7-28-2022

Evaluation of preview cues to enhance recall of auditory sequential information

Tsz Chun Marco LAU

Terms of Use

EVALUATION OF PREVIEW CUES TO ENHANCE RECALL OF AUDITORY SEQUENTIAL INFORMATION

LAU TSZ CHUN MARCO

MPHIL

LINGNAN UNIVERSITY

2022

EVALUATION OF PREVIEW CUES TO ENHANCE RECALL OF AUDITORY SEQUENTIAL INFORMATION

by
LAU Tsz Chun Marco
劉梓俊

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
Master of Philosophy in Psychology

Lingnan University

2022

ABSTRACT

Evaluation of Preview Cues To
Enhance Recall of
Auditory Sequential Information

by

LAU Tsz Chun Marco

Master of Philosophy

Background: In previous work, an auditory vital sign display of five patients was developed. Sounds denoting the vital signs of each patient were delivered in order, with a special sound for any patient whose vital signs were all normal. Although the display was effective, accuracy decreased as the number of abnormal patients increased. We wondered whether accuracy would improve with a preview sound indicating the number of patients with abnormal vital signs in the upcoming sequence by reducing working memory load. We also wondered whether the preview sound would affect the performance of responding to concurrent task.

Methods: A 3 (preview cue type) x 4 (number of abnormal patients) mixed-factorial design was adopted. Preview cue type (between-subjects) was either time-compressed speech or an abstract sound containing white noise pulses to indicate the upcoming number of abnormal patients, or no preview cue. The number of abnormal patients (within-subjects) was zero, one, two, or three.

Results: Preview cue did not improve non-clinician participants' ability to identify the location in the sequence or the vital signs of patients with abnormal vital signs. Response accuracy dropped as the number of patients with abnormal vital signs increased. The preview cue types did not affect the accuracy of responding to the concurrent task, However, the users tended to ignore the concurrent task when preview cue created by abstract sound with white noise pulses was used .

Conclusion: The current preview cue did not improve or hurt the performance of identifying abnormal patients' locations and vital signs. However, it would degrade the concurrent task performance. Therefore, the current design of preview cue can be eliminated in future auditory display design.

DECLARATION

I declare that this is an original work based primarily on my own research, and I warrant that all citations of previous research, published or unpublished, have been duly acknowledged.

SIGNED

(LAU Tsz Chun Marco)
Date: 26/08/2022

CERTIFICATE OF APPROVAL OF THESIS

EVALUATION OF PREVIEW CUES TO
ENHANCE RECALL OF
AUDITORY SEQUENTIAL INFORMATION
by
LAU Tsz Chun Marco

Master of Philosophy

Panel of Examiners :

SIGNED

(Prof. CHEN Hon Fai) (Chairman)

SIGNED

(Prof. OR K.L. Calvin) (External Member)

SIGNED

(Prof. LEE Lap Fai Alan) (Internal Member)

SIGNED

(Prof. HUANG Yi) (Internal Member)

Supervisor :

Prof. LEE Lap Fai Alan

Approved for the Senate :

SIGNED

(Prof. MOK Ka Ho Joshua)
Chairman, Postgraduate Studies Committee

28 JUL 2022

Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The work described in this thesis was fully supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project No.: LU 13601919). The study presented in this thesis is published in *Auditory Perception and Cognition*.

I would like to express my sincere gratitude to my supervisors, Dr. Simon Li, and Dr. Alan Lee. I have learnt a lot about research design, implementation, data analysis, and project management from them. They always support me fully not only in conducting and presenting the present study, but also in my career preparation.

I would also like to thank Professor Penelope Sanderson and Dr. Robert Butch Loeb. They offered a lot of useful advice in designing the study and preparing the manuscript submitted to *Auditory Perception and Cognition*. Also, I want to thank Chiu Wing Sum Jenny for assistance in collecting the data for this study, and Harald Waxenegger for software programming in this study.

Other than my great colleagues, I would also like to thank the professors and administrative staffs in Department of Applied Psychology, my schoolmates, and the cleansing workers for WYL303 in Lingnan. The professors are very kind in offering me feedbacks on my research, which allows me to know others' perspectives on the findings. The administrative staffs always help with the necessary administrative works during the MPhil study period. Also, I am happy to have some lovely schoolmates in WYL303 to support each other in any aspects of life. With the help of cleansing workers, I could have a clean office and warm emotional support in my daily work.

1. INTRODUCTION

Clinicians usually monitor up to five patients in medical-surgical ward (McHugh et al., 2021). They rely on medical alarms to alert them to otherwise unnoticed patient deteriorations while they work on other tasks (Hendrich, 2008). The auditory modality can help clinicians monitor patients in an eyes-free manner while performing visually-intensive tasks. In previous work, an auditory display of the status of multiple patients was developed, using time-compressed speech (Li et al., 2019a; Sanderson et al., 2019). Although results were encouraging, working memory limitations (Cowan, 2010) seemed to compromise the display's effectiveness as the number of patients with abnormal vital signs increased (Hickling et al., 2017; Li et al., 2019a). Accordingly, in the present study we investigated whether an auditory preview cue indicating the number of patients with abnormal vital signs would help participants orient their attention better when listening to the upcoming display of multiple patients, and reduce the load on working memory.

2. LITERATURE REVIEW

In hospital wards, medical alarms are used to alert clinicians if the patient's vital sign is deviated from normal range or a medical device is not functioning properly. However, most of these are false alarms, which do not require any action from clinicians (Ruskin & Hueske-Kraus, 2015). The frequent exposure to false alarms is a threat to patient safety because clinicians will learn to delay their responses or ignore the alarms (Ruskin & Hueske-Kraus, 2015; Dewan et al., 2019). The patient may die due to the delayed treatment.

Auditory Display

Rather than redesigning the medical alarm, we proposed to use auditory display to monitor patients. Auditory display presents the patients' status, which is the vital sign levels, continuously and intermittently (Walker et al., 2013; Li et al., 2017; Li et al., 2019a). The vital sign levels are broadcasted sequentially based on their patient ID. By using auditory display, clinicians will know the actual vital sign levels of the patient when the patient's situation is deteriorating (Li et al., 2017). Also, clinicians can confirm the normality of patients with auditory display because it is announced in the display, but not in medical alarm.

Earcon and Spearcon are two types of auditory display. In previous research, earcon and spearcon were used to present the oxygen saturation (SpO_2) and heart rate (HR) levels of patients. Earcon conveys those vital signs by non-speech sound with manipulation of its parameters such as tremolo and timbre (Blattner et al., 1989; Hickling et al., 2017; Li et al., 2017). Spearcon presents the vital signs through time-compressed speech of user's native language (Davidson et al., 2019; Li et al., 2017; Li et al., 2019a; Li et al., 2019b; Sanderson et al., 2019; Walker et al., 2013). Training related to the mapping between the sounds and its meanings is necessary for target users to understand the messages from both earcon and spearcon. But spearcon is easier to learn compared with earcon because the mapping of spearcon is self-explanatory (Li et al., 2017; Walker et al., 2013). Also, comparing to conversational speech, the content presented by spearcon could be understood by target users only (Li et al., 2017; Sanderson et al., 2019). Then, in healthcare context, the spearcon would not disturb the other people but alert the clinicians only. Although earcons do not convey their meanings as readily as spearcons do, they are useful as families of alerts or simple notifications, and users can be trained to understand their meanings (Brewster, 1998).

In Hickling et al. (2017), a sequence of earcons was proposed to present the patients' status at regular interval, including a specific sound for patient with all normal vital signs. In subsequent work, and current study, spearcons are used to represent the SpO_2 and HR levels in a patient sequence (Li et al., 2019a; Sanderson et al., 2019). Non-clinician's participants

were required to identify the location in the sequence and vital sign levels of any patients with any abnormal vital signs.

Although the above results were promising, individuals' ability to identify patient's location in the sound sequence and abnormal vital signs decreased as the number of patients with abnormal vital signs increased (Hickling et al., 2017; Li et al., 2019a). The decrease was even greater when concurrent tasks required individuals to use verbal working memory (Li, et al., 2019b; Davidson et al., 2019). Limited working memory is the obstacle for users to recall the location and vital signs accurately in multitasking context (Davidson et al., 2019; Hickling et al., 2017; Li et al., 2019a; Li et al., 2019b).

Hickling et al. (2017) suggested that increasing the time interval between two sounds of patients' status in patient sequence could preserve the accuracy of identifying the location and vital signs of abnormal patients. It allowed users to consolidate and maintain the vital signs in working memory (Hickling et al., 2017).

However, this may not be an effective solution for monitoring multiple patients in multitasking context because of the following two reasons. First, the working memory allocated to the patient sequence may be less than the condition of having no current task because part of the verbal working memory is used for the current task in multitasking (Baddeley et al., 2001; Li et al., 2019b; Lin et al., 2016). It may require even longer time interval between two patients' status to preserve the identification performance as in Hickling et al. study. Second, the clinicians could not notice the urgent change of patients due to the delayed presentation of their vital signs with the even longer time interval, which may lead to delayed treatment. Therefore, it is necessary to find a way, other than increasing the time interval between two sounds of patients' status, to preserve participants' ability to report the location and vital signs as the number of abnormal patients increases.

Adding "Preview Cue" in Auditory Display

One potential solution is to borrow the idea from visual display design of a preview that gives users an expectation of the forthcoming details. Preview provides initial high-level information about the pattern or ordering of data in summary format (Greene et al., 2000; Shneiderman, 1996) whereas the subsequent display gives the detail of the data (Hornbæk & Hertzum, 2011). The users could determine whether or not to look at the details by having a brief prior understanding from the preview (Greene et al., 2000; Hasani et al., 2018). The preview can be presented in either visual or auditory modality, but regardless of the modality, the users cannot revisit the preview once it has been presented (Greene et al., 2000).

A preview cue for a visual interface has been shown to preserve both single-tasking (Qvarfordt et al., 2013; Schraefel et al., 2003) and multitasking performance (Reissland &

Manzey, 2016) regardless of the modality of the preview cue. Qvarfordt et al. (2013) showed that the visual preview helped users discover more useful information from the visual display compared with a no preview cue condition. In addition, Schraefel et al. (2003) found that when search for a target in a visual display, users were quicker to locate the target with the aid of auditory preview cue than with no preview cue. Further, in multitasking situations, Reissland and Manzey (2016) found that preview cue reduced users' task-switching time. Reissland and Manzey suggested that the preview cue provided the users expectation of an interruption, which allowed them to mentally prepare for the pending task switch while working on a concurrent task.

We postulated that this idea, and possibly its benefits, could be extended to the design of auditory displays, including our patient sequence display (Hickling et al., 2017; Li et al., 2019a; Sanderson et al., 2019; Davidson et al., 2019). We prolongate our patient sequence display by inserting an auditory preview cue before presenting the details of patients' status. However, the effectiveness of having a preview cue is unknown for a complex auditory display such as a patient sequence. Auditory displays using sound sequences to represent multiple patients' information can help listeners locate and identify abnormal vital signs with above-chance accuracy (Li et al., 2019a). But the participants note that locating patients with abnormal vital signs is not easy and identifying those vital signs is even harder (Li et al., 2019a). Therefore, it is unclear whether a preview would reduce the load on working memory sufficiently for performance to improve.

One implementation of a preview cue could indicate, in advance, how many patients have abnormal vital signs. It would suggest the amount of attention needed to find all necessary information from the upcoming patient sequence. This might reduce uncertainty and allow clinicians to anticipate the attention and working memory demands in the upcoming sequence. They could then allocate sufficient mental resource to identify the patients' status from the patient sequence.

A preview cue could take the form of either a spearcon or earcon as noted before and either could indicate the number of patients in an upcoming patient sequence who have abnormal vital signs. The meaning of a spearcon is more readily apparent than the meaning of an earcon (Li et al., 2017; Sanderson et al., 2019). Therefore, a spearcon-based preview cue (spearcon preview cue) may be more effective than earcon-based preview cue (earcon preview cue) at reducing the working memory demands associated with multiple-patient monitoring.

Effect on Concurrent Task Performance

In this study, the participants are required to carry out a concurrent task—a forced-pace arithmetic task—while monitoring the vital signs of multiple patients. The concurrent task performance in spearcon preview cue condition may be the worst compared with other preview cue types due to its higher emotional arousal. Röer et al. (2017) found that the focal attention on the concurrent task would be captured by changing and emotional arousing speech in the background, which degraded the concurrent task performance. Compare with earcon preview cue, spearcon preview cue is more emotional arousing because it presents the urgency of patient sequence in speech across time. It may capture the focal attention from concurrent task and allocate to patient sequence as in Röer et al. study. Therefore, the concurrent task performance in spearcon preview cue condition may be worse than earcon preview cue condition and no-cue condition.

Hypotheses

In the study reported herein, the primary outcomes are participant's accuracy at identifying (1) the location of abnormal patients in patient sequence ("patient location accuracy") and (2) the vital sign levels of abnormal patients ("vital sign accuracy"). We hypothesize (H1) that both accuracies will be better with a preview cue than without, and the increases will be larger for the spearcon preview cue than the earcon preview cue. We also hypothesize (H2) an interaction between preview cue type condition and number of abnormal patients. The accuracy improvements due to the presence of preview cue will be the smallest when there is no or only one abnormal patient in the sequence, and will increase as the number of abnormal patients in the sequence increases. The no-cue condition will show the greatest decreases in accuracy as the number of abnormal patients increases. This interaction is hypothesized because there is very little load on working memory to report the location and vital signs of one abnormal patient, but an increasing load on working memory as the number of abnormal patients increases (Hickling et al., 2017; Li et al., 2019a; Li et al., 2019b; Sanderson et al., 2019).

The secondary outcomes are related to the concurrent task performance. They are the accuracy of providing correct responses ("arithmetic attempted accuracy") and the rate of giving no response per minute ("arithmetic non-response rate) in the concurrent arithmetic task during the broadcast of patient sequence. Speech is particularly effective at capturing focal attention from the concurrent arithmetic task due to its linguistic form, which disrupts the performance of the concurrent task related to memory and number (Röer et al., 2017). Possibly, the spearcon preview cue may impair both arithmetic attempted accuracy and arithmetic non-response rate. The users may focus more on the patient sequence and neglect

the concurrent task after knowing the number of abnormal patients in the upcoming sequence from the spearcon preview cue. Therefore, the arithmetic attempted accuracy and arithmetic non-response rate are predicted (H3) to be worse with the spearcon preview cue than with the earcon preview cue and no-cue conditions.

3. METHOD

Design

The experiment used a 3 (preview cue type) x 4 (number of abnormal patients) mixed-factorial design. Preview cue type was a between-subjects factor with three levels: no-cue, earcon preview cue and spearcon preview cue. Number of abnormal patients in a sequence was a within-subjects factor with four levels: zero, one, two, or three abnormal patients. The experiment was conducted through Zoom because of COVID-19 restrictions on face-to-face testing. All tasks were completed on participant's computer under the instruction of an experimenter who followed a standardized running protocol.

