Perception of Mobile Emergency Alert Sounds

Olamide Olatoye
University of Waterloo
Waterloo, Canada
oolatoye@uwaterloo.ca

Clara Lau
University of Waterloo
Waterloo, Canada
cswlau@uwaterloo.ca

Neuel Sam
University of Waterloo
Waterloo, Canada
nsam@uwaterloo.ca

ABSTRACT

Canada launched Alert Ready, an emergency alert system that directly communicates emergency warnings and information to citizens via their personal mobile devices in 2015 but has received a large volume of criticism. Some of the major criticism from the citizens as found by a nationwide survey performed on November 27, 2019, were the need to improve sound for the alerts. A lab study was conducted to investigate how participants perceived and interacted with simulated mobile alert messages with three different sounds while engaged in a gaming task. Participants were particularly observed for reactions and dismissal times during the study. They were provided a post-task questionnaire to report their perceived annoyance, perceived urgency, and positive and negative affect after the task. Results found significant differences in dismissal times, perceived urgency and positive and negative affect between the three sounds. There was no significant difference observed for perceived annoyance. With these results, the authors discuss various ideas for future studies on mobile emergency alert sound design.

Author Keywords

Mobile Emergency Alerts; Warning Sound Design; Auditory Perception; Behavioural Measures; Lab Study

CSS Concepts

•**Human-centered computing**
~Ubiquitous and mobile computing
~Empirical studies in ubiquitous and mobile computing

INTRODUCTION

In 2015, Canada launched Alert Ready, an emergency alert system that directly communicates emergency warnings (e.g. Amber alerts, environmental hazards, etc.) and information to citizens via their personal mobile devices with LTE connection [1]. Unlike radios and television sets, which are traditionally used to broadcast emergency alerts, mobile devices such as cell phones and tablets are rarely ever turned off and a 2013 poll found that approximately 42% of Canadians have their devices on hand 24/7 [2]. The introduction of mobile emergency alerts has changed the way people interact with emergency messages. Unfortunately, even though this change has the potential to increase public safety and the effectiveness of early emergency warnings, it has also received a large volume of criticism.

The current Alert Ready mobile alerts already have a harsh reception from the general public for various reasons such as causing unnecessary disruptions and having unclear messages. This has prompted citizens to file many disruptive complaints directly to the police and complain on social media [3]. To dissuade this behaviour and prevent citizens from dismissing the importance of future emergency messages, it is important to understand why the general public is upset with the current design. A nationwide survey performed on an emergency test alert sent out on November 27, 2019, found that two of the main things that respondents wished to change about the alerts were to "Improve the sound played for the Test Alert" and "Have different sounds for different types of Alerts" [4]. However, it is still unknown exactly how to change the sound of these alerts.

To address this, a lab study was performed on how participants perceived and interacted with simulated mobile alert messages with three different sounds while engaged in a gaming task. Particularly, alert dismissal times were measured, and participants were provided a post-task questionnaire to report their perceived annoyance, perceived urgency, and positive and negative affect immediately after the task. The authors found significant differences in dismissal times, perceived urgency and positive and negative affect between the three sounds. No significant difference was observed for perceived annoyance. Based on these results, the authors discuss the implications of using text-to-speech and perceptually neutral sounds for emergency alarms, and various ideas for future studies on mobile emergency alert tone design.

RELATED WORK

The body of literature pertaining to mobile emergency alerts is still growing. However, much of the design for these mobile alerts are based on older works on disaster response evaluations and alert design [5]. The following sections will discuss some of the related literature motivating the study performed in this paper.

Existing studies on government-issued mobile emergency alerts

To begin with, Canada is not the only country with a mobile emergency alert system. The USA has the Wireless Emergency Alerting (WEA) system and Australia has the Emergency Alert System. These countries' governments have executed their own surveys, systematic reviews and assessments of their respective mobile alert systems [4, 6-8]. These reports are created based on large data samples representative of the general public and provide insight into whether alerts were successfully received and interpreted. This research provides generalizable results to understand how the system is working as a whole but due to the limitations of using generic surveys, they lack specific details on people's experiences and how to improve the alerts. They also focus more on assessing the infrastructure stability and wording of messages rather than the sound used for the alerts. Last, this research does not directly observe how people interact with mobile alerts as they are occurring and rather probe citizens for their input at a later time (for example the next day in [4]).

Fortunately, there are research publications in HCI, and usability focused on assessing the interaction between the general public and emergency alert systems. [9] is a notable study titled "*Rethinking the Future of Wireless Emergency Alerts: A Comprehensive Study of Technical and Conceptual Improvements*" published in 2018. In this work, Kumar et al. worked with alert creators and performed two public usability studies spanning one year to evaluate various extensions to the capabilities of the WEA system. They created a testbed application to simulate wireless alerts [9] and found that the actionability and accuracy of interpretation of wireless alerts could be improved by implementing precise geo-targeting, location maps, and digestible message formats. The findings of this study are similar to those found in a usability study performed on the Commercial Mobile Alert System (CMAS) where the authors created a prototype application supporting an enhanced message structure and allowing users to filter through their alert messages [10]. These publications provide rich information on the interaction aspect of these mobile alerts and how public feedback can inform the design of mobile emergency alerts. However, they still do not focus on the actual sound of the alert.

