0. Sample Demands Table of the Participants

Date	Gender	ID	Condition	Mental Demands	Physical Demands	Temporal Demands	Task Complexity	Situational Stress	Distractions
###	F	P1	Noise	7	9	...			
###	M	P2	Silent	8	9	...			

The expert evaluations followed a rating scheme standard which is supposed to evaluate the quality and efficiency of the mesh alignment, knotting, and go-around tasks. For example, there were some indicators to evaluate quality of the knot at the lowest level: Not square, loose and cut ends too long/short, at the middle level: partially square, somewhat loose, cut ends OK length, and at the top level: square, snug and, cut ends proper length. Every type of task was evaluated on the following rating scale for each participant:

*1*__________*2*__________*3*__________*4*__________*5*

Not square	Partially square	Square
Loose	Somewhat loose	Snug
Cut ends too long/short	Cut ends OK length	Cuts proper length

Figure 1.4. Quality of the Knot Evaluation

We made certain preliminary visualizations on the experiment outcomes and biometrics data which helped us direct our focus on the most relevant parts concerning our objectives. We chose the performance scores to have outside, expert perspective in our assessment and the self-reported demands as supplementary context for the experience of the surgeons. We deemed the biometrics data unsuitable for our objectives as it was hard to connect to the surgeons' performance and generally hard to interpret. The physiological signals that were measured were heart rate, body temperature, muscle activity of the hand movements. However, these signals can be affected by the other environmental factors such as environmental heat, excitement etc. that are hard to control for our study. Thus we limited our research to simpler self-reported and outwardly assessed variables.

As the main statistical test, we chose Student's t-test for paired samples on the reported demands and the assessed performance, which shows whether the effect of the environmental noise is systematic or mere statistical noise. T-test is suitable for smaller sample sizes (e.g. $n < 30$), and where they can be assumed to follow a normal distribution and have an equal variance. [12] To illustrate the actual difference in magnitude then, we found the relative difference between the means of the variables prior and during noise disturbance.

3.4 Tools

This research study was working on the Jupyter Notebook and R Notebook with R and python languages. The visualization libraries were as follows: Matplotlib, seaborn, plotly, and ggplot2. As for the statistical calculations: Numpy and Pandas were used.
The links to the tools can be found as follows,

1. https://matplotlib.org/gallery/index.html
2. https://seaborn.pydata.org/
3. https://plot.ly/python/
4. https://ggplot2.tidyverse.org/
5. https://numpy.org/
6. https://pandas.pydata.org/

3.5 Sampling

In this research purposive, convenient nonprobability sampling was implemented because we were interested particularly in surgeons' performance under noise. It is not in the interest of this research to infer about the effect in general population which due to lack of specialization might be lesser, but rather about the relevant practitioners in the field it. The sampling frame for the research are the residents at Kuopio University Hospital where 21 people had volunteered to be part of the research. [8]

3.6 Research Ethics

At the beginning of the experiment, participants were introduced to the study, sensors, and use of the collected data. After the introduction, participants written informed consent was obtained prior to their inclusion in the study, the consent form detailed the anonymity, use and disposal of the participants' data collected during the study. The video footage shows only the hands of the surgeons during the task, the video footage was muted for the assessment.

4. Results

To statistically evaluate the significance of the difference between the silent and noisy environment we performed one-tailed two sample t-test on our data, assuming normal distribution, equal variance and independence of the samples [12], [13].

Table 2.1. Statistical Two-Sample T-test Results Table

	Mental Demands	Physical Demands	Temporal Demands	Distractions	Quality of knot	Efficiency	Handling
T-score	1.08	3.42	1.11	3.16	-0.118	-1.87	-0.378
P-value	0.148	0.00145	0.141	0.00258	0.454	0.0386	0.355

Each of the variables' t-scores and p-values are presented in table 2.1. The higher the t-score, the higher the probability (p-value) implying that noise has positively affected the variables' measure and vice versa.
Thus, if we fix the significance level to 5% we notice that out of all the presented variables, physical demands, distractions and efficiency cross the threshold. Therefore on the given p-values, we can reject the null hypothesis stating that the means of the silent and noisy environment samples are the same, and accept the alternative hypothesis that they are systematically higher for physical demands and distractions and lower for efficiency.