The primary outcomes were the accuracy with which participants located abnormal patients in the five-patient sequence ("patient location accuracy") and identified vital sign levels for patients who had at least one abnormal vital sign ("vital sign accuracy"). The secondary outcomes were participants' accuracy at the concurrent arithmetic task ("arithmetic attempted accuracy"), the rate of giving non-responses to the arithmetic task per minute ("arithmetic non-response rate") during the broadcast of the patient sequence, and participants' responses to questions about the sounds.

Power Analysis

A pilot study with 18 non-clinician's participants (six being randomly assigned to each of the three conditions) was conducted. Mixed ANOVA was used to analyse the data for primary outcomes. The main effect of preview cue type on patient location accuracy was not significant, $F(2, 15)=1.43$, $p=.27$, $\eta_p^2=.16$. The interaction between number of abnormal patients and preview cue type on patient location accuracy was also not significant, $F(6, 45)=0.93$, $p=.49$, $\eta_p^2=.11$. However, the effect of number of abnormal patients on patient location accuracy was significant, $F(3, 45)=10.42$, $p<.001$, $\eta_p^2=.41$. For vital sign accuracy, the main effect of preview cue type was not significant, $F(2, 15)=1.15$, $p=.34$, $\eta_p^2=.13$. Also, the interaction between number of abnormal patients and preview cue type on vital sign accuracy was not significant, $F(6, 45)=0.47$, $p=.82$, $\eta_p^2=.06$. Similarly, the effect of number of abnormal patients on vital sign accuracy was significant, $F(3, 45)=49.35$, $p<.001$, $\eta_p^2=.77$.

Since we decided to detect the effect with given effect sizes and desired statistical power, a power analysis was conducted to calculate the required sample size for the main experiment. Based on the interaction between preview cue type and number of abnormal patients, the effect size (f) of patient location accuracy was 0.35 and the effect size of vital sign accuracy was 0.25. To provide sufficient power to find any significant interactions in both patient location accuracy and vital sign accuracy, we based the power analysis on the

smaller effect size of vital sign accuracy. Using G*Power with power = 0.95, $\alpha = 0.025$, the minimum total sample size required was 90 (Faul et al., 2009).

Participants

Ethical approval was obtained from Lingnan University's Sub-Committee on Research Ethics (reference number: EC031/1819). Email informed consent (See Appendix A2) was received from each participant before the experiment.

Ninety-three non-clinician participants were recruited in total through mass email sent to all students in Lingnan University, including 33 males and 60 females. The age of participants ranged from 17 to 38 years ($M = 20.7$, $SD = 2.9$). Inclusion criteria were: (a) ability to provide a Windows PC with headphones or earphones, (b) current enrolment at Lingnan University, (c) native speaker of Cantonese, (d) self-reported normal or corrected-to-normal hearing ability, (e) competence at using mouse and keyboard, and (f) no prior participation in any auditory display study. Participants who completed the experiment received HKD$150 (USD$19.25).

Since the present study was related to patient monitoring in hospital ward, it was more suitable to recruit clinicians as participants. However, clinicians was not available for the experiment due to COVID-19 pandemic. Also, the present study aimed at testing the usability of the preview cue by experiment, which could be achieved by testing non-clinician participants. Therefore, we recruited non-clinician in the present study and clinicians could be recruited for future investigations.

Sound Stimuli

Figure 1 shows a typical sequence of sounds in each patient sequence. First, there was an *alert signal* to alert users to the start of patient sequence. The alert signal was a 'ding' sound with a fundamental frequency of f_0=2926 Hz that lasted for 700 ms.

In the preview cue type conditions, a *preview cue* then indicated the number of abnormal patients in the upcoming patient sequence. For the spearcon preview cue, the number was presented by Cantonese time-compressed speech: it was either "all normal" (「全正常」) or if there were one, two, or three abnormal patients, the phrase would be "one person" (「一個」), "two people" (「兩個」), or "three people" (「三個」) respectively. After compression, the duration for the "all normal" spearcon preview cue was 477 ms, and for the one, two, or three abnormal patients spearcon preview cue was 257 ms, 254 ms, or 248 ms respectively.

Each earcon preview cue consisted of a 355 Hz sine wave tone punctuated by zero, one, two, or three pulses of white noise, representing the number of patients with abnormal vital

signs in the upcoming patient sequence. The sine wave tone lasted for 474 ms and the duration for one, two, or three abnormal patients earcon preview cue was 471 ms, 465 ms, or 497 ms respectively. If there was no abnormal patient, there would be no white noise and only the sine wave tone would be played.

Cantonese Spearcons. The Cantonese spearcons were the same as in Li et al. (2019a). The vital signs for each patient included SpO_2 (represented by「氧」, meaning "oxygen") and HR (represented by「心」, meaning "heart"). Each vital sign has five levels: "very high" (「好高」), "high" (「高」), "normal" (「正常」), "low" (「低」), and "very low" (「好低」). For each patient, participants heard a sequence of four spearcon elements: the SpO_2 label, SpO_2 level, the HR label, and HR level. The spearcon elements were recordings of a native Cantonese-speaking female voice and each was time compressed to 25% of its original length yielding spearcons that were around 500 ms long. Spearcons were used only for patients with one or more abnormal vital signs. If both vital signs were normal for a patient, a 500-ms "boop" tone would be played instead of a "SpO_2 normal HR normal" spearcon.

Patient Sequence. The preview cue started playing 1 s after the alert signal. The sequence of spearcons started playing 1.5 s after the preview cue (see Figure 1). In no-cue condition, the sequence of spearcons started playing 1 s after the alert signal. There were 800-ms between spearcons. Because the preview cue had different durations, the duration of patient sequences differed across preview cue type conditions. The average duration of patient sequences without preview cue was 7.4 s, while in the spearcon preview cue and earcon preview cue conditions it was 9.2 s and 9.4 s respectively.

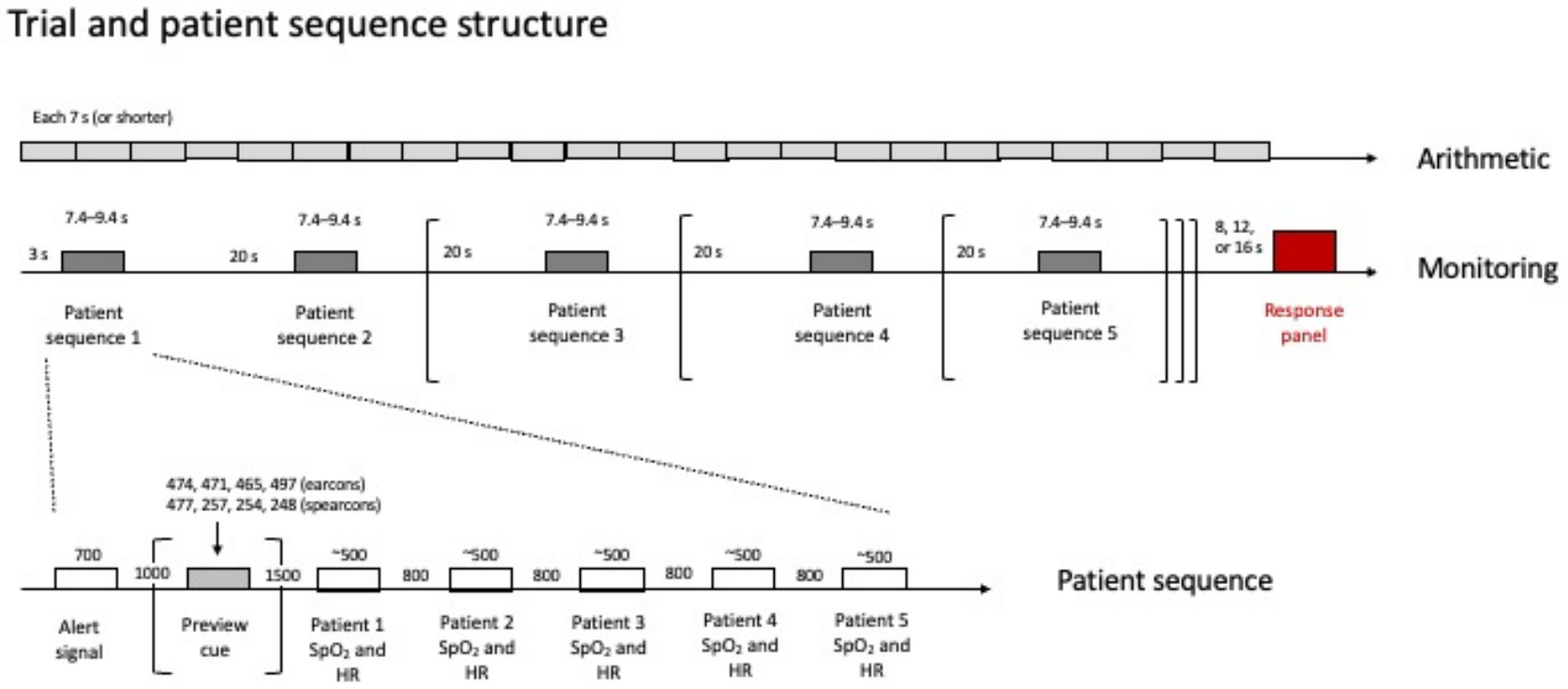


Figure 1. Structure of experimental trials. Each trial presented either two, three, four or five patient sequences before the response panel appeared. Each patient sequence was structured as in the lower timeline, with or without a preview cue, where times are indicated in ms. Earcon and spearcon preview cue durations are given for zero, one, two, and three patients with abnormal vital signs. The arithmetic task presented equations every 7 s, unless the participants responded before 7 s, at which point a new equation was immediately presented.

Tasks

Monitoring Task. Participants listened to a series of patient sequences, and at an unpredictable point after two, three, four, or five patient sequences, a response screen appeared (see Figure 1). The response screen provided a panel for each the five patients (see Figure 2). Participants identified the level of both vital signs (SpO_2 and HR) of any patient who had either one or two abnormal vital signs, including any vital sign that was normal for that patient. For patients whose vital signs were both normal, however, participants were told not to enter responses. The monitoring task resumed as soon as the participant clicked the "Back to testing" button.

Concurrent Arithmetic Task. The concurrent arithmetic task was a metaphor of the administrative tasks in a medical-surgical ward, such as performing drug dosage calculations. An equation with two-digit addition or subtraction was displayed in the center of the screen (see Figure 3). For each equation, participants determined its correctness and clicked either the "TRUE" or "FALSE" button using the mouse. Feedback stating that the answer was "CORRECT" or "WRONG" immediately appeared under the two buttons for around 1 s before the next equation appeared. If a participant did not select an answer within 7 s a new equation appeared; this would be counted as a non-response.

Trial structure. Following Li et al. (2019b), each trial contained two to five patient sequences followed by a response panel at 8, 12, or 16 s after the last patient sequence. As shown in Figure 1, the first patient sequence played 3 s after the beginning of trial after which the inter-sequence interval was set at 20 s. The order in which trials were presented was randomly generated by the software to eliminate learning effects or bias.

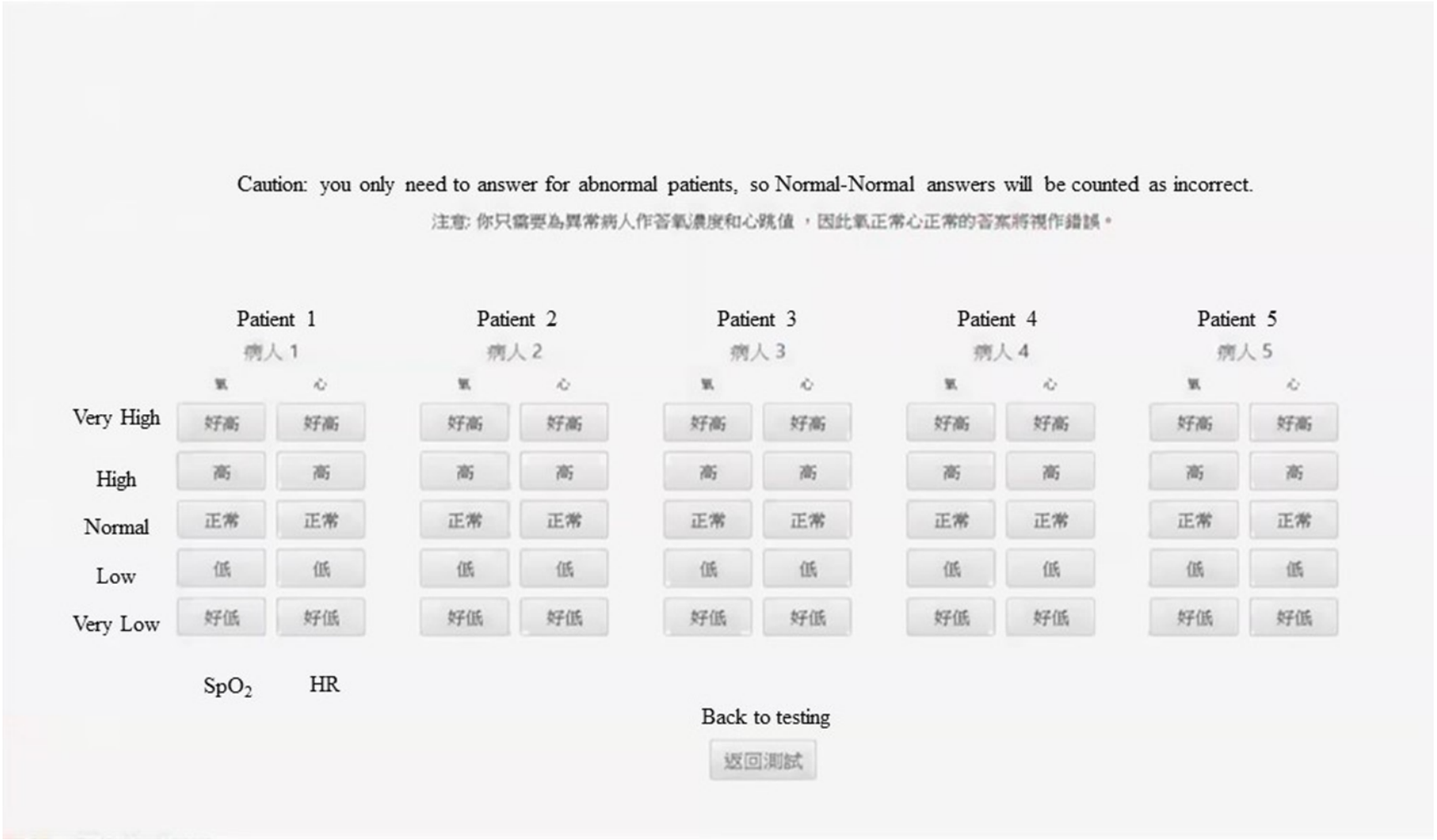


Figure 2. Response panel for the monitoring task

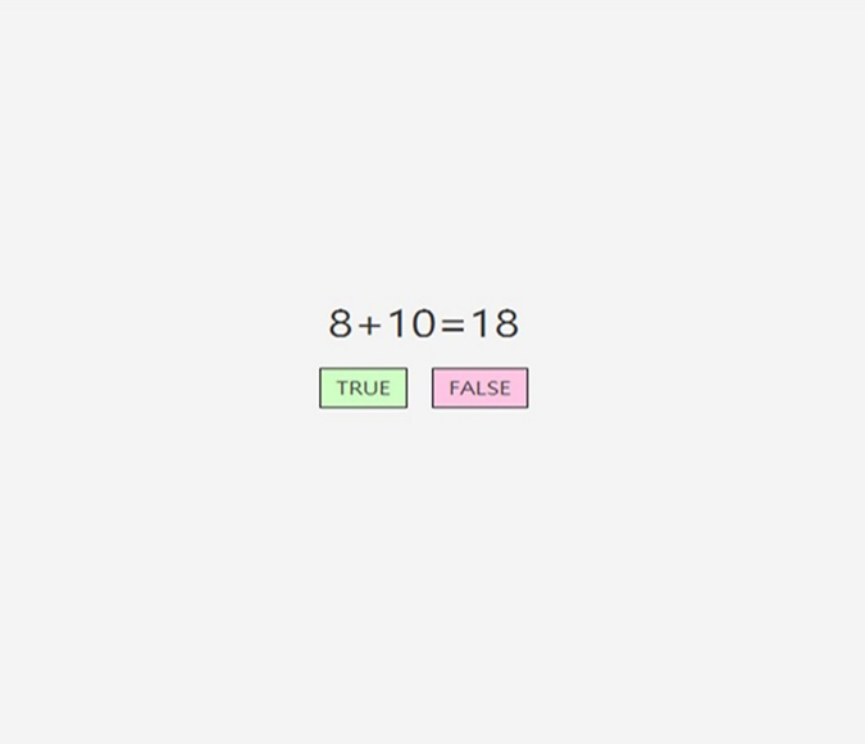


Figure 3. Software interface for the concurrent arithmetic task.