Alarm design outside of Mobile Emergency Alerts

To better understand the effects of sound on human perception and behaviour, it is instead beneficial to look at the large body of existing human factors research on alarms. The purpose of alarms is to alert and evoke urgency or actions in receivers in the immediate future and an individual's experience with alarms is influenced by a variety of psychoacoustic factors. Various publications discussed in [11, 12] provide information on the general best practices for alarm design and the psychoacoustic properties of different sounds. For example, a publication by Patterson in 1990 presents guidelines for alarm design to improve their ergonomics [13]. These guidelines establish the possibility of achieving different urgency mappings by designing distinct alarms [13] resonating with the general public's current suggestions to use different sounds for the alerts.

Meanwhile, experiments such as [14-16] evaluate how acoustic qualities (e.g. pitch, timbre, resonance) are perceived and affect alarm performance and learnability. This includes studies that assess alarms for specific occupations such as medical operation room monitor alarms [17,18] and warning tones for workers backing up vehicles [19]. In the space of alarms used by the general public, there are various works examining the notification tones used in consumer mobile applications [20] and alarms for waking up [21] also presents an interesting clinical research study on fire alarms finding that children were more responsive to personalized fire alarms recorded with their parents' voices as opposed to a conventional emergency tone [22]. The above experiments evaluate sounds by measuring sound pressure levels, alarm response times, task completion times before and after alarms and subjective ratings of perceived urgency and annoyance. In this paper, we specifically study tones used for mobile emergency alert messages in a controlled simulation recording similar metrics and forming our hypotheses on the generalizable findings from this existing literature.

EXPERIMENT DESIGN

To observe how people interact with mobile alerts with different sounds, a within-participants design lab study was performed. Participant recruitment and study execution was performed in compliance with the course-approved research ethics outlined for the MSCI 630 course offered at the University of Waterloo.

Participants

14 participants (11 Males, 3 Females) aged between 20 and 55 were recruited from a local university in abidance by research ethics. No compensation was provided for participation.

Pilot Study Participants

Out of the 4 participants (3 Males, 1 Female) aged between 24 and 55, 3 used alarms often and 3 had pop-up notifications activated on their phone.

Research Data Participants

Out of the 10 participants (8 Males, 2 Females) aged between 20 and 40, 9 used alarms often and all of them had pop-up notifications activated on their phone. All 10 participants reported normal hearing abilities.

Method

Sound selection

Three sounds were selected based on the assumption that extremely disruptive emergency alert sounds would be perceived negatively and result in behaviours such as dismissing alerts without reading them. The first two sounds chosen for the study simulate auditory tones representing extremes from "pleasant" to "disturbing" as perceived by humans. The sounds were selected by picking multiple sounds that the researchers perceived as "pleasant" and "disturbing" from http://www.ZapSplat.com, a free online library of sound effects. Three third-party acquaintances were also asked to confirm that there was a large audible difference or contrast between the sounds.

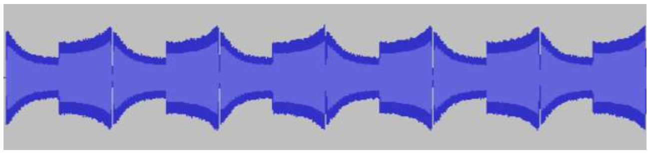

Figure 1: Siren Alert

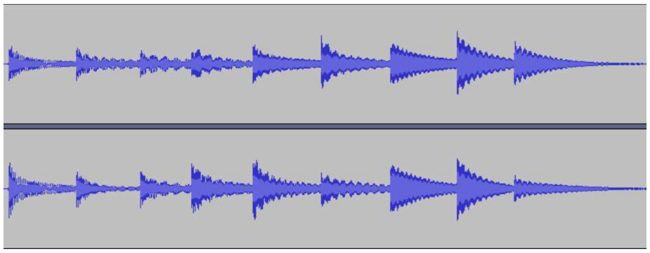

Figure 2: Chime Alert

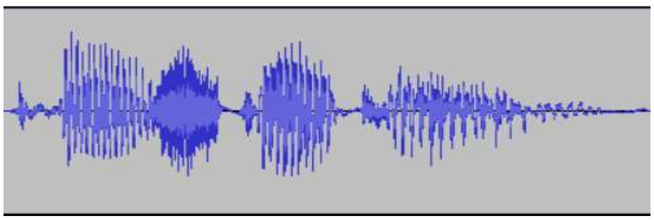

Figure 3: Text-to-Speech Alert

Text-to-speech alarms have also received increasing attention with advances in technology [23] and thus a third text-to-speech alarm was also included in the study. The "Siren"[1] sound (Figure 1) was selected to resemble the existing Canadian Alerting Attention Signal. It is a repeating fast pitched siren alarm that gives the notion of urgency. The "Chime"[2] sound (Figure 2) is a repeating tone resembling alarm tones used to calmly wake people up in the morning. The "Text-to-speech" sound (Figure 3) is a male voice derived from the NaturalSoft (https://www.naturalreaders.com/online/) online text-to-speech engine and repeats "This is an emergency alert. Please stay calm."