Table 2.2. Statistical Two-Sample T-test Relative Differences Table

	Mental Demands	Physical Demands	Temporal Demands	Distractions	Quality of Knot	Efficiency	Handling
Relative difference	1.24	1.88	1.36	2.56	0.989	0.850	0.964

T-test scores do not intuitively show the difference in magnitude between the samples, they merely state that the signal is sufficiently larger than the noise. Table 2.2 illuminates the relative difference of the means of the two sample groups. In the case efficiency, the reduction is nearly 15%, whilst the self-reported physical demands are up nearly 90% and the distractions have more than two-folded

The demands are visualized as in the following figure 2.0. grouped by the environmental noise condition. There is an illustrated difference between the silent and noisy environment according to the spreading area of the box plots. The silent environment box plots spread out in the lower end and the noisy box plots spread out in the higher end of the scales.

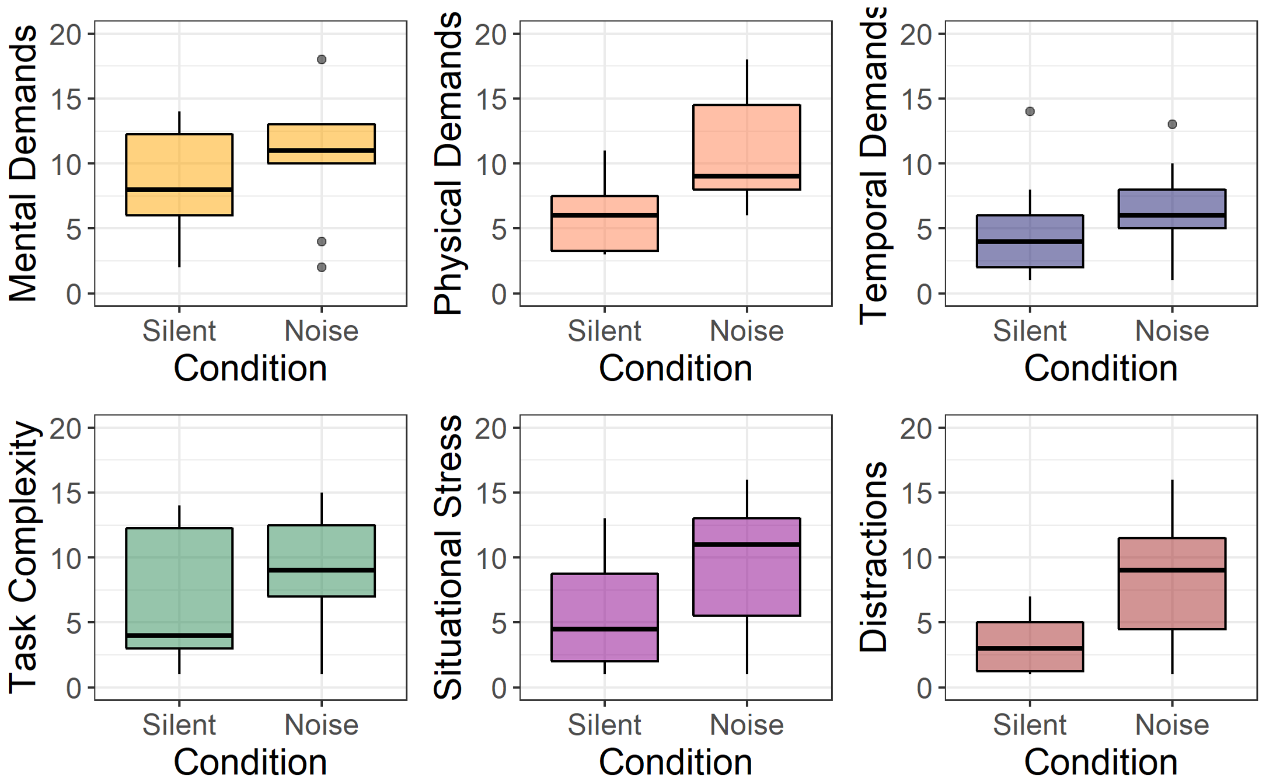

Figure 2.0. The demands of the participants grouped by condition

In conclusion, we found a statistically significant difference (p=0.05) in the self-reported physical demands and distractions over the outwardly assessed task performance efficiency. The relative difference of the statistically significant variables were 88% higher for physical demands and 156% higher for distractions, whilst the expert-assessed efficiency declined for 15%

5. Limitations and Future Work

Our research study focused mostly on observed data in the form of self-reportage and blinded experts, but there was a lot of data that was left unused for our purposes.

Some of the unused logs and biometric data were collected by specialized instruments that could be used to eliminate subjective bias from our analysis. Thus, make more objective claims for the effects of noise disturbance [7]. Although interpreting this data effectively would require the researcher to have access to professional with deep domain expertise in the relevant fields of medicine.

Additionally, to extend this study to analyze the effect of noise at different scales, future research work could test varying volumes and styles of noise with a larger sample size. It should also be noted that since our participants were students, we can not infer how seasoned professionals are affected by the noise.