Questionnaires

Pre-testing Questionnaire. The pre-testing questionnaire collected data about each participant's age, gender, music training experience, and hearing ability (See Appendix A3).

Post-testing Questionnaire. To understand user's subjective feeling of identifying the location and vital signs of abnormal patients with or without preview cue, the post-testing questionnaire investigated participants' perceived difficulty and confidence of completing the monitoring task and the concurrent arithmetic task. Participants gave their responses on an ordinal scale ranging from 1 (not difficult at all/ not confident at all) to 9 (extremely difficult/ extremely confident). Open-ended questions probed participants' strategy for remembering the location and vital signs of abnormal patients, and their judgment of the effectiveness of the preview cue on the monitoring task if they were in spearcon or earcon preview cue condition (See Appendix A5).

Outcome Measures

Participants' responses in both concurrent arithmetic task and monitoring task were collected by the custom software. As shown in Figure 1, a task trial started when the concurrent arithmetic task started and ended after the participant entered their responses in the monitoring task response panel. Performance in each trial was determined as follows.[1]

Patient Location Accuracy. Patient location accuracy measured participants' accuracy at identifying the location of any patients in the sequence with abnormal vital signs. To determine this, each patient in the sequence was scored as either correct or incorrect. A normal patient was scored as correct if the participant had made no entry for abnormal level (i.e., Very High, High, Low, Very Low) in either vital sign column. An abnormal patient was scored correct for *location* if the participant made an entry for abnormal vital sign level (i.e., Very High, High, Low, Very Low) in either vital sign column and press "Normal" in another column or in both vital sign columns, even if the wrong vital sign level was selected. Normal patients with abnormal vital sign entries, and abnormal patients with no vital sign entries were scored as incorrect.

Patient location accuracy for each trial was calculated as follows.

$$Patient\ Location\ Accuracy(\%) = \frac{No.\ of\ correctly\ located\ patients}{5} \times 100\% \quad \text{(Eq. 1)}$$

[1] Formulae are slightly different from those in Hickling et al. (2017), Li et al. (2019a), Sanderson et al. (2019) and others. The present study included trials in which there were no abnormal vital signs, which was not the case for the previous studies. Figures of merit had to reflect accurate responding in such trials to test the effect of preview cue even for all-normal patient sequence.

Vital Sign Accuracy. Vital sign accuracy measured the performance of identifying 10 vital sign levels. The vital sign was correctly identified only if the participants chose the correct vital sign level for the correctly located patient.

Vital sign accuracy for each trial was calculated as follows.

$$Vital\ Sign\ Accuracy(\%) = \frac{No.\ of\ correctly\ identified\ vital\ signs}{10} \times 100\% \quad \text{(Eq. 2)}$$

Arithmetic Attempted Accuracy. Performance on the concurrent arithmetic task was calculated as proportion accuracy across all trials. Number of attempts was the total number of correct and incorrect responses in the concurrent arithmetic task. Number of correct responses was the total number of correct responses to the concurrent arithmetic task.

$$Arithmetic\ Attempted\ Accuracy(\%) = \frac{No.\ of\ correct\ responses}{No.\ of\ attempts} \times 100\% \quad \text{(Eq. 3)}$$

Arithmetic Non-Response Rate. The total number of equations that the participant did not respond to during all trials was counted. To avoid the confound of time differences in different patient sequences, the total number of non-responses was divided by the product of average duration of patient sequence in the same preview cue conditions and 56 which was the total number of patient sequences. By multiplying by 60 seconds, the arithmetic non-response rate can be interpreted as the rate per minute at which the participant was ignoring the concurrent arithmetic task while patient sequences were being played.

$$Arithmetic\ Non-Response\ Rate = \left(\frac{Total\ no.\ of\ non-responses}{Average\ duration\ of\ patient\ sequence \times 56}\right) \times 60 \quad \text{(Eq. 4)}$$

Procedure

There were four phases in the experiment: (i) introduction, (ii) familiarization, (iii) testing, and (iv) debriefing.

Introduction. Eligibility for participation was confirmed verbally. The participant completed the pre-testing questionnaire and the experimenter explained the principle of using spearcons to monitor multiple patients.

Familiarization. In each subpart of the familiarization phase, the participant learned about one aspect of the testing phase and could practice it. They received feedback and could have a second attempt if they answered incorrectly the first time. First, the participant watched a video demonstrating how to complete the concurrent arithmetic task, then they practiced the task for 30 s. Second, the participant was introduced to all the spearcons, learned how to monitor one patient using spearcon, and completed four practice trials at identifying the spearcon. Third, the participant learned how to monitor multiple patients using the spearcon or earcon preview cue and then the patient sequence and they completed two practice trials. Finally, the participant learned how to do the monitoring and concurrent arithmetic tasks at the same time, and practiced it (See Appendix A4 for the training answer sheet).

Testing. Participants performed 16 trials of the monitoring task. During each trial, participants also performed the concurrent arithmetic task. When the 16 trials were completed, participants completed the post-testing questionnaire.

Debriefing. The participant was compensated for their participation and informed about the purpose of this study by an educational debrief sheet (See Appendix A6).

Statistical analysis

For the primary outcomes, mixed ANOVAs were used with a between-subjects factor of preview cue type condition and within-subjects factor of number of abnormal patients. For the secondary outcomes, a one-way ANOVA was used with a between-subjects factor of preview cue type condition. The Greenhouse-Geisser correction was adopted if sphericity was violated. Residuals analyses were carried out to test conformity of data to assumptions on ANOVA. Mixed ANOVA with experimenter and preview cue type conditions as the between-subject factors and number of abnormal patients as within-subject factor was used to examine whether there was experimenter effect on all outcomes. The Likert-scale responses from post-testing questionnaire were analysed with Kruskal-Wallis test if the normality assumption was violated in one-way ANOVA.

4. RESULTS

Ninety participants were initially recruited. The data from three participants were excluded either for technical reasons or because the participant did not follow instructions. For example, one participants revealed that she used pen and paper to note down the answers for monitoring task while listening to the patient sequence, which was prohibited. They were replaced by another three participants who were recruited during the data collection period. Data from two participants were also excluded due to evidence that they had responded at random on both tasks as found in data analysis process. Four participants did not follow the instructions in answering the monitoring task on one trial by pressing "Normal" in both vital sign columns for a patient, so those specific data points were excluded. An extra trial with one abnormal patient occurred in one session, and the data point from that trial was also excluded. The final number of patients for data analysis was N=88.

Although the main experiment was executed by two experimenters, there was no experimenter effect on all outcomes (See Appendix B1).

Residuals were negatively skewed for all ANOVA analyses, except the arithmetic non-response rate which was positively skewed. Exponential transformation and square root transformation were performed respectively, but the change of residual distributions was minor with no effect on the substantive outcomes (See Appendix B2). Therefore, the untransformed results are reported.

Most of the outcomes met the assumption of homogeneity of variance, as assessed with the Levene's test. The assumption of homogeneity of variance was not met for the location and vital sign accuracy in the zero abnormal patient condition, because all but three participants obtained maximum scores of 100%. Given the very small difference in sample size in each preview cue type condition, it was concluded that the non-homogeneity of variance in the zero abnormal patient condition would not distort the substantive outcome.

Primary Outcomes

Patient Location Accuracy. Results are summarized in Table 1 and Figure 4. The main effect of preview cue type on patient location accuracy was not significant, $F(2, 85)=0.65$, $p=.52$, $\eta_p^2=.02$, and the interaction between number of abnormal patients and preview cue type was not significant, $F(4.92, 208.88)=0.45$, $p=.81$, $\eta_p^2=.01$. However, participants' accuracy at identifying the location of abnormal patients in the sequence decreased as the number of abnormal patients increased, $F(2.46, 208.88)=87.33$, $p<.001$, $\eta_p^2=.51$. Post-hoc tests showed that overall patient location accuracy decreased as the number of abnormal patients increased from zero ($M = 99\%$, $SD = 3\%$) to one ($M = 93\%$, $SD = 9\%$), $t(85)=7.49$, $p<.001$, from one to two ($M = 86\%$, $SD = 16\%$), $t(85)=4.46$, $p<.001$, and from two to three ($M = 78\%$, $SD = 15\%$), $t(85)=5.87$, $p<.001$.

Vital Sign Accuracy. Results are summarized in Table 1 and Figure 5. The main effect of preview cue type on vital sign accuracy was not significant, $F(2, 85)=0.14$, $p=.87$, $\eta_p^2 =.003$ and the interaction between number of abnormal patients and preview cue type was not significant, $F(4.49, 190.97)=0.06$, $p=1.00$, $\eta_p^2 =.001$. Again, vital sign accuracy dropped as the number of abnormal patients increased, $F(2.24, 190.97)=314.07$, $p<.001$, $\eta_p^2 =.79$. Post-hoc tests showed that overall vital sign accuracy decreased as the number of abnormal patients increased from zero ($M = 100\%$, $SD = 2\%$)[2] to one ($M = 94\%$, $SD = 7\%$), $t(85)=8.56$, $p<.001$, from one to two ($M = 82\%$, $SD = 12\%$), $t(85)=11.17$, $p<.001$, and from two to three ($M = 64\%$, $SD = 14\%$), $t(85)=12.34$, $p<.001$.

Secondary Outcomes

There was also no significant difference in arithmetic attempted accuracy across preview cue type conditions, $F(2, 85)=0.25$, $p=.78$, $\eta_p^2 =.01$ (see Table 1 and Figure 6). However, there was a significant difference in the arithmetic non-response rate across preview cue type conditions, $F(2, 85)=3.12$, $p=.049$, $\eta_p^2=.07$, with non-responses being more frequent in the earcon preview cue condition ($M = 2.21$, $SD = 1.72$) than in the no-cue condition ($M = 1.20$, $SD = 1.45$), $t(85)=-2.50$, $p=.04$ (see Table 1 and Figure 7).

Table 2 shows the results for the post-testing questionnaire. There was no significant difference among preview cue type conditions in the scores on all questions (subjective difficulty and confidence in identifying the locations, vital sign levels, and completing the concurrent arithmetic task).

Equivalence Test

To avoid an inappropriate interpretation of null results, an equivalence test proposed by Lakens et al. (2018) was conducted. The equivalence bounds were set as the raw mean difference between two preview cue type conditions. The equivalence bounds were ± 10% because Lim and Sanderson (2019) adopted 10% as a criterion for clinical relevance when examining whether an alternative earcon improved the accuracy of identifying the vital signs compared with a control earcon.

Table 3 showed the results of equivalence tests in pairwise comparisons across three preview cue types. Location accuracy and vital sign accuracy were statistically equivalent across three preview cue type conditions for all numbers of abnormal patients except for location accuracy with two abnormal patients, where the non-significance of the equivalence

[2] In this case, the mean and standard deviation were, in fact, 99.7% and 1.7% respectively corrected to one decimal place. To keep reporting the percentage in the main text consistent, the means and standard deviations were rounded to the nearest whole number.

test was caused by one outlying data point that could not be excluded under our criteria. The effect of preview cue type on arithmetic attempted accuracy was also statistically equivalent.

Table 1

Descriptive statistics for all performance measures, by preview cue type and by number of abnormal patients.

	No-Cue	Earcon Preview	Spearcon Preview
	M (SD) [CI]	M (SD) [CI]	M (SD) [CI]
Patient location accuracy			
0 abnormal patient	100% (0%) [100%, 100%]	98% (5%) [96%, 100%]	100% (0%) [100%, 100%]
1 abnormal patient	92% (8%) [89%, 95%]	92% (10%) [88%, 96%]	93% (10%) [90%, 97%]
2 abnormal patients	88% (14%) [83%, 93%]	83% (18%) [76%, 90%]	88% (14%) [83%, 93%]
3 abnormal patients	77% (14%) [72%, 82%]	77% (16%) [71%, 82%]	79% (15%) [73%, 84%]
Vital sign accuracy			
0 abnormal patient	100% (0%) [100%, 100%]	99% (3%) [98%, 100%]	100% (0%) [100%, 100%]
1 abnormal patient	94% (6%) [92%, 96%]	93% (8%) [90%, 96%]	94% (8%) [92%, 97%]
2 abnormal patients	82% (11%) [78%, 86%]	81% (13%) [77%, 86%]	83% (11%) [79%, 87%]
3 abnormal patients	64% (14%) [59%, 69%]	64% (14%) [59%, 69%]	65% (16%) [59%, 70%]
Arithmetic attempted accuracy			
Overall	86% (10%) [82%, 89%]	84% (12%) [80%, 88%]	85% (10%) [82%, 89%]
Arithmetic non-response rate (per minute)			
Overall	1.20 (1.45) [0.68, 1.72]	2.21 (1.72) [1.58, 2.83]	1.68 (1.47) [1.14, 2.22]

Note. No-Cue = no-cue condition; Earcon Preview = earcon preview cue condition; Spearcon Preview = spearcon preview cue condition; CI = 95% confidence interval for mean.

Table 2

Descriptive statistics and Kruskal-Wallis Test results of post-testing questionnaire responses, by preview cue type.