Alert setup

An android personal alarm application called AMdroid Alarm Clock[3] app was used to simulate the alerts. The application allows users to set alarms with personalized ringtones and background images. The authors originally considered using the testbed WEA simulation application provided in [9]. However, it was determined that an alarm application could provide the necessary functionality with a less complex setup.

During the studies, four alarms were configured in the AMdroid app (one test alarm with an unrelated sound to show participants how to dismiss the alarms and the three experiment sounds). The alarms were set at fixed uneven intervals (3 minutes after the beginning of the study, 1 minute after the 1st alarm and 2 minutes after the second alarm) for all participants. To reduce ordering effects, the alarms were counterbalanced following the six possible combinations shown in Table 1 on the next page.

On screen, participants were shown a red "Dismiss" button in the bottom right corner and three emergency alert messages that appeared in the last five years in Ontario [24-26]. Each alert sound had a different alert message and the alerts were displayed with white backgrounds in order to increase the focus on the text displayed.

Task Design

Based on the objective of making participants engaged in given task, an online version of the *Zuma's Revenge* game by PopCap Games was chosen (Figure 4). The objective of *Zuma*'s *Revenge* is to eliminate all of the balls rolling around the screen along a given path before they reach a skull at the end of the path. To eliminate the coloured balls, the player can fire a coloured ball from the stone frog idol's mouth towards the chain of balls. When three or more balls of the same colour collide, they explode, possibly triggering other explosions resulting in a combo. The level is completed when the player eliminates all the balls on the screen [27].

Figure 4: Zuma Game running on a PC[4]

[1] ZapSplat. *Science fiction alarm or siren with a slightly harsh tone*. Retrieved March 26, 2020 from https://www.zapsplat.com/?s=Science+fiction+alarm+or+siren+with+a+slightly+harsh+tone&post_type=music&sound-effect-category-id=

[2] ZapSplat. *Game sound, climbing mallet, ascend 5.* Retrieved March 26, 2020 from https://www.zapsplat.com/music/game-sound-climbing-mallet-ascend-5/

[3] AMdroid Alarm Clock App. *Alarm Clock for Heavy Sleepers — Loud + Smart Math* Retrieved March 26, 2020 from https://play.google.com/store/apps/details?id=com.amdroidalarmclock.amdroid&hl=en .

[4] MSN Games, *Zuma's Revenge!* Retrieved March 26, 2020 from https://zone.msn.com/gameplayer/gameplayer.aspx?game=zumasrevenge&instance=default .

In 3 minutes	1 minute after 1st	2 minutes after the 2nd
Chime	Siren	Text-to-speech
Siren	Text-to-speech	Chime
Text-to-speech	Chime	Siren
Chime	Text-to-speech	Siren
Siren	Chime	Text-to-speech
Text-to-speech	Siren	Chime

Table 1: Alert Sound Order for Experiments

Control decisions

The pause button in the game is not emphasized graphically and a conscious decision to not tell the participants about the pause button was taken in order to make the experiment as close to real-life as possible. In this case, people doing everyday tasks would not be interested in pausing their activity in order to respond to an alert and if they want to, they will check if a pause button exists.

The smartphone was kept on the side where the mouse of the computer was kept so as to minimize quick dismissal using the non-dominant hand. This was only possible for the first alert and participants kept the phone on either side of the computer after handling the phone when the first alert went off. The participants were asked to treat the phone as their own. Due to concerns over privacy and experiment setup, participant phones could not be used to run the experiment.

To reduce ordering effects, the alarms were counterbalanced following the six combinations shown in Table 1.

Dismissal Times

Dismissal times from when the alarm went off until the participant actively stopped it were recorded for each sound. These times were recorded by one researcher who reviewed all of the videos. The authors were interested in observing how much attention participants placed on the mobile phone and whether participants paid attention to the onscreen messages as opposed to dismissing the alarm as soon as possible. One concern with the existing emergency alert tone is that it may over irritate the receiver and shifts their attention away from the emergency message towards trying to stop the sound. Stopping the sound consequently dismisses the message and therefore may result in emergency information being only partially received or ignored. Thus, in this study, significantly longer dismissal times accompanied by observations of whether the participant looked at the onscreen message were interpreted as positive behaviour signifying longer exposure to the emergency message increasing the chances of proper communication.

Post-task questionnaire

A 3-page post task questionnaire (see Appendix A) was given to participants with one page for each of the three alert sounds. The purpose of the questionnaire was to collect self-reported information reflecting the participant's perceptions of each sound and later evaluate the author's hypotheses.

The questionnaire begins with two open ended questions on the alert sounds they heard. Next, participants were also asked to fill out an opinion scale on the score 0-10 based on the level of urgency and annoyance they felt for the specific sound. These ratings were asked for using the ICBEN's recommended standardized general-purpose noise reaction questions [28] format but reworded slightly to match the context of the study. These metrics were selected as the purpose of an alarm is to not only alert the receiver but also communicate a specific level of urgency associated with its message. Assessing perceived annoyance was also expected to provide insight on how perception of the alert sounds could be "improved" to address the public's negative feedback. Similar measures have been used in other studies [15, 16] to evaluate alarm designs.