6. References

1. Hodge B, Thompson JF.,1990, Noise pollution in the operating theatre. Lancet 1990;335:891-4
2. Shapiro RA, Berland T. Noise in the operating room. 1972, N Engl J Med 1972;287:1236-8.
3. Cabrera IN, Lee MH., 2020,Reducing noise pollution in the hospital setting by establishing a department of sound: a survey of recent research on the effects of noise and music in health care. Prev Med 2000;30:339-45.
4. West J, Busch-Vishniac I, King J, Levit N. Noise reduction in an operating room: a case study. J Acoust Soc Am 2008;123:3677.
5.Justus J. Randolph, Chapter 4 in Randolph (2008) (Data collection in educational technology research, p. 76-77)
6. Kenneth Holmqvist, Marcus Nystrom, Fiona Mulvey, 2012, Eye tracker data quality: What it is and how to measure it.
7. Ella Haig, Stephan Weibelzahl, 2006, Can Log Files Analysis Estimate Learners‘ Level of Motivation?.
8. R.S. Brown (2010, p. 142-146, v.3), International Encyclopedia of Education https://www-sciencedirect-com.ezproxy.uef.fi:2443/science/article/pii/B9780080448947002943.
9. Siu KC1, Suh IH, Mukherjee M, Oleynikov D, Stergiou N, The impact of environmental noise on robot-assisted laparoscopic surgical performance.
10. Persoon MC, Broos HJ, Witjes JA, Hendrikx AJ, Scherpbier AJ, 2010,The effect of distractions in the operating room during endourological procedures. Surg Endosc. 2011;25(2):437–443. doi:10.1007/s00464-010-1186-8 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3032204/)
11. Jotautienė, E., Bernotienė, I., Liebuvienė, J., Bernotas, A. (2019). Quantitative study of environmental noise on dovilu village population. In Engineering for rural development [elektroninis išteklius]: 18th international scientific conference: proceedings, May 22-24, 2019. Jelgava, 2019, Vol. 18.
12. Kim, T. K. (2015). T test as a parametric statistic. Korean journal of anesthesiology, 68(6), 540.
13. Fay, M. P., Proschan, M. A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Statistics surveys, 4, 1.
14. Temple, C. L., Ross, D. C. (2011). A new, validated instrument to evaluate competency in microsurgery: the University of Western Ontario Microsurgical Skills Acquisition/Assessment instrument [outcomes article]. Plastic and reconstructive surgery, 127(1), 215-222.