	No-Cue		Earcon Preview		Spearcon Preview		Kruskal-Wallis Test		
Question	Median	CI	Median	CI	Median	CI	χ^2 (2)	*p* value	Effect size (ε^2)
1: Subjective difficulty identifying location	6	[5, 7]	6	[4, 7]	6	[5, 7]	0.37	.83	.004
2: Confidence in identifying location	5	[4, 6]	5	[4, 6]	6	[4, 6]	1.16	.56	.01
3: Subjective difficulty identifying vital signs	7	[7, 7]	6	[5, 7]	7	[7, 8]	4.02	.13	.05
4: Confidence in identifying vital signs	4	[3, 5]	4	[3, 5]	4	[3, 5]	1.08	.58	.01
5: Subjective difficulty in arithmetic task	7	[6, 8]	7	[6, 8]	7	[6, 8]	0.39	.82	.004
6: Confidence in arithmetic task	5	[4, 6]	4	[3, 5]	4	[3, 5]	2.28	.32	.03

Note. CI = 95% confidence interval for median

Table 3

Equivalence test results of location accuracy, vital sign accuracy, and arithmetic attempted accuracy, by preview cue types' pairwise comparisons.

		No-Cue vs Earcon Preview		No-Cue vs Spearcon Preview		Earcon preview vs Spearcon preview	
Outcome Measures	Number of abnormal patients	t(57)	*p* value	t(57)	*p* value	t(56[a])	*p* value
Location accuracy	0	-12.30	<.001**	N/A[b]	N/A[b]	-12.09	<.001**
	1	-4.09	<.001**	-4.88	<.001**	3.22	.001**
	2	1.24	.11	-2.77	.004**	1.19	.12
	3	2.38	.01*	2.21	.02*	1.92	.03*
Vital sign accuracy	0	-20.92	<.001**	N/A[b]	N/A[b]	-20.56	<.001**
	1	-6.06	<.001**	-6.21	<.001**	-5.66	<.001**
	2	3.03	.002**	3.23	.001**	2.79	.004**
	3	-2.68	.01*	2.47	.01*	2.56	.01*
Arithmetic attempted accuracy	N/A[c]	2.87	.003**	-3.62	<.001**	2.95	.002**

Note. * = p < .05; ** = p < .01;

[a]Two participants (one in earcon preview and one in spearcon preview) were removed from data analysis due to reasons outlined in Results;

[b]The data for both no-cue and spearcon preview cue conditions were the same, therefore the effect of preview cue types would be the same and the equivalence test was not applicable;

[c]The effect of number of abnormal patients was not examined in arithmetic attempted accuracy, therefore it is not applicable.

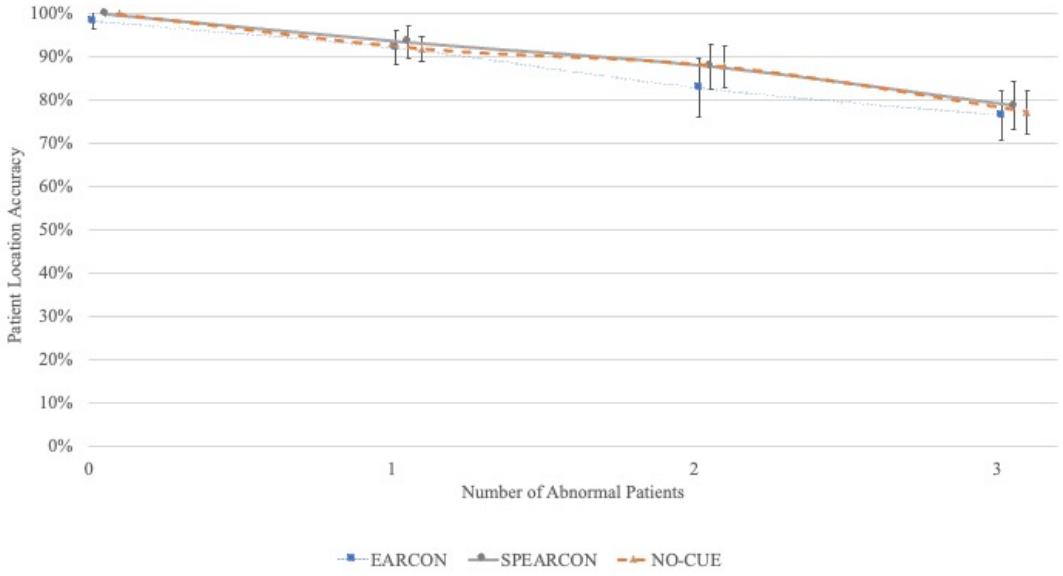


Figure 4. Accuracy of identifying the locations of abnormal patients with no-cue, earcon preview cue, and spearcon preview cue. Error bars are 95% confidence intervals for mean.

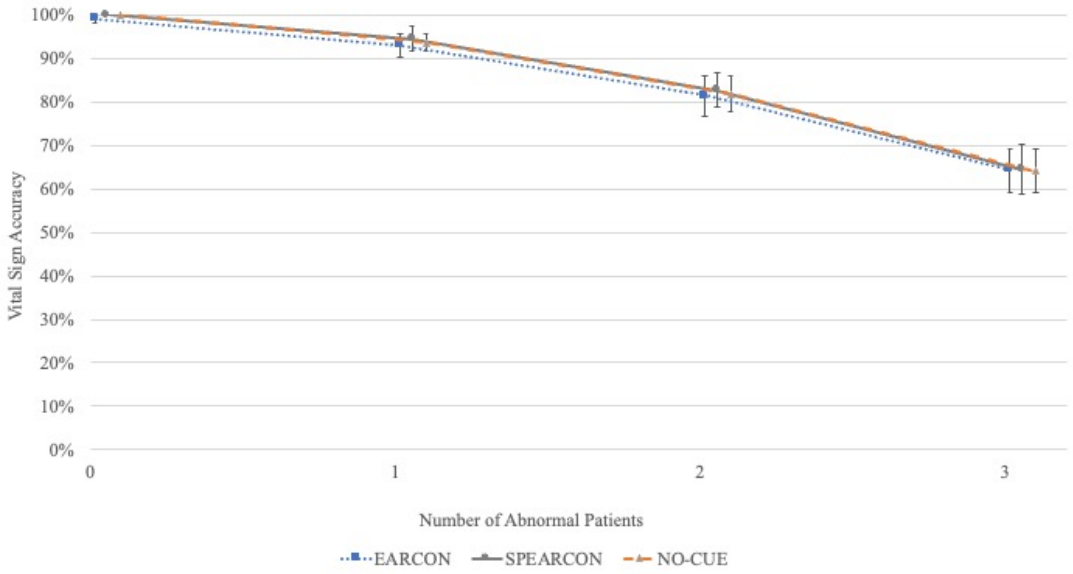


Figure 5. Accuracy of identifying the vital sign levels of abnormal patients with no-cue, earcon preview cue, and spearcon preview cue. Error bars are 95% confidence intervals for mean.

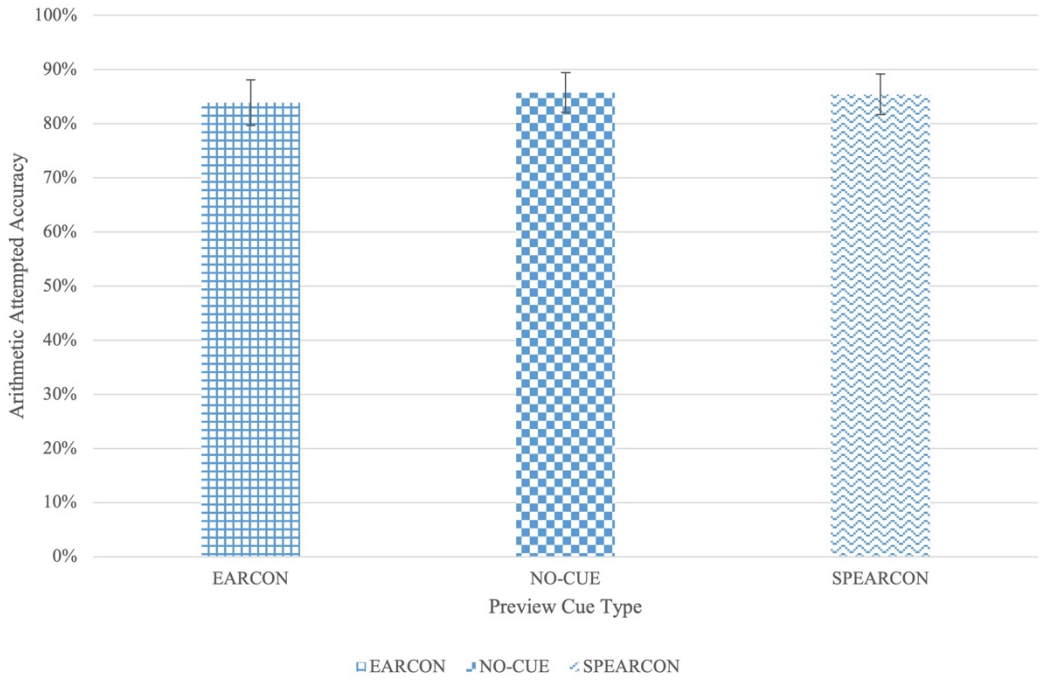


Figure 6. Accuracy of responding to the concurrent arithmetic task during the broadcast of patient sequence with no-cue, earcon preview cue, and spearcon preview cue. Error bars are 95% confidence intervals for mean.

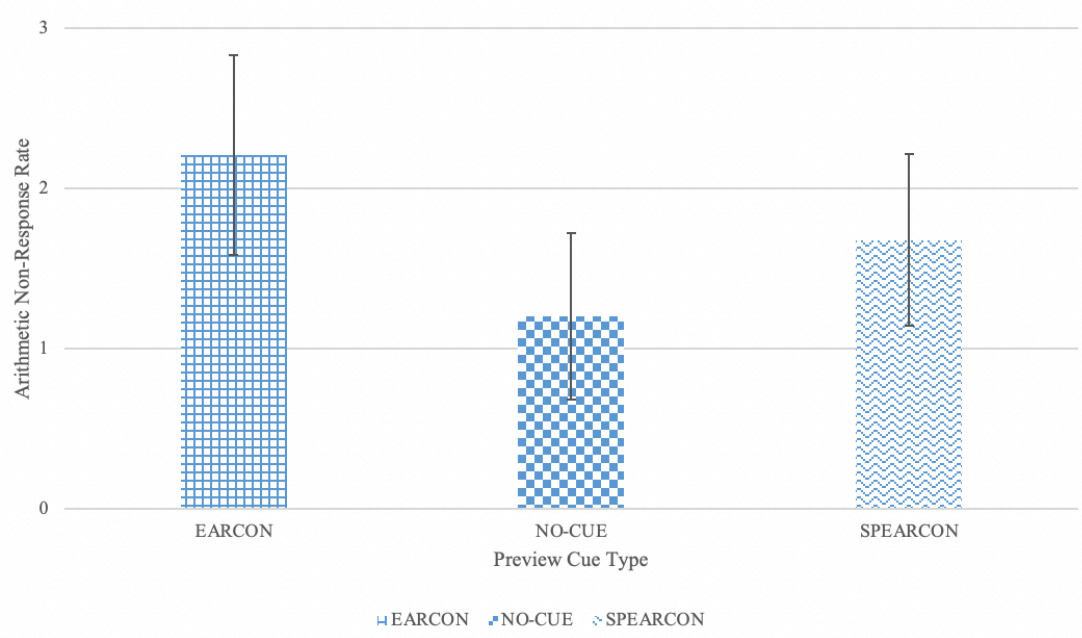


Figure 7. Rate of giving no responses in concurrent arithmetic task per minute during the broadcast of patient sequence, with no-cue, earcon preview cue, and spearcon preview cue. Error bars are 95% confidence intervals for mean.

5. DISCUSSION

Overall, the current preview cue designs, with earcon or spearcon, did not improve the performance of monitoring multiple patients. Contrary to H1, the experiment did not show that the spearcon or earcon preview cue improved location accuracy or vital sign accuracy, compared with the no-cue condition. Moreover, contrary to H2, as the number of abnormal patients increased, accuracy decreased to the same extent for all preview cue type conditions rather than to a greater extent in the no-cue condition than the other conditions. There was also no evidence to suggest the current spearcon preview cue led to worse concurrent arithmetic task performance than other preview cue type conditions.

Limitations of Current Preview Cues

Recalling patient information from auditory patient sequence is still a challenge even in the current study. Prior evidence suggests that similar working memory challenges exist when the patient sequence is presented visually (Sanderson et al., 2019). People can store three to five chunks of items in working memory. But if the items to be recalled are presented very closely in time in either modality, there is interference among them and the participant does not have enough time to encode them (Cowan, 2010). Given that there were only 800 ms between spearcons in the patient sequences in the present study, the participants might not have enough time to encode both the location and vital sign levels of abnormal patients. Therefore, the patient location and vital sign accuracy were similar to previous studies (Hickling et al., 2017; Li et al., 2019a; Li et al., 2019b; Sanderson et al., 2019).

The suitable method to preserve the location and vital sign accuracy still remains unknown. As argued, increasing the time interval between two sounds of patients' status in patient sequence is not effective in clinical context due to the possibility of delayed treatment. Also, the present study found that inserting the current preview cue in patient sequence, which indicated the number of abnormal patients, did not preserve the accuracies. There are two reasons for this phenomenon.

First, the current auditory preview cue provided insufficient information to guide the users to find the answers directly. The preview cue in current study provided only minimum information about the upcoming patient sequence—the *number* of abnormal patients—whereas participants were asked about the *location* of abnormal patients in the sequence and the levels of their vital signs, both of which changed across trials. Preview cues usually work by informing users to the area in which the answer is to be found (Singleton, 1971) or by indicating the general pattern of results at a high level (Shneiderman, 1996). However, the current auditory preview cue did not assist in constraining the possible responses in these ways; no direct information was given about patient location or about vital sign levels. In the post-testing questionnaire, a participant from earcon preview cue condition said the preview

cue was not useful for identifying the location of abnormal patients because "(I can) only know the number of (abnormal) patients, but not their location" from the preview cue. Also, the participants from both spearcon and earcon preview cue conditions revealed that the preview cue did not help identify the abnormal vital signs because "the oxygen saturation and heart rate for each individual were different." These responses showed that the participants still had to retrieve all the information needed for their responses from the *detail* part of the patient sequence even the preview cue was included. Therefore, the current preview cue did not preserve the location and vital sign accuracy by guiding their attention allocation to find the necessary information of abnormal patients in patient sequence. In future design, the information provided by the preview cue should match with the user's needs in completing the monitoring task, which is to include sufficient information about the details of patient sequence for the identifications.