Finally, to assess the emotional affect of each alert sound, participants were also asked to fill out the PANAS-SF [29], a 10-item short-form version of the PANAS (Personal Affect Negative Affect Schedule) used to rank the positive affect and negative affect of events. This scale is typically used for mood assessments however, participants were asked to fill out the scale based on their experience with the three different alert sounds. The PANAS was selected over a custom satisfaction or emotion questionnaire as it is a validated scale. Specifically, the short-form was selected over the 20-item original version in attempt to reduce questionnaire fatigue on the participant.

Procedure

Pilot Studies were run in a noisy open environment on a university campus to simulate real-life environments. The study was run on four participants to evaluate the experiment design and identify changes that may affect the data derived. Based on the pilot study and the participant feedback from the study, the following changes were made to the experimental design:

- Change of setting from a noisy environment to an isolated meeting room for better control over what participants hear. In the noisy environment, participants misrecognized the alert sounds from the phone as noises in the background.
- Change of protocol/script to explicitly tell participants to treat the phone as their own. This was done as participants

ignored alerts on the phone thinking it was not necessary to interact with the phone as it is not their own. Also, this made sure participants would not pick up the phone and ask the researchers for help.

- Revision to some of the questionnaire questions which included changes to wordings and clarifications on the experiment.
- The participants were explicitly asked to reveal if they had hearing disabilities based on the pilot study conducted to ensure that the data recorded closely resembled real-life cases. The vibrator (haptic feedback) was also turned on to ensure that the participants noticed the alert went off and to include the accessibility features.

The revised study was conducted in a small conference room that is sized for 6 people. Each participant was welcomed and asked to sit in the area set up for the study. The equipment set-up for the study was a comfortable chair, table, laptop computer, mouse, Android smartphone (Samsung S10) and a camera on a tripod stand. The experiment was reintroduced to participants and tasks required of them were explained. The facilitator explained that the team is investigating how people interact with alerts on their mobile phones and the study was going to take approximately 15 minutes. The facilitator further explained that they would be asked to play a computer game on the laptop and given questionnaires to fill afterward while answering a few non-structured interview questions. Participants were duly informed of the video recordings by the camera which would not show their face, the anonymity and safekeeping of their data and their right to leave whenever they cannot proceed with the study. Participants were instructed to turn off their phones or put on silent/airplane modes. The facilitator also asked and checked for any hearing impairments before proceeding further with the study.

Each participant was asked to fill both a demographic questionnaire and a consent form. The demographic questionnaire assessed data like age, gender, daily mobile phone, and alert usage. The instructions for the game were also thoroughly explained to participants as they were told to proceed as far as possible in the game within the 7 minutes of the study. Participants were also told that while playing the game, alerts may go off on the mobile phone placed right on the table beside the laptop. Facilitators clearly provided a disclaimer on alerts not representing a real emergency and that if there be any actual emergency, participants would be informed immediately. Participants were asked to treat the phone as their own and act accordingly when interacting with it. They were shown how to dismiss alerts on the phone if needed and finally given a few minutes to familiarise themselves with the game. While participants were getting familiar with the game, the facilitator with the assistance of other team members set the alert times.

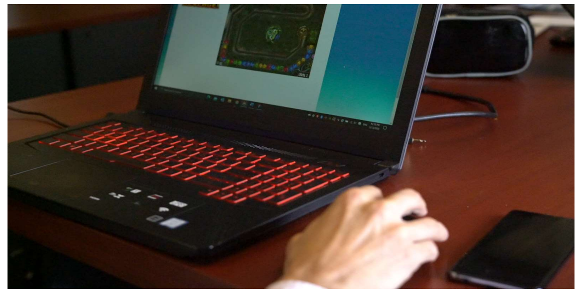

Figure 5: Participant playing Zuma with the phone placed next to the mouse

The experiment began the moment the mobile phone was returned to next to the participants already interacting with the game (Figure 5).

The participants were then left to play the game while the emergency alerts went off. The camera recorded full interaction for each participant for further observation data and assessing response times. After 7 minutes elapsed and with all three alerts gone off, participants were debriefed and asked if they remembered all sounds distinctly, how each made them feel and if the game was engaging. A few other non-structured interview questions were asked for observation and control purposes. The facilitator then handed out the post-task questionnaires for each sound. Once participants completed the questionnaires and confirmed they had understood the questions, the researchers expressed their thanks and escorted the participants out of the conference room.

HYPOTHESES

The review of the literature and previous studies done regarding alert sounds have informed the following hypotheses tests to be drawn from the study:

H1: the siren sound has shorter dismissal times compared to the other two sounds.

H2: the siren sound provokes higher perceived annoyance compared to the other two sounds

H3: the siren sound provokes higher perceived urgency compared to the other two sound

H4: the siren sound has a higher negative affect on the PANAS than the other two sounds.

H5: the chime sound has lower perceived urgency compared to the other two sounds.