Second, since the monitoring task was excessively difficult, the preview cue could not preserve the performance of identifying abnormal patient's location and vital signs. With three abnormal patients and therefore six vital signs to recall, participants were reaching the limits of dynamic working memory and were probably experiencing some degree of output interference (Hickling et al., 2017). With output interference, the items' content will interfere the memory performance because of its similarity. The latest vital signs might not be sufficiently consolidated in verbal working memory and they might be overwritten by previously memorized vital signs consequently when there were more abnormal patients (Hickling et al., 2017; Roediger, 1974). Also, the participants were lack of confidence in identifying the location and vital signs and felt difficult on both identifications (See Table 2). With such high task difficulty, the current additional preview may even lead to cognitive fatigue (Chen et al., 2018). Hence, the preview cue did not work as predicted when the monitoring task was extremely difficult. The task difficulty should be considered when we design a preview cue that intends to preserve the monitoring task performance in the future.

Effect of Current Preview Cue on Concurrent Task

The current preview cue did not affect how accurate the participants responded to the concurrent arithmetic task during patient monitoring. But there were more non-responses in earcon preview cue condition than no-cue condition.

There was similar arithmetic attempted accuracy in concurrent arithmetic task because the participants in different preview cue type conditions might undergo the same process when they were responding to it. Regardless of the preview cue type, the participants needed to listen and rehearse the location and vital signs from patient sequence. Concurrently, they needed to respond to the concurrent arithmetic task. Both tasks burden the phonological loop, which is responsible to process verbal and auditory information. In this case, the

performance of the concurrent arithmetic task might be interfered by the patient sequence because of the resource competition in phonological loop (FÜrst & Hitch, 2000). Twenty-five participants across different preview cue type conditions revealed that they would "keep murmuring the oxygen saturation and heart rate levels they have heard repeatedly" while responding to the concurrent arithmetic task in the post-testing questionnaire. Therefore, both concurrent arithmetic task and monitoring task burdened the phonological loop heavily regardless of the preview cue type, which led to similar results in arithmetic attempted accuracy.

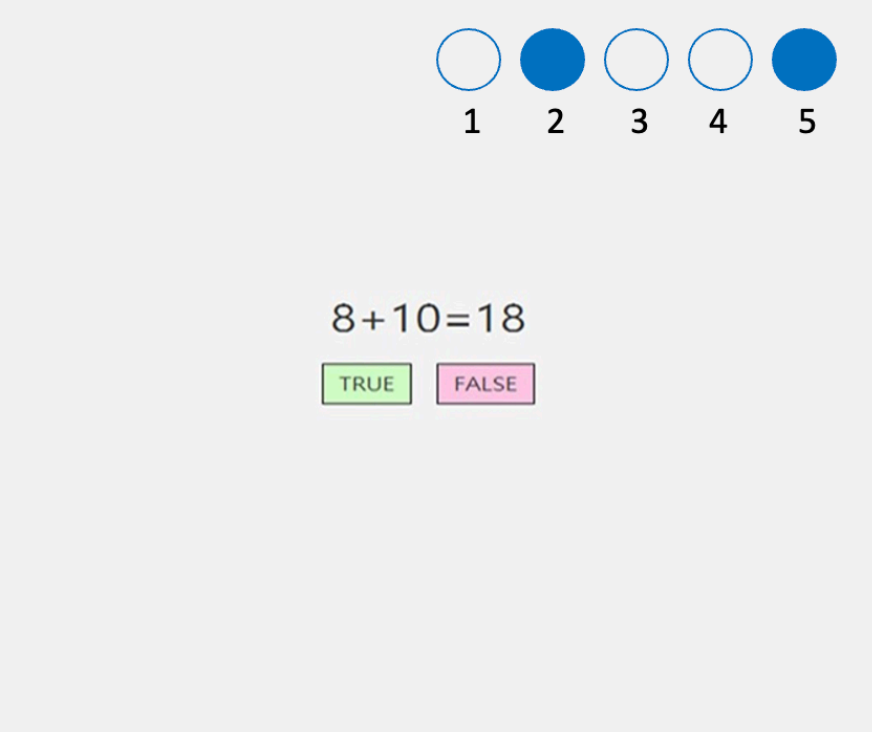
Although the arithmetic attempted accuracy was similar across preview cue type conditions, more non-responses appeared in earcon preview cue condition than no-cue condition. It implied that the participants tended not to respond to the concurrent arithmetic task if they were using earcon preview cue compared with no-cue. It is possible that the abstractness of earcon preview cue may motivate participants sometimes to suspend their arithmetic responses. Given that the earcon is an abstract sound, users have to associate its sound to its meaning when they process it (Li et al., 2017). As noted by Watson and Gill (2004), participants can become distracted when they must process a newly-learned earcon sound. In our experiment, when participants heard the earcon preview cue, they might suspend the response to the concurrent arithmetic task in order to preserve their performance on the monitoring task. Although the responding accuracy of concurrent arithmetic task would not be affected by the preview cue, the preview cue may motivate the participants to suspend the response in arithmetic task, which impairs the concurrent task performance in another way. Hence, the preview cue may harm the concurrent task performance by motivating to suspend the response.

Design Implication

Auditory sequences representing complex information can impose heavy demands on listeners. A design implication is to have specific but cognitive undemanding content in an auditory preview cue, in this case indicating which patient(s) have abnormal vital signs rather than the number of patients with abnormal vital signs. For example, the content presented by the auditory preview cue could be more specific, e.g., "Two-Five", to indicate patients in the second and fifth location are in abnormal states. Another example is to provide a visual preview cue which allows the users to see the ID of abnormal patients in the upcoming patient sequence on both screens for concurrent arithmetic task and monitoring task until the new patient sequence is about to broadcast (See Figure 8). But the specific auditory preview cue is preferred because it supports eyes-free monitoring, which may not impair the visual concurrent task compared with the visual preview cue. With the new auditory preview cue, the performance of identifying the location and vital signs of abnormal

patients may be preserved by guiding the attention allocation and reducing much memory load.

However, the specificity of preview cue should be carefully balanced with respect to the detailed view. Otherwise, it loses the purpose of being a preview, which is to give listeners a high-level summary before the details. Further testing about the effectiveness of auditory display with the new preview cue should be conducted before implementing it in the healthcare context.



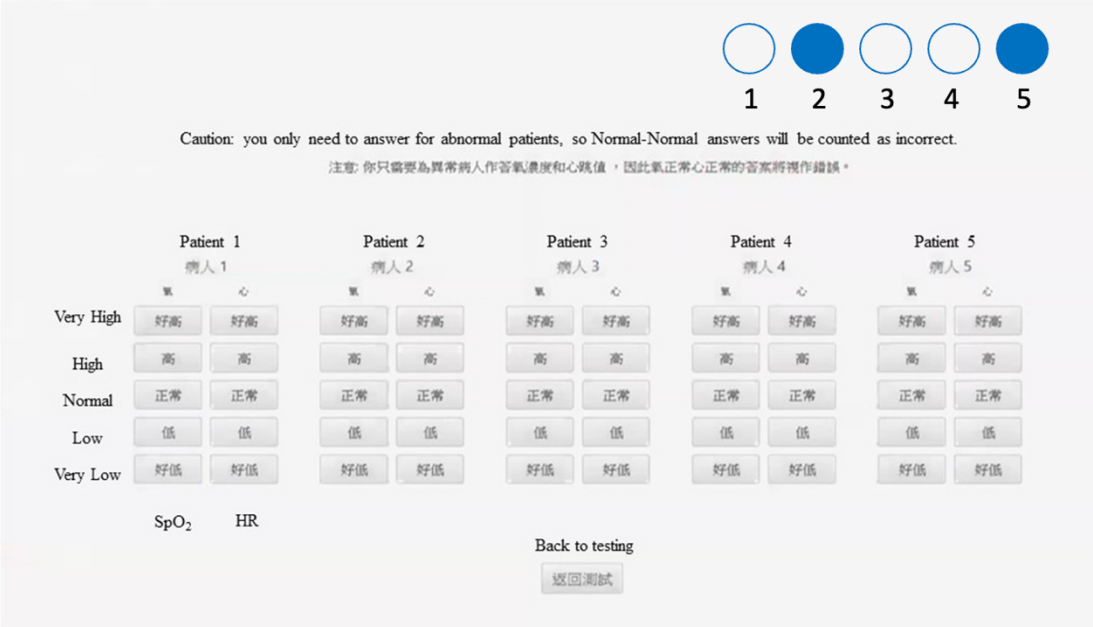


Figure 8. The illustrations of having visual preview cue (at the top right-hand corner) on screen. It indicates that patient two and patient five are abnormal in the current patient sequence. It appears on both screens for concurrent arithmetic task and monitoring task until the new patient sequence is about to broadcast.

6. LIMITATIONS AND FURTHER STUDIES

There are four limitations that can be addressed in future research. First, participants revealed that they memorized the locations of abnormal patients using their fingers and auditory imaginary to offload memory in the post-testing questionnaire. But this strategy may not be available in clinical context. The clinicians may have other ways to remember the details of abnormal patients because of their professional experience. In the future studies, clinicians working in hospital wards could be the participants. Nevertheless, the current study has shown the limitations of the current preview cue design and implied that redesigning the preview cue is necessary before conducting further testing.

Second, the response demand imposed on participants in the experiment were extreme. Clinicians would seldom need to report the locations and vital signs of multiple patients at a time, but instead may focus on the patient with highest acuity or the most concerning change. However, the current study aimed at examining the effect of auditory preview cue in initial stage, which could not be accomplished if the participants are asked to report partial information about abnormal patients only. When the new preview cue has been found to be effective, the participants can be asked to report partial, rather than full, information, which is more typical of clinical use, in the further studies about the preview cue.

Third, more representative clinical tasks were not used in this study. The concurrent arithmetic task used in this study served as a metaphor of administrative tasks in hospital ward. So, the interface of concurrent task was not clinically relevant. Since the participants in the present study were not clinicians, clinical task was not used. It would take longer time for participants to learn how to perform a clinical task if it was adopted in the current study. Also, the jargon in clinical task might distract the participants during the experiment, which could affect the concurrent task performance as an extraneous variable. In future experiments with clinicians as the participants, realistic clinical task should be used to examine the effect of auditory display on concurrent task performance.

Fourth, there might be some undesirable impacts of testing participants through zoom on the study outcomes. For example, participants might be distracted by the notifications on the computer. They might also feel fatigued with 1.5-hour video call. These could affect participant's responses to the tasks. However, following the study protocol, the experimenters would ask the participants to turn off notification alert function before the experiment started. Also, 2-minute breaks were provided to participants between two phases. Then, the impacts of using zoom to test could be minimized. Face-to-face testing is preferred in future experiment, depending on the restrictions under COVID-19.

7. CONCLUSION

In summary, using spearcons or earcons as an auditory preview cue indicating the number of patients with abnormal vital signs in an upcoming sequence does not help participants locate abnormal patients or identify the vital sign levels of abnormal patients. The preview cue tested in this study may make participants aware of the number of patients with abnormal vital signs, indicating the ease or difficulty of reporting the upcoming patient sequence, but it does not increase their accuracy at reporting the details. Moreover, the current design of earcon preview cue suggested that a poorly-designed preview cue may even demand further cognitive processing in order to be understood, causing participants to suspend performance unnecessarily on concurrent tasks, which is the antithesis of what effective auditory displays should do. All in all, neither the spearcon-based nor the earcon-based preview cue of the present design (i.e., indicating number of abnormal patients) improves performance over no cue, so such design can be eliminated as an effective option when designing auditory display for multiple-patient monitoring in future.

APPENDIX A: PAPER MATERIALS

A1. Information Sheet

Lingnan 嶺南大學 University

Department of Applied Psychology

Participant Information Sheet
Multiple-Patient Monitoring Experiment (PEAD03)

The purpose of the study
The purpose of this study is to examine the effectiveness of adding preview cue in patient sequence to monitor multiple patients. This study is being conducted by LAU Tsz Chun Marco under the supervision of Dr. Alan Lee.

Participation and withdrawal
Participation in this study is completely voluntary and you are free to withdraw from this study at any time without prejudice or penalty. If you wish to withdraw, simply stop completing the exercises. If you do withdraw from the study, the materials that you have completed to that point will be deleted and will not be included in the study.

What is involved?
Once you have consented, you will receive some training on the judgements you need to make about the patients. You will be given plenty of practice before the real testing begins. During the test trials, you may be asked to indicate the properties or meanings of auditory and/or visual displays. These displays could be words, pictures, or sound clips. The sounds may represent the status of patients (e.g., heart rate, level of oxygen saturation), Participation in this study is around 1.5 hour.

Risks
Participation in this study should involve no physical or mental discomfort, and no risks beyond those of everyday living. You may feel tired or bored during the experiment. If, however, you should find any question or procedure to be invasive or offensive, you are free to omit answering or participating in that aspect of the study.

Confidentiality and security of data
All data collected in this study will be stored confidentially. Only members of the research team will have access to any identified material such as consent forms. All data will be coded in a de-identified manner and subsequently analysed and reported in such a way that responses cannot be linked to any individual. The data you provide will only be used for the specific research purposes of this study.

Ethics Clearance and Contacts
This study has been cleared in accordance with the ethical review processes of Lingnan University. You are, of course, free to discuss your participation with project staff.

If you would like to learn the outcome of the study in which you are participating, you can contact one of us on the email listed below, and we will send you an Abstract of the study and findings.

Thank you for your participation in this study.
Dr. Alan L. F. Lee & Lau Tsz Chun Marco

A2. Consent Form

Lingnan 嶺南大學 University

Department of Applied Psychology

Participant Informed Consent Form

Project Title: Multiple-Patient Monitoring Experiment (PEAD03)

Your informed consent to participate in this study is needed. Please read the following statements. If you agree with them, please copy/paste the following text into an email and send it to the experimenter at marcolau@ln.hk / chiuwingsum@ln.hk .

==========

With regards to the experiment titled

Multiple-Patient Monitoring Experiment (PEAD03)

I, [insert your full name] on [insert today's date] agree that:

- The nature of this project has been explained to me and I have read and understood the Participant Information Sheet provided.
- I agree to participate in the study as described in the Participant Information Sheet.
- I understand that my participation in this study is voluntary and that I am free to withdraw from the study at any time, without penalty and without needing to provide any reason.
- I understand that any personal data collected at the start of the study will remain confidential and that data collected during the study is de-identified.
- I have been informed that I can contact the researcher if I want feedback on this study.

==========

Researcher Details:
Professor Penelope Sanderson, Professor, School of Psychology, UQ
Dr Simon Y. W. Li, Senior Lecturer, School of Psychological Science, UWA
Professor Robert Loeb, Clinical Professor of Anesthesiology, University of Florida, Gainesville.
Dr. Lee Lap Fai Alan, Department of Applied Psychology, Lingnan University
Marco Lau, MPhil (Psychology) student, Department of Applied Psychology, Lingnan University
Chiu Wing Sum, Research Officer, Department of Applied Psychology, Lingnan University

A3. Pre-testing Questionnaire

PEAD03 測試前問卷

Participant ID:
Condition: Spearcon Cueing / Earcon Cueing / No Cueing

<u>音樂訓練經驗</u>

1. (a) 你是否曾接受一年或以上的正式音樂訓練（例如學樂器或唱歌）？
   是 / 否

   (b) 如是，請問曾接受多少年的正式音樂訓練？_____

2. (a) 你現在是否正接受一年或以上的正式音樂訓練（例如學樂器或唱歌）？
   是 / 否

   (b) 如是，已完成最高等級為 _____

3. (a) 你自己有否進行定期練習（如一星期一次）
   有 / 沒有

<u>聽力</u>

1. 你有沒有正常或已被糾正的聽力？
   有 / 沒有

年齡：_____

性別：男 / 女 / 不願透露

A4. Training answer sheet

Basic Information

PEAD03 Training Answer Sheet (Main Study)

Participant ID

Condition

Spearcon Cueing
Earcon Cueing
No Cueing

Training 2.1

練習二 —— 第一題 （第一次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐

低	☐	☐
好低	☐	☐

練習二 —— 第一題 （第二次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐
低	☐	☐
好低	☐	☐

Training 2.2

練習二 —— 第二題 （第一次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐
低	☐	☐

好低	☐	☐

練習二 —— 第二題 （第二次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐
低	☐	☐
好低	☐	☐

Training 2.3

練習二 —— 第三題 （第一次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐
低	☐	☐
好低	☐	☐

練習二 —— 第三題 （第二次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐
低	☐	☐
好低	☐	☐

Training 2.4

練習二 —— 第四題 （第一次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐
低	☐	☐
好低	☐	☐

練習二 —— 第四題 （第二次作答）

	請選擇水平	
	氧	心
好高	☐	☐
高	☐	☐
正常	☐	☐
低	☐	☐
好低	☐	☐

Training 3.1

練習三 —— 第一題 （第一次作答）

	病人一		病人二		病人三		病人四		病人五	
	氧	心	氧	心	氧	心	氧	心	氧	心
好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

練習三 —— 第一題 （第二次作答）

	病人一		病人二		病人三		病人四		病人五	
	氧	心	氧	心	氧	心	氧	心	氧	心
好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Training 3.2

練習三 —— 第二題 （第一次作答）

	病人一		病人二		病人三		病人四		病人五	
	氧	心	氧	心	氧	心	氧	心	氧	心
好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

練習三 —— 第二題 （第二次作答）

	病人一		病人二		病人三		病人四		病人五	
	氧	心	氧	心	氧	心	氧	心	氧	心
好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Training 4.1

練習四 —— 第一題 （第一次作答）

	病人一		病人二		病人三		病人四		病人五	
	氧	心	氧	心	氧	心	氧	心	氧	心
好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

練習四 —— 第一題 （第二次作答）

	病人一	病人二	病人三	病人四	病人五

	氧	心	氧	心	氧	心	氧	心	氧	心
好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Training 4.2

練習四 —— 第二題 （第一次作答）

	病人一		病人二		病人三		病人四		病人五	
	氧	心	氧	心	氧	心	氧	心	氧	心
好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

練習四 —— 第二題 （第二次作答）

	病人一		病人二		病人三		病人四		病人五	
	氧	心	氧	心	氧	心	氧	心	氧	心

好高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
高	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
正常	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
好低	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

A5. Post-testing Questionnaire

PEAD03 測試後問卷

Participant ID:

Condition: Spearcon Cueing / Earcon Cueing / No Cueing

1. 在整個測試階段，你認為辨認異常病人的位置有多大難度？

極不困難 極困難

1_____2_____3_____4_____5_____6_____7_____8_____9

2. 在整個測試階段，你認為辨認異常病人的位置有多大信心？

極沒信心 極有信心

1_____2_____3_____4_____5_____6_____7_____8_____9

3. 在整個測試階段，你認為辨認異常病人的氧濃度水平和心跳值水平有多大難度？

極不困難 極困難

1_____2_____3_____4_____5_____6_____7_____8_____9

4. 在整個測試階段，你認為辨認異常病人的氧濃度水平和心跳值水平有多大信心？

極沒信心 極有信心

1_____2_____3_____4_____5_____6_____7_____8_____9

5. 在整個測試階段，當多名病人狀態聲音正在播放時，你認為要在心算遊戲中提供正確答案有多大難度？

極不困難 極困難

1_____2_____3_____4_____5_____6_____7_____8_____9

6. 在整個測試階段，當多名病人狀態聲音正在播放時，你認為要在心算遊戲中提供正確答案有多大信心？

極沒信心 極有信心

1_____2_____3_____4_____5_____6_____7_____8_____9

7. 你在測試中用了什麼方法來辨認異常病人的位置？

8. 你在測試中用了什麼方法來辨認異常病人的氧濃度水平和心跳值水平？

For Spearcon Cueing or Earcon Cueing condition:

9. 你覺得異常病人數目的聲音對你辨別異常病人的位置有幫助嗎？請簡短解釋。

10. 你覺得異常病人數目的聲音對你辨別異常病人的氧濃度水平和心跳值水平有幫助嗎？請簡短解釋。

A6. Debrief Sheet

Lingnan 嶺南大學 University

Department of Applied Psychology

Educational Debrief Sheet

Multiple-Patient Monitoring Experiment (PEAD03)

Earcons are short structured sequences of artificial tones that can be used to convey information and combined to provide multiple messages (Brewster et al., 1995). Unlike visual displays, earcons allow eyes-free monitoring. The operator can attend to other tasks while being peripherally aware of the status and trends of vital information. Earcons have been tested and found to be a successful method for communicating complex messages in sound (Brewster et al., 1994).

"**Spearcons**" (speech earcons) is one type of earcons. They contain the name of an alarm (e.g., "pulse high") pronounced in human voice, but in a time-compressed format. Recent studies have shown that spearcons also offer an effective way of monitoring patients as the users could achieve very high accuracy in perceiving the message (Li et al., 2019; Lim and Sanderson, 2019).

As you have experienced, the clinicians need to identify the vital signs of abnormal patients from the patient sequence in multiple-patient monitoring. The patient sequence is the sequence of sounds you heard in the experiment. Li et al. (2019) showed that the accuracy of identifying the vital signs drops when there are more abnormal patients due to the limited memory. But the decrement of accuracy may be lightened if we indicate the number of abnormal patients in the patient sequence. It is because the clinicians can determine whether they focus more on listening to the sequence or not based on the number of abnormal patients provided (Shneiderman, 2003; Jeon et al., 2009). Our study examines the effectiveness of providing that information in the patient sequence on the performance of multiple-patient monitoring.

The number of abnormal patients is called "preview cue". It is an preview of the coming patient sequence. This cue is designed with spearcon and earcon as described in the above. Spearcon may be a better tool for providing the preview because of the benefit of spoken language and intuitiveness. (Lucas, 1984; Thorn et al., 2002). In this study, we also compare the effectiveness of spearcon preview cue and earcon preview cue on multiple-patient monitoring.

This research is an example of human factors psychology. It is only one of the many projects being undertaken by a team consisting of clinicians at the New Territories West cluster of Hospital Authority and researchers from The University of Queensland, University of Western Australia, University of Florida, and Lingnan University. If you would like to learn the outcome of the study in which you are participating, you can contact Dr. Alan Lee at alanlee@ln.edu.hk, and he will send you an Abstract of the study and findings.

Dr. Lee Lap Fai Alan & Lau Tsz Chun Marco
alanlee@ln.edu.hk ; marcolau@ln.hk

References

Brewster, S., Wright, P. C. & Edwards, A. D. N. (1994). A detailed investigation into the effectiveness of earcons, *Auditory Display, 18,*471- 498.

Brewster, S., Wright, P. C. & Edwards, A. D. N. (1995). Parallel Earcons – Reducing the length of audio messages. *International Journal of Human- Computer Studies, 43*(2), 153-175. doi: 10.1006/Ijhc.1995.1039

Jeon, M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2009, September). Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. In *Proceedings of the 1st international conference on automotive user interfaces and interactive vehicular applications* (pp. 91-98).

Li, S. Y., Tse, M. K., Brecknell, B., & Sanderson, P. M. (2019). Spearcon sequences for monitoring multiple patients: Laboratory investigation comparing two auditory display designs. *Human factors*, *61*(2), 288-304.

Lim, H. P., & Sanderson, P. (2019). A comparison of two designs for earcons conveying pulse oximetry information. *Applied ergonomics*, *78*, 110-119.

Lucas, P. A. (1994). An evaluation of the communicative ability of auditory icons and earcons. Georgia Institute of Technology.

Sanderson, P. (2006). The multimodal world of medical monitoring displays. *Applied Ergonomics, 37*(4), 501-512.

Shneiderman, B. (2003). The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization* (pp. 364-371). Morgan Kaufmann.

Thorn, A. S., Gathercole, S. E., & Frankish, C. R. (2002). Language familiarity effects in short-term memory: The role of output delay and long-term knowledge. *The Quarterly Journal of Experimental Psychology: Section A*, *55*(4), 1363-1383.

APPENDIX B: STATISTICAL TABLES AND GRAPHS

B1. Experimenter Effect

Repeated Measures ANOVA (Experimenter Effect on Location Accuracy)

Examine the interaction among number of abnormal patients, preview cue type, and experimenter in affecting the location accuracy.

If there was an experimenter effect, the between subject effect, which was "Experimenter" in the table, would be significant. If there was not, it would be insignificant.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Number of ABN patients	2.199	3	0.733	82.422	<.001	0.501
Number of ABN patients ✻ Preview Cue Type	0.030	6	0.005	0.570	0.754	0.014
Number of ABN patients ✻ Experimenter	0.028	3	0.009	1.051	0.370	0.013
Number of ABN patients ✻ Preview Cue Type ✻ Experimenter	0.025	6	0.004	0.476	0.826	0.011
Residual	2.188	246	0.009			

Note. Type 3 Sums of Squares

[3]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Preview Cue Type	0.050	2	0.025	0.812	0.447	0.019
Experimenter	0.013	1	0.013	0.436	0.511	0.005
Preview Cue Type ✻ Experimenter	0.040	2	0.020	0.647	0.526	0.016
Residual	2.521	82	0.031			

Note. Type 3 Sums of Squares

Repeated Measures ANOVA (Experimenter Effect on Vital Sign Accuracy)

Examine the interaction among number of abnormal patients, preview cue type, and experimenter in affecting the vital sign accuracy.

If there was an experimenter effect, the between subject effect, which was "Experimenter" in the table, would be significant. If there was not, it would be insignificant.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Number of ABN patients	6.205	3	2.068	304.809	<.001	0.788
Number of ABN patients ✻ Preview Cue Type	0.003	6	0.000	0.067	0.999	0.002
Number of ABN patients ✻ Experimenter	0.034	3	0.011	1.691	0.170	0.020
Number of ABN patients ✻ Preview Cue Type ✻ Experimenter	0.035	6	0.006	0.859	0.525	0.021
Residual	1.669	246	0.007			

Note. Type 3 Sums of Squares

[3]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Preview Cue Type	0.011	2	0.005	0.284	0.754	0.007
Experimenter	0.029	1	0.029	1.510	0.223	0.018
Preview Cue Type ✻ Experimenter	0.067	2	0.033	1.763	0.178	0.041
Residual	1.551	82	0.019			

Note. Type 3 Sums of Squares

ANOVA (Experimenter Effect on Arithmetic Attempted Accuracy)

Arithmetic Attempted Accuracy is a proportion accuracy showing how many correct responses were provided out of the total number of attempt during the broadcast of patient sequence in the whole experiment in percentage.

If there was an experimenter effect, the between subject effect, which was "Experimenter" in the table, would be significant. If there was not, it would be insignificant.

ANOVA - Arithmetic Attempted Accuracy

	Sum of Squares	df	Mean Square	F	p	η²p
Preview Cue Type	0.009	2	0.005	0.398	0.673	0.010
Experimenter	0.009	1	0.009	0.790	0.377	0.010
Preview Cue Type ✻ Experimenter	0.009	2	0.005	0.403	0.670	0.010
Residuals	0.955	82	0.012			

ANOVA (Experimenter Effect on Arithmetic Non-Response Rate)

Arithmetic Non-Response Rate is the rate of giving non-responses per minute in concurrent arithmetic task during the broadcast of patient sequence in the whole experiment.

If the participant missed the response, the datafile would record "NoResponse" in User Response column.

If there was an experimenter effect, the between subject effect, which was "Experimenter" in the table, would be significant. If there was not, it would be insignificant.

ANOVA - Arithmetic Non-Response Rate

	Sum of Squares	df	Mean Square	F	p	η²p
Preview Cue Type	13.450	2	6.725	2.742	0.070	0.063
Experimenter	0.446	1	0.446	0.182	0.671	0.002
Preview Cue Type ✻ Experimenter	2.398	2	1.199	0.489	0.615	0.012
Residuals	201.085	82	2.452			

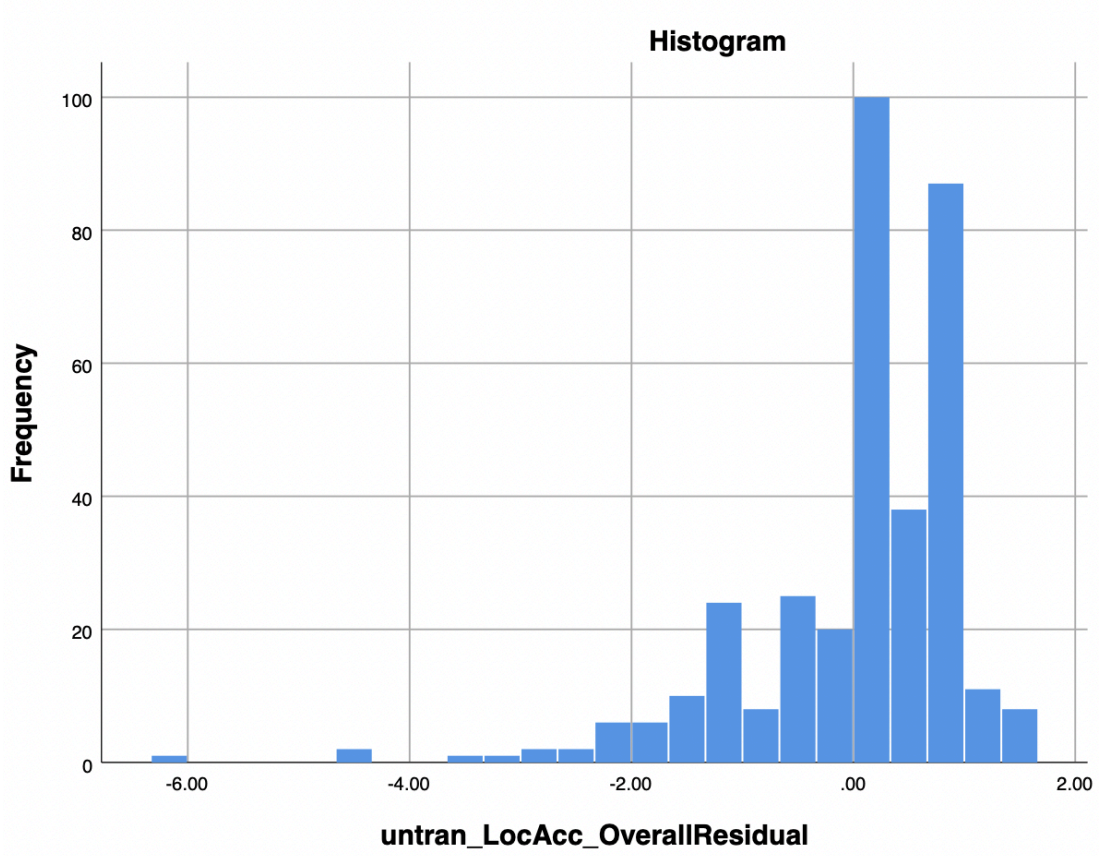
Note. Number of ABN patients = Number of abnormal patients

B2. Residual Distribution Graphs

- The histogram under "Untransformed" section shows the residual distribution of raw data.
- The histogram under "Exponential Transformation" section presents the residual distribution of the data being transformed by exponential transformation.
- The histogram under "Square Root Transformation" section displays the residual distribution of the data being transformed by square root transformation.