H6: the text-to-speech sound has longer dismissal times compared to the other two sounds.

Based on the current public discontent towards the mobile emergency alerts and the alert tone, we propose H1, H2, H4. We expected the siren sound, similar to the current alert tone, to be perceived negatively, be perceived as annoying, and be stopped more quickly.

H3 and H5 are proposed in research suggesting that sounds that are faster, higher pitch and more unpredictable tend to be perceived with higher urgency [12]. The siren sound chosen is the fastest and highest pitch out of the three sounds and thus it was expected to have significantly higher urgency than the other two sounds. Meanwhile, the chime sound is slower and more abstract than the other two sounds and thus it was expected to have a lower perceived urgency.

Last, existing research on text-to-speech alerts suggests that they require longer response times as more time is needed to listen to the message [24]. Thus, we expected participants to respond and in extension, dismiss the text-to-speech alert more slowly.

RESULTS

Data Entry

After the studies were completed, the questionnaire responses (identified by participant numbers) were recorded. One researcher reviewed the video footage and recorded the dismissal times (from when the sounds went off to when they stopped). They also created a log of qualitative observations for each participant. The video files could not be stored on a cloud storage service and due to the COVID-19 campus closure, only one researcher was able to review this data.

Removing Outliers

The recorded measurements were then reviewed for outliers. In one study, the researchers failed to inform the participant on the exact method to dismiss the alert. The participant noticeably struggled to dismiss the first alarm during the experiment resulting in an abnormal dismissal time (36 seconds as opposed to 12 and 15 seconds for the second and third alarms). Therefore, the researchers reached a consensus to mark this participant's dismissal times as invalid and removed it from the sample. All other measurements were deemed as valid.

Statistical Analysis

With a sample size of n=9 for dismissal times and n=10 for all other measurements, the mean values and standard deviation values were calculated. (Appendix B) . Figures 6-9 displaying the results are shown on the next page.

	Siren vs. Chime	Siren vs. Text-to-speech	Chime vs. Text-to-Speech
Dismissal Time	0.45	**0.001**	**0.01**
Urgency	**0.03**	0.97	**0.02**
Positive Affect	0.11	0.86	**0.04**
Negative Affect	**0.01**	0.16	0.25

Table 2: P-values for Post-hoc Tukey's Test

To compare the performance of the different alarm sounds, one-way ANOVA for repeated measures was computed (α=0.05). P-values >0.05 were found for all dependent variables except for annoyance (p=0.078).

Next, post-hoc Tukey's HSD tests were computed to make pairwise comparisons for the variables demonstrating significant differences. The resulting p-values are recorded in Table 2 with statistically significant values <0.05 shown in bold.

Dismissal Times (Figure 6)

As observed from the p-values, the participants took longer to dismiss the text-to-speech sound compared to the siren (p=0.001) and compared to the chime sound (p=0.01). This supports H6. However, there was no significant difference in response times between the siren and the calm sound thus H1 is not supported.

Perceived Annoyance (Figure 7)

The siren sound did receive the highest mean value of perceived annoyance. However, there was no significant difference in perceived annoyance while the participants performed the gaming task. This fails to support H2. Note from Figure 7 that the interquartile range is wide for the annoyance ratings showing large differences between individual ratings with participants providing responses at both extremes for all three sounds.

Perceived Urgency (Figure 8)

Even though no significance was observed for reported annoyance, the perceived urgency of the chime sound was significantly lower than the other two sounds (chime-siren: p=0.03, chime-speech: p= 0.02). This supports H5. The perceived urgency of the siren was only significantly different from the chime sound and therefore, H3 is not fully supported.

PANAS-SF Responses (Figure 9)

The PANAS-SF scores are computed by taking the sum of the responses for the five 5-point positive items to produce a Positive Affect (PA) score and summing the responses for the five negative items to produce a Negative Affect (NA) score. The scores range from 5-25 with a higher score indicating a higher affect.

In general, the chime sound received lower PA and NA scores than the other two sounds. Particularly, it received a significantly lower PA score than the text-to-speech sound (p=0.04) and a significantly lower NA score than the siren (p=0.01). The lower scores received by the chime sound indicate interaction with this sound was more neutral and has a lower emotional affect on the receiver. Meanwhile, the siren had a stronger negative affect on the receiver.

Though no significant difference was found in affect scores between the siren and the text-to-speech sound, participants provided additional information on their feelings towards the sounds in the unstructured interviews after the study. Participants generally expressed discontent towards the siren sound pointing out its similarity to the existing emergency alert tone reminding them of negative experiences with mobile alerts (e.g. startling them and providing irrelevant information). Meanwhile, two participants indicated that they had a positive impression of the text-to-speech sound as they found it "calming" and "informative" compared to the other sounds.