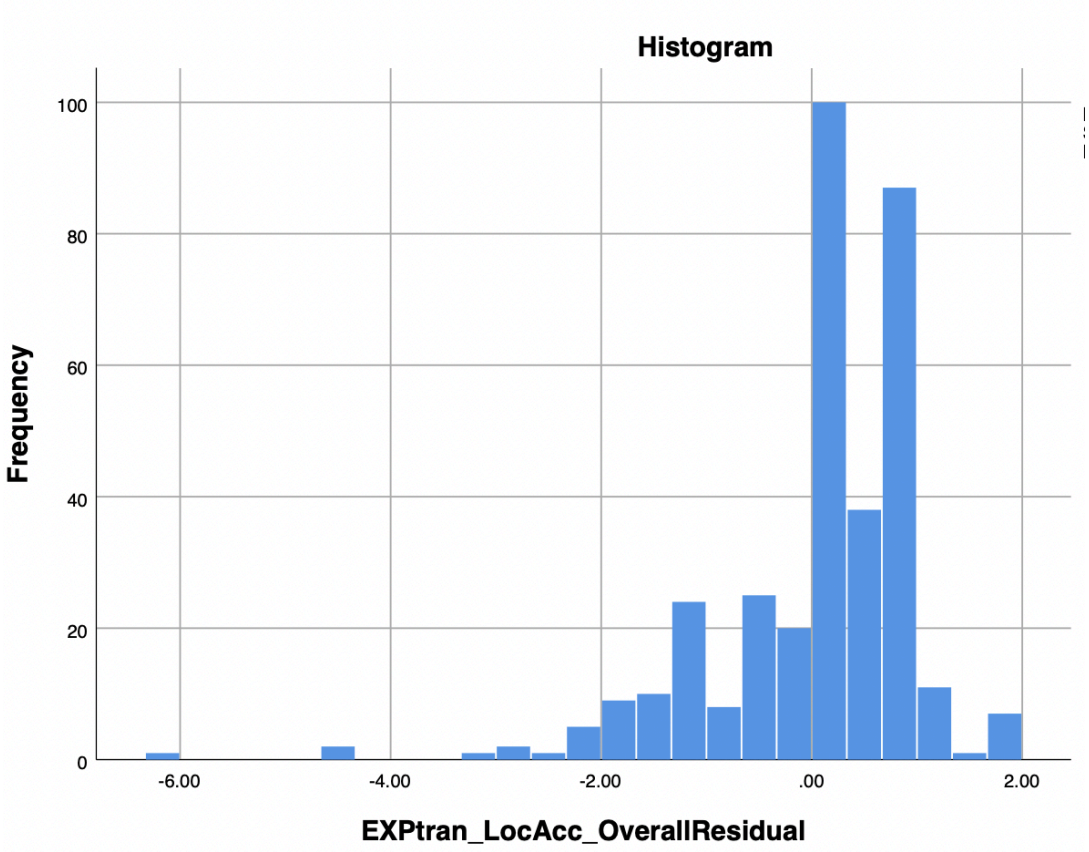
(a) Patient location accuracy

i. Untransformed



As shown in the above histogram, the residuals of raw data for patient location accuracy was negatively skewed. Therefore, exponential transformation was used to transform the raw data in order to get normal distributed data.

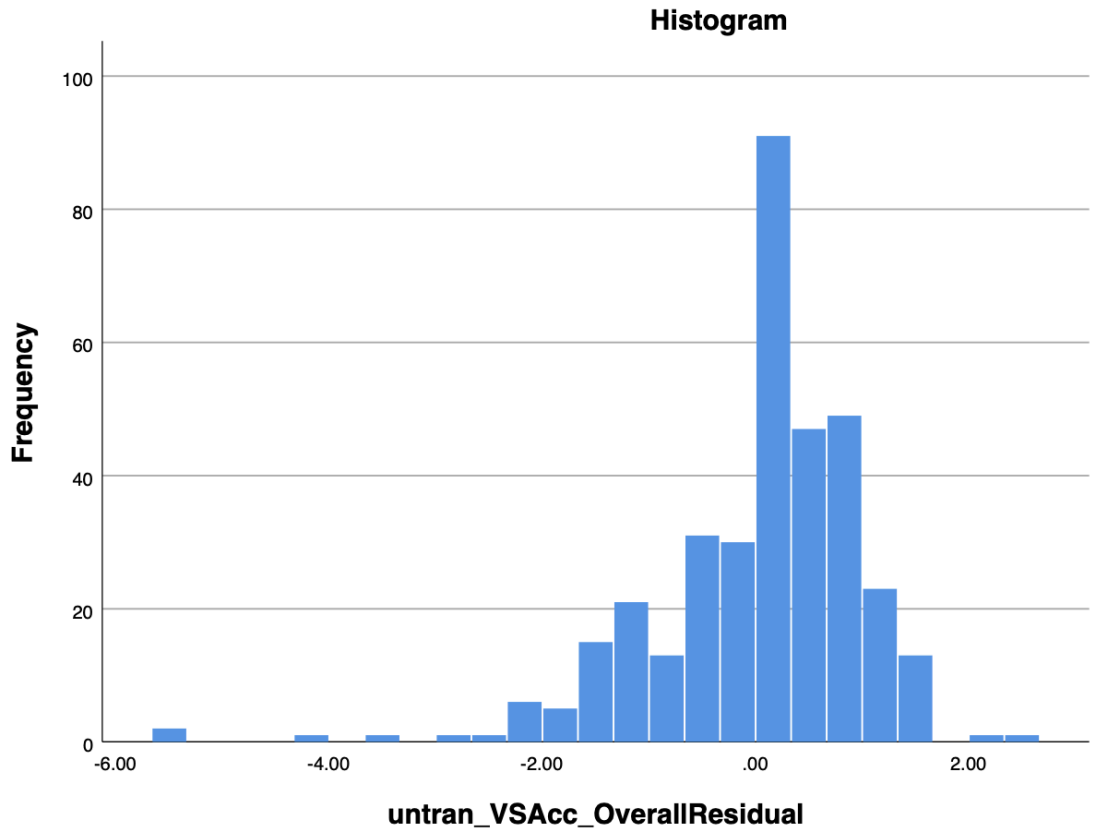
ii. Exponential Transformation



As presented in the above histogram, the residuals for patient location accuracy were still negatively skewed after conducting exponential transformation. Hence, the exponential transformation did not work.

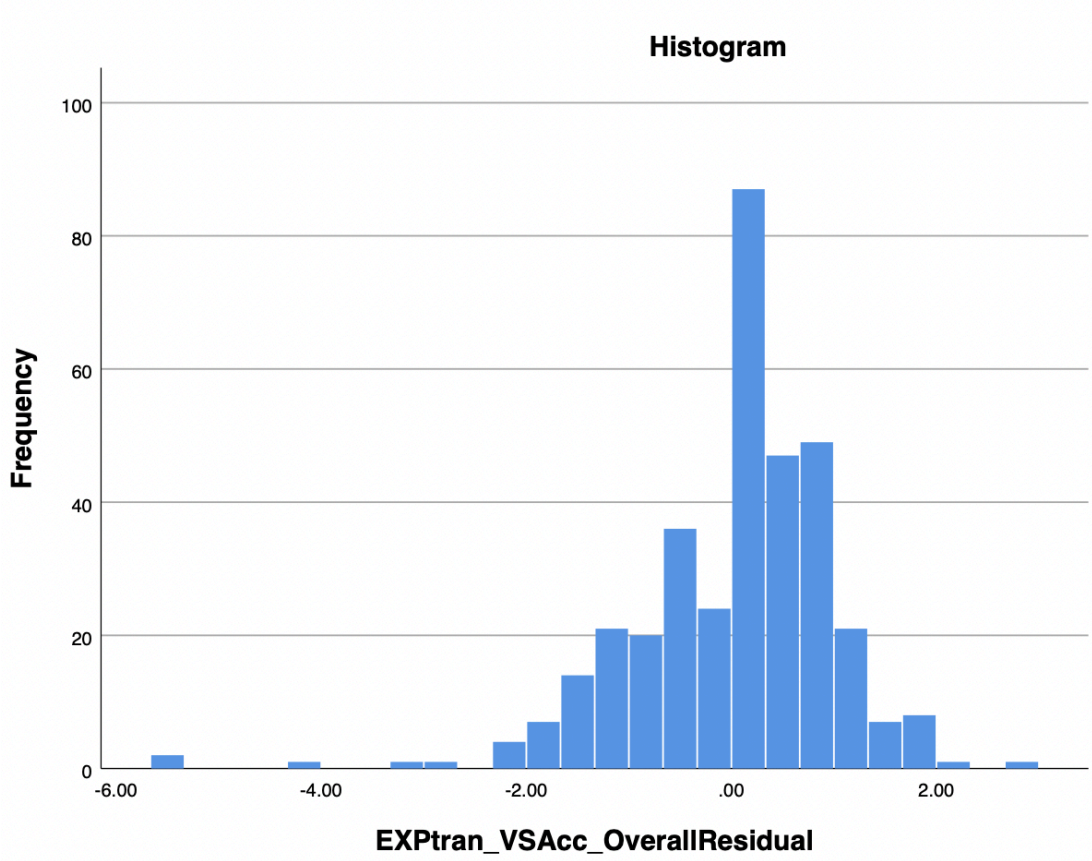
(b) Vital sign accuracy

i. Untransformed



As shown in the above histogram, the residuals of raw data for vital sign accuracy was negatively skewed. Therefore, exponential transformation was used to transform the raw data in order to get normal distributed data.

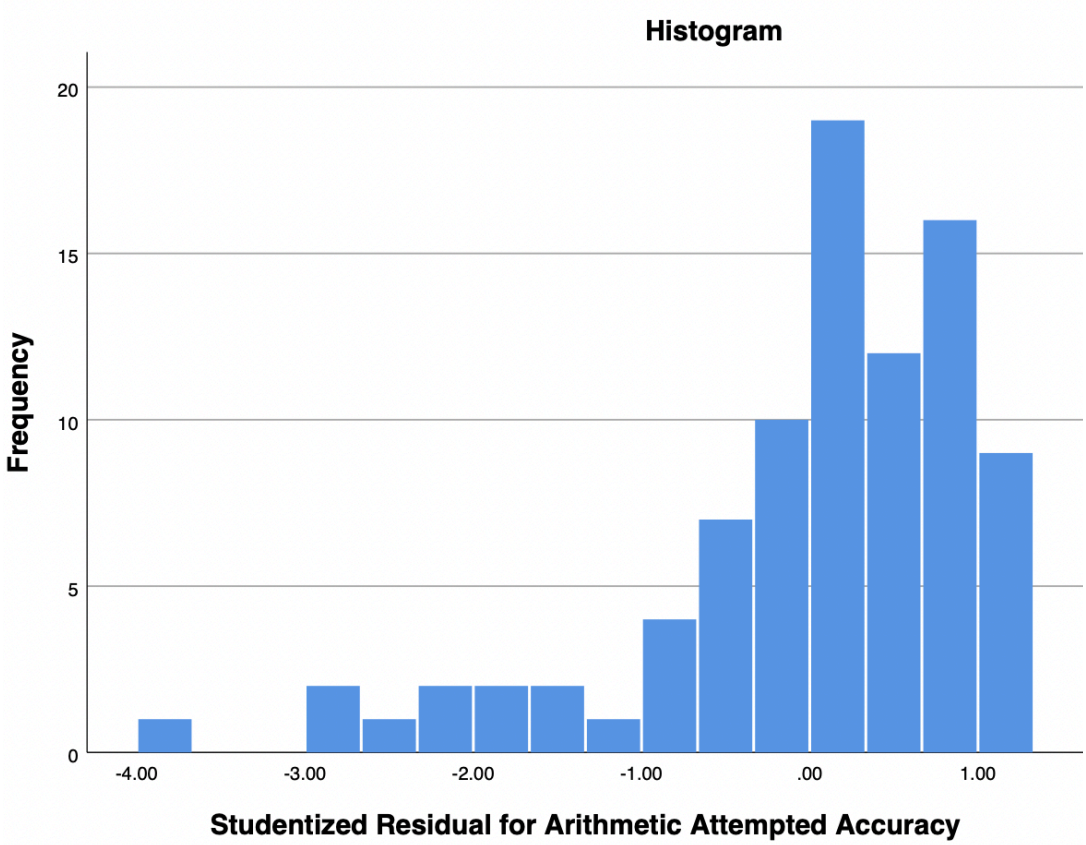
ii. Exponential Transformation



As presented in the above histogram, the residuals for vital sign accuracy were still negatively skewed after conducting exponential transformation. Hence, the exponential transformation did not work.

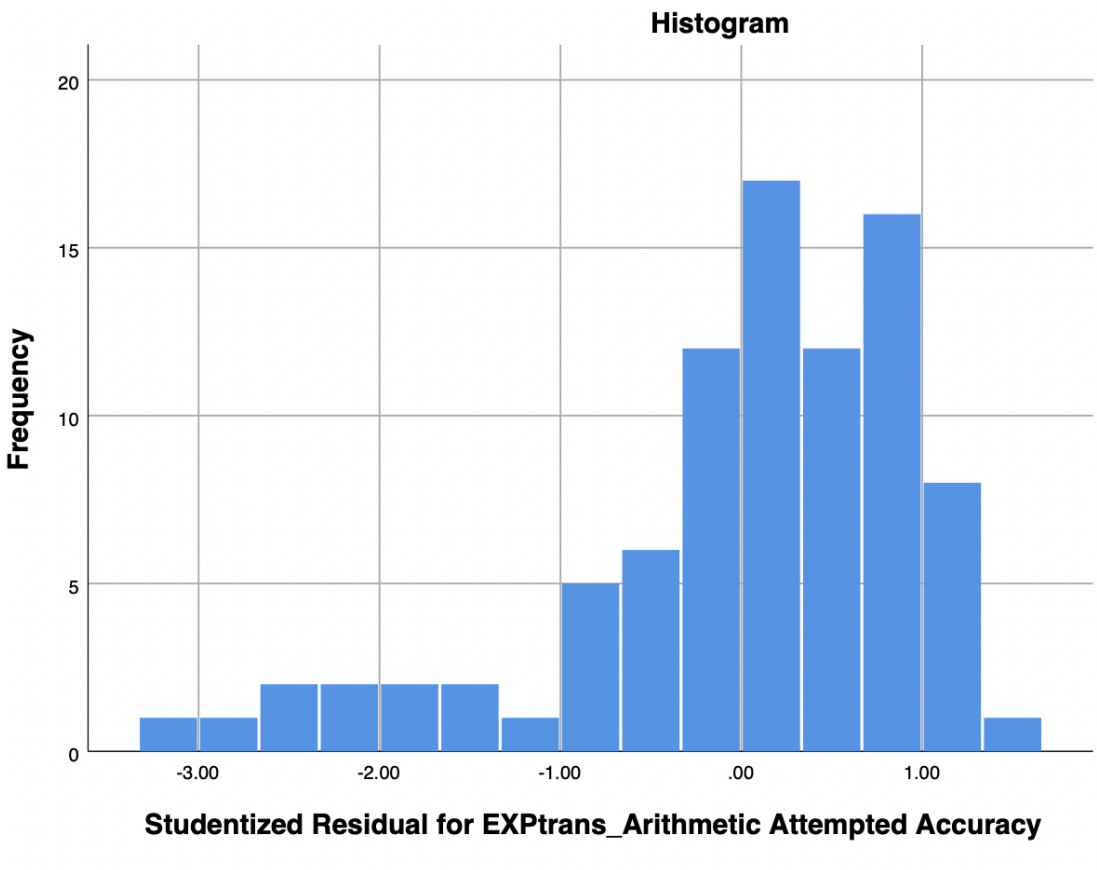
(c) Arithmetic attempted accuracy

i. Untransformed



As shown in the above histogram, the residuals of raw data for arithmetic attempted accuracy was negatively skewed. Therefore, exponential transformation was used to transform the raw data in order to get normal distributed data.

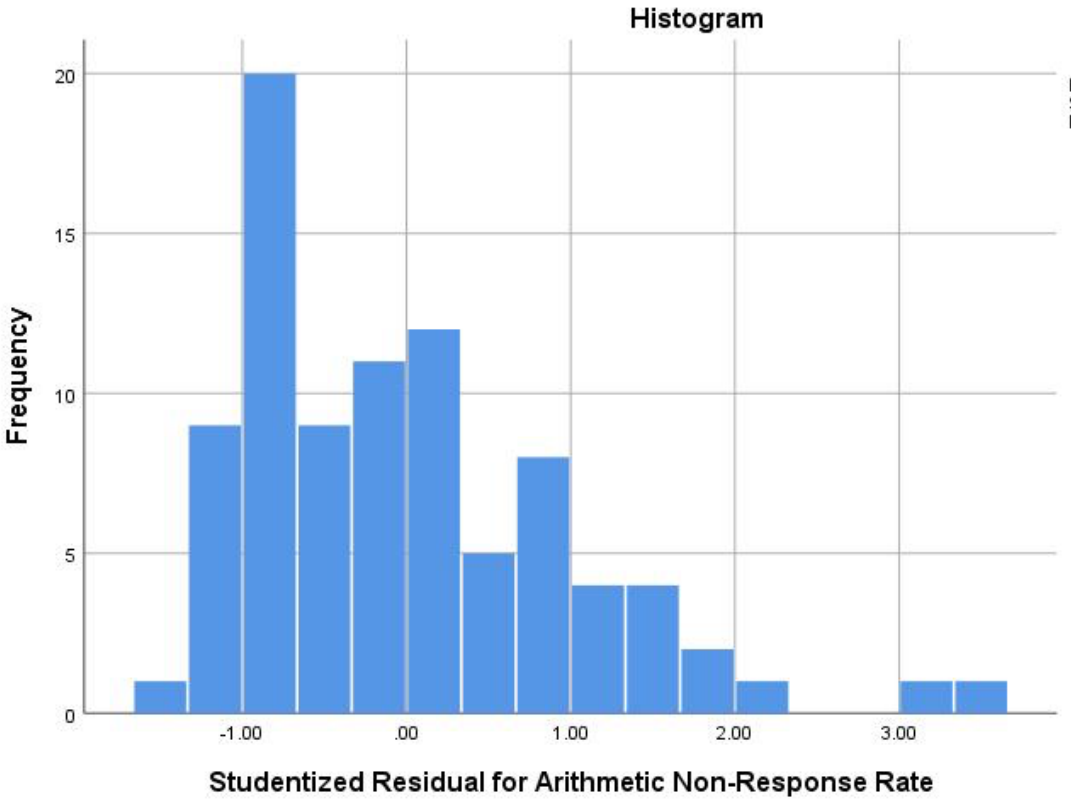
ii. Exponential Transformation



As presented in the above histogram, the residuals for arithmetic attempted accuracy were still negatively skewed after conducting exponential transformation. Hence, the exponential transformation did not work.

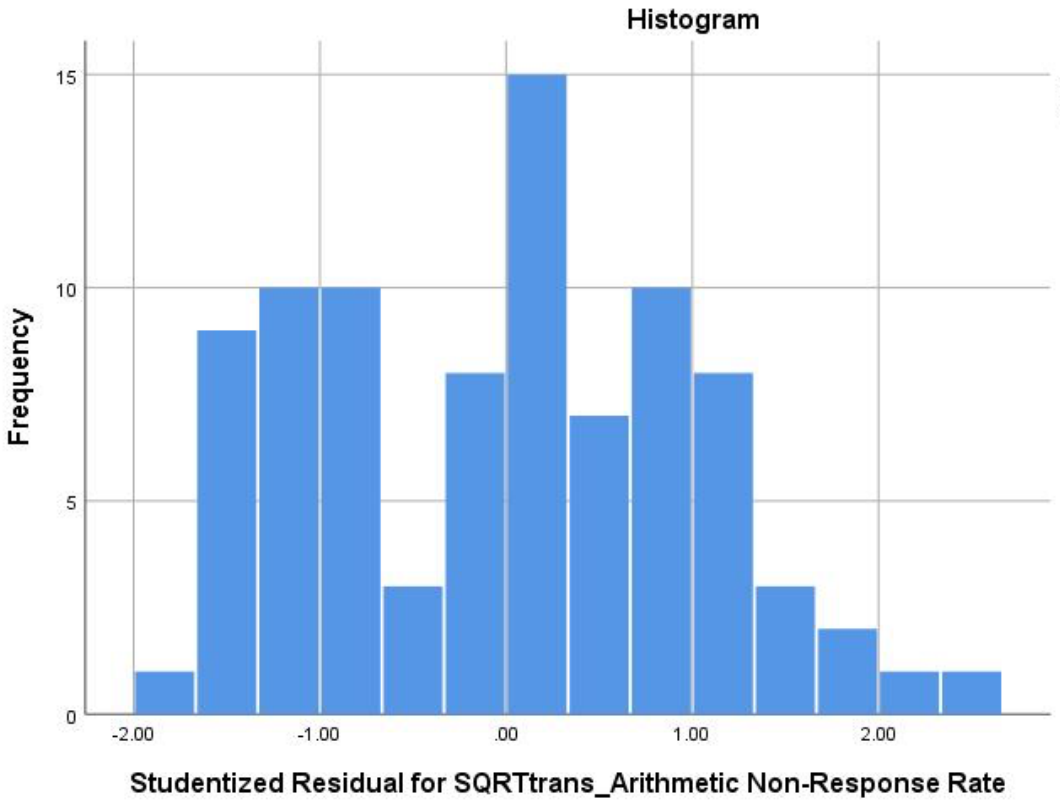
(d) Arithmetic non-response rate

i. Untransformed



As suggested in the above histogram, the residuals of raw data for arithmetic non-response rate were positively skewed. Therefore, square root transformation was used in order to get normal distributed data.

ii. Square Root Transformation



As presented in the above histogram, the residuals for arithmetic non-response rate were still positively skewed after conducting square root transformation. Hence, the square root transformation did not work.

LIST OF REFERENCES

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*(4), 641–657. https://doi.org/10.1037/0096-3445.130.4.641

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, *4*(1), 11–44. https://doi.org/10.1207/s15327051hci0401_1

Brewster, S. A. (1998). Using nonspeech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction*, *5*(3), 224–259. https://doi.org/10.1145/292834.292839

Chen, T., Wu, Y.-S., & Zhu, K. (2018). Investigating different modalities of directional cues for multi-task visual-searching scenario in virtual reality. *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 1–5. https://doi.org/10.1145/3281505.3281516

Cowan, N. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Davidson, T., Ryu, Y., Brecknell, B., Loeb, R. G., & Sanderson, P. (2019). The impact of concurrent linguistic tasks on participants' identification of spearcons. *Applied Ergonomics*, *81*, 16. https://doi.org/10.1016/j.apergo.2019.102895

Dewan, M., Cipriani, L., Boyer, J., Stark, J., Seger, B., & Tegtmeyer, K. (2019). Reducing Redundant Alarms in the Pediatric ICU. *Multimodal Technologies and Interaction*, *3*(1), 11. https://doi.org/10.3390/mti3010011

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

FÜrst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, *28*(5), 774–782. https://doi.org/10.3758/BF03198412

Greene, S., Marchionini, G., Plaisant, C., & Shneiderman, B. (2000). Previews and Overviews in Digital Libraries: Designing Surrogates to Support Visual Information Seeking. *Journal of the American Society for Information Science*, *51*(4), 380–393. https://doi.org/10.1002/(SICI)1097-4571(2000)51:4<380::AID-ASI7>3.0.CO;2-5

Hasani, S., Yan, N., & Li, C. (2018). TableView: A Visual Interface for Generating Preview Tables of Entity Graphs. *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 1617–1620. https://doi.org/10.1109/ICDE.2018.00190

Hendrich, A. (2008). A 36-Hospital Time and Motion Study: How Do Medical-Surgical Nurses Spend Their Time? *The Permanente Journal*, 25–34. https://doi.org/10.7812/TPP/08-021

Hickling, A., Brecknell, B., Loeb, R. G., & Sanderson, P. (2017). Using a Sequence of Earcons to Monitor Multiple Simulated Patients. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(2), 268–288. https://doi.org/10.1177/0018720816670986

Hornbæk, K., & Hertzum, M. (2011). The notion of overview in information visualization. *International Journal of Human-Computer Studies*, *69*(7–8), 509–525. https://doi.org/10.1016/j.ijhcs.2011.02.007

Lakens, D., Scheel, A., & Isager, P. (2018). *Equivalence Testing for Psychological Research: A Tutorial*. https://doi.org/10.17605/OSF.IO/QAMC6

Li, S. Y. W., Tang, T.-L., Hickling, A., Yau, S., Brecknell, B., & Sanderson, P. M. (2017). Spearcons for Patient Monitoring: Laboratory Investigation Comparing Earcons and Spearcons. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(5), 765–781. https://doi.org/10.1177/0018720817697536

Li, S. Y. W., Tse, M.-K., Brecknell, B., & Sanderson, P. M. (2019a). Spearcon Sequences for Monitoring Multiple Patients: Laboratory Investigation Comparing Two Auditory Display Designs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(2), 288–304. https://doi.org/10.1177/0018720818797502

Li, S. Y. W., Yeung, C.-W., Davidson, T., Ryu, Y., Srbinovska, M., Salisbury, I., Loeb, R. G., & Sanderson, P. M. (2019b). Spearcons for Patient Monitoring: Program of Laboratory-Based Feasibility Studies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 663–667. https://doi.org/10.1177/1071181319631258

Lim, H.-P., & Sanderson, P. (2019). A comparison of two designs for earcons conveying pulse oximetry information. *Applied Ergonomics*, *78*, 110–119. https://doi.org/10.1016/j.apergo.2019.01.013

Lin, L., Cockerham, D., Chang, Z., & Natividad, G. (2016). Task Speed and Accuracy Decrease When Multitasking. *Technology, Knowledge and Learning*, *21*(3), 307–323. https://doi.org/10.1007/s10758-015-9266-4

McHugh, M. D., Aiken, L. H., Sloane, D. M., Windsor, C., Douglas, C., & Yates, P. (2021). Effects of nurse-to-patient ratio legislation on nurse staffing and patient mortality, readmissions, and length of stay: A prospective study in a panel of hospitals. *The Lancet*, *397*(10288), 1905–1913. https://doi.org/10.1016/S0140-6736(21)00768-6

Qvarfordt, P., Golovchinsky, G., Dunnigan, T., & Agapie, E. (2013). Looking ahead: Query preview in exploratory search. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 243–252. https://doi.org/10.1145/2484028.2484084

Reissland, J., & Manzey, D. (2016). Serial or overlapping processing in multitasking as individual preference: Effects of stimulus preview on task switching and concurrent dual-task performance. *Acta Psychologica*, *168*, 27–40. https://doi.org/10.1016/j.actpsy.2016.04.010

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*(2), 261–269. https://doi.org/10.3758/BF03208993

Röer, J. P., Körner, U., Buchner, A., & Bell, R. (2017). Attentional capture by taboo words: A functional view of auditory distraction. *Emotion*, *17*(4), 740–750. https://doi.org/10.1037/emo0000274

Ruskin, K. J., & Hueske-Kraus, D. (2015). Alarm fatigue: Impacts on patient safety. *Current Opinion in Anaesthesiology*, *28*(6), 685–690. https://doi.org/10.1097/ACO.0000000000000260

Sanderson, P. M., Brecknell, B., Leong, S., Klueber, S., Wolf, E., Hickling, A., Tang, T.-L., Bell, E., Li, S. Y. W., & Loeb, R. G. (2019). Monitoring vital signs with time-compressed speech. *Journal of Experimental Psychology: Applied*, *25*(4), 647–673. https://doi.org/10.1037/xap0000217

Schraefel, M. C., Karam, M., & Zhao, S. (2003). Listen to the Music: Audio Preview Cues for Exploration of Online Music. *Proceedings of Interact 2003-Bringing the Bits Together*.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–343. https://doi.org/10.1109/VL.1996.545307

Singleton, W. T. (1971). The ergonomics of information presentation. *Applied Ergonomics*, *2*(4), 213–220. https://doi.org/10.1016/0003-6870(71)90108-6

Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2013). Spearcons (Speech-Based Earcons) Improve Navigation Performance in Advanced Auditory Menus. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(1), 157–182. https://doi.org/10.1177/0018720812450587

Watson, D. M., & Gill, M. T. (2004). Earcon for Intermittent Information in Monitoring Environments. *Proceedings of the 2004 Annual Conference of the Computer-Human*

Interaction Special Interest Group of the Human Factors and Ergonomics Society of Australia (OzCHI 2004), 4. CD-ROM.