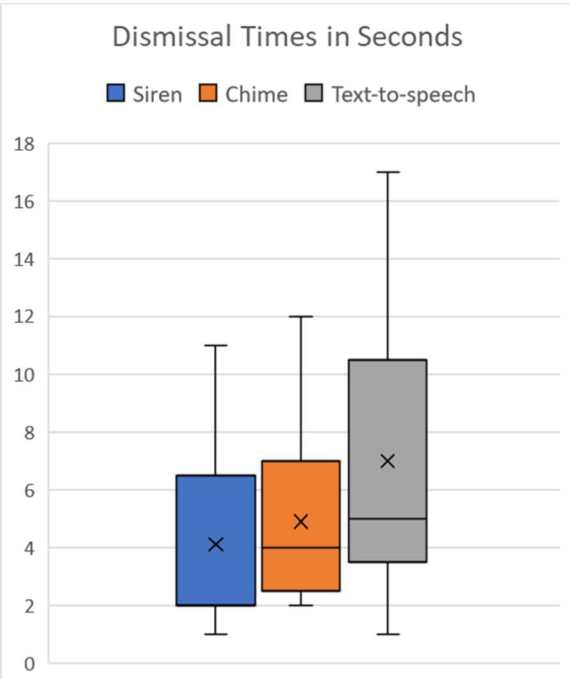

Figure 6: Dismissal Times in Seconds

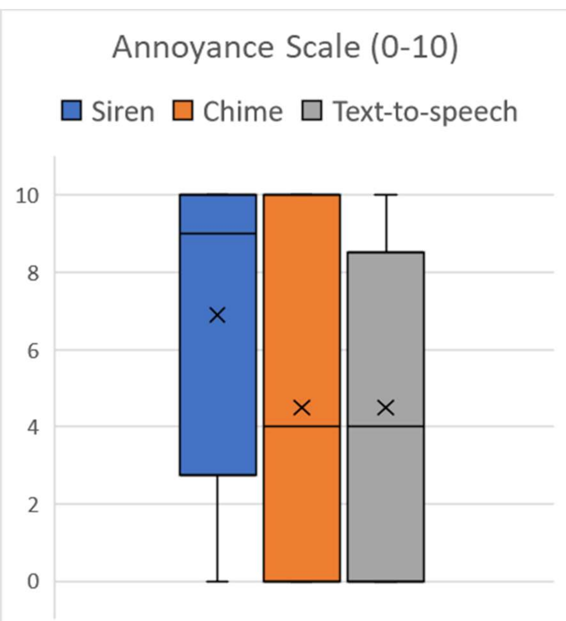

Figure 7: Perceived Annoyance Scale Ratings

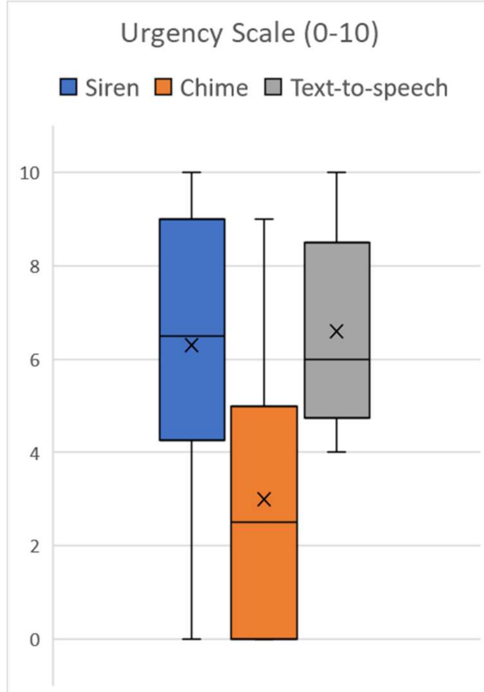

Figure 8: Perceived Urgency Scale Ratings

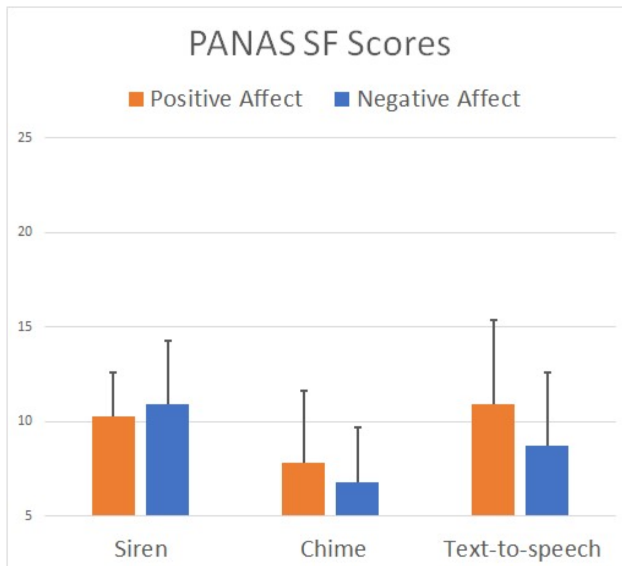

Figure 9: Perceived Annoyance Scale Ratings

LIMITATIONS

The first important limitation was the COVID-19 pandemic and lockdown limiting time to properly prepare and have more control over the factors of the study. A consideration for the between-participant study was an option for more control as compared to the within-participant design we eventually opted for due to time and available participants. For example, the order and number of alarms would have been better as exposure to the three sounds in series raised the expectations of participants and may have affected their reaction to the second or third alarm. In general, a larger sample size would also improve the validity of our statistical results.

The choice of the three sounds for this study was dependent on researchers' judgement and perception of pleasant and unpleasant. Future studies may look into literature on the psychology and perception of sounds or investigate choice of sounds with a study. The effect of the game engagement on participants was also a limitation as some participants admitted to not caring about the sounds or the alerts just so they could continue with the game and get high scores. Another contributing factor to that limitation was the not so obvious pause button for the game.

Another observation from one of the participants was that the sound from the game and the alerts clashed. He explained how he could not figure the sounds out at some point, which we believe is likely to have affected his response time. Individuals participating in the study had personal preferences and experiences which may have affected their interaction with these alerts. For example, some participants mentioned that they always had the habit of dismissing the siren sound for current emergency alerts anytime it comes up on their phones. Annoyance and urgency have subjective interpretations from individuals which may have reflected in the scaling. The lab setting used may not have been a perfect simulation of a real-world scenario. The participants may have been consciously aware of being watched and some found it a little difficult to perceive the mobile phone as theirs which may have affected their responses.

DISCUSSION AND FUTURE WORK

From the study results, the text-to-speech alert had the longest dismissal times and a perceived urgency not significantly different from the siren sound. The similar perceived urgency ratings resemble the findings in [23] where comparison between speech and non-speech warnings yielded no difference between the modes in conveying urgency. The longer dismissal times and observations during the study suggest that participants directed more attention to the text-to-speech alert and took more time to read the alert although the actual impact of the alert tone on whether receivers read messages and message retention is an area that needs to be studied in separate future studies.

To continue, the text-to-speech alert also scored low on annoyance which may have beneficial effects such as reducing the public's negative perception of the alerts and hopefully prevent people from dismissing the importance of the alerts. The results are promising and put a strong case to further research text-to-speech alerts for mobile emergency alert systems. This being said, speech warning design is complicated and changing properties of speech warnings such as the length of the message, intonation, perceived gender could yield vastly different results. Thus, this study reveals benefits to text-to-speech warning sounds such as lower annoyance with similar urgency and the author's suggest further consideration of their usage in public warning systems.

Another notable observation from the study is that sounds like the chime sound and other sounds used for mobile notifications, are perceived as neutral and easily blend with the surroundings or existing application sounds. The neutrality of the sound (lower positive and negative affect) lowers the perceived urgency of the sound too much. But, participants still found the sound "annoying" which suggests that sounds that are too neutral and should be avoided for emergency alerts. This is similar to the findings by Stephen et al. as described in [12] where abstract alarms were found to have poorer performance and memorability.

Next, it is interesting to note that participants provided feedback that altering haptic feedback may improve their experience with mobile emergency alerts. One particular participant specifically mentioned how they prefer haptic feedback for receiving messages and how a unique haptic pattern would annoy them less and likely still catch their attention. Since there was no significant difference in perceived annoyance while varying the sound, it might be interesting to see how changing a different dimension of the alerts (haptic feedback) may influence perceived annoyance. This would need further future research.

Last, meaningful situational effects were also observed in the participants when the alerts went off. Participant reaction to the sounds was very dependent on their immersion in the gaming task. Participants who were more engaged in the game reported that they did not recognize that different sounds were used. Engaged participants also reported higher annoyance which led them to quickly dismiss alerts or completely ignore the alert sound. As noted in [30] emotions are a key component of auditory perception. As the motivations behind this study are to understand how to improve public opinion towards warning sound, future work may wish to focus on implementing better controls to observe participants under a consistent immersive and emotional state.

ACKNOWLEDGEMENTS

We would like to acknowledge Professor Mark Hancock, PhD Candidate and Teaching Assistant Milad Soroush, our MSCI 630 classmates, and study participants for their kind cooperation in developing and executing this study. Despite the unexpected circumstances due to the pandemic, this project has provided us with an enriched learning experience in the field of HCI and we greatly appreciate all the feedback and support we have received.

APPENDIX

A. Post-Task Questionnaire

1. *Briefly describe the alert sound. How did it make you feel?* (open answer)
2. *Did you stop (dismiss) the alert with this sound? Why or why not?* (open answer)

3. *On a 0-10 opinion scale how much did this alert sound **annoy** or **disturb** you? If you are not at all annoyed choose zero, if you are extremely annoyed choose ten, if you are somewhere in between choose a number between zero and ten.*

Answer: ______

4. *On a 0-10 opinion scale, rate the **urgency** of the alert with this sound If it was not urgent at all choose zero, if it was extremely urgent choose ten, if you are somewhere in between choose a number between zero and ten.*

Answer: ______

5. *Thinking back on the experiment, please circle the extent to which you felt this way when hearing this alert sound. If the word does not apply, circle 1.*

	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
Upset	1	2	3	4	5
Hostile	1	2	3	4	5
Alert	1	2	3	4	5
Ashamed	1	2	3	4	5
Inspired	1	2	3	4	5
Nervous	1	2	3	4	5
Determined	1	2	3	4	5
Attentive	1	2	3	4	5
Afraid	1	2	3	4	5
Active	1	2	3	4	5

B. MEAN (SD) VALUES

Variable	Siren	Chime	Text-to-Speech
Dismissal Times (s)	4.11 (3.30)	4.89 (3.59)	7.00 (4.97)
Annoyance	6.90 (3.90)	4.50 (4.33)	4.50 (4.20)
Urgency	6.30 (3.27)	3.00 (3.33)	6.60 (2.32)
Positive Affect	10.30 (2.26)	7.80 (3.79)	10.90 (4.46)
Negative Affect	10.90 (3.35)	6.80 (2.86)	8.70 (3.86)

REFERENCES

[1] 2020. Alert Ready, Canada's New Emergency Alert System. Retrieved January 25, 2020 from https://www.alertready.ca/

[2] Michael Oliveira. 2013. Nearly 50% of Canadian mobile users say their device is on hand 24/7: poll. Retrieved March 26, 2020 from https://toronto.citynews.ca/2013/12/27/nearly-50-of-canadian-mobile-users-say-their-device-is-on-hand-247-poll/

[3] Nick Westoll. 2019. Peel police detail complaints made to 911 about Amber Alert for 11-year-old girl who was murdered. *Global News*. Retrieved January 25, 2020 from https://globalnews.ca/news/5087521/riya-rajkumar-amber-alert-peel-police-911-complaints/

[4] PEASI. 2019. *Survey Results Canada's Alert Ready Test November 27, 2019*. Calgary. Retrieved March 26, 2020 from https://peasi.com/wp-content/uploads/2019/12/Alert-Ready-November-27-Test-Survey-Results.pdf

[5] 2018. *Emergency alert and warning systems: Current knowledge and future research directions*. National Academies Press. https://doi.org/10.17226/24935

[6] Federal Communications Commission. 2019. *Report: October 3, 2018 Nationwide WEA and EAS Test* Washington DC. Retrieved Jan 25, 2020. https://docs.fcc.gov/public/attachments/DOC-356902A1.pdf

[7] RMIT University. 2011. *Systematic Review of Reports on Emergency Alert (Prepared for the Office of the Emergency Services Commissioner by the Centre for Risk and Community Safety, RMIT University.)*. Victoria, Australia.

[8] Alastair McAslan. 2011. *Assessment of the Effectiveness of Emergency Alert - Final Report*. Victoria. Retrieved March 26, 2020 from http://www.emergencyalert.gov.au/images/stories/Assessment_of_the_Effectiveness_of_Emergency_Alert_July_2011.pdf

[9] Sumeet Kumar, Hakan Erdogmus, Bob Iannucci, Martin Griss, Martin Griss Associates, Usa João Diogo Falcão, and João Diogo Falcão. 2018. Rethinking the Future of Wireless Emergency Alerts: A Comprehensive Study of Technical and Conceptual Improvements. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol* 2, 2: 71. https://doi.org/10.1145/3214274

[10] Paul Ngo and Duminda Wijesekera. 2011. Enhancing the usability of the commercial mobile alert system. In *IFIP Advances in Information and Communication Technology*, 137–149. https://doi.org/10.1007/978-3-642-24864-1_10

[11] Neville A. Stanton and Judy. Edworthy. 1999. *Human factors in auditory warnings*. Ashgate.

[12] A Guillaume and C Drake. 2006. Non Vocal Auditory Warning Signals. In *Advances in Psychology Research*, Alexandra Columbus (ed.). Nova Science Publishers, Incorporated, New York. Retrieved from https://books.google.ca/books?id=3-M8viRhZ9gC

[13] R. D. Patterson. 1990. Auditory warning sounds in the work environment. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 327, 1241: 485–492. https://doi.org/10.1098/rstb.1990.0091

[14] Judy Edworthy, Elizabeth Hellier, Kathryn Walters, Ben Weedon, and Austin Adams. 2000. The Relationship between Task Performance, Reaction Time, and Perceived Urgency in Nonverbal Auditory Warnings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44, 22: 674–677. https://doi.org/10.1177/154193120004402247

[15] Ellen C. Haas and John G. Casali. 1995. Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. *Ergonomics* 38, 11: 2313–2326. https://doi.org/10.1080/00140139508925270

[16] Asa Skagerstrand, Susanne Köbler, and Stefan Stenfelt. 2017. Loudness and annoyance of disturbing sounds – perception by normal hearing subjects. *International Journal of Audiology* 56, 10: 775–783. https://doi.org/10.1080/14992027.2017.1321790

[17] 1. Todd A. Mondor and G. Allen Finley. 2003. The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian Journal of Anesthesia* 50, 3: 221–228. https://doi.org/10.1007/BF03017788

[18] F E Block Jr. 1992. Evaluation of users' abilities to recognize musical alarm tones. *Journal of Clinical Monitoring* 8, 4: 285–290. Retrieved March 26, 2020 from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=1453188

[19] Hugues Nélisse, Chantal Laroche, Pascal Laferrière, Christian Giguère, Jérôme Boutin, and Véronique Vaillancourt. 2013. Comparison of sound propagation and perception of three types of backup alarms with regards to worker safety. *Noise and Health* 15, 67: 420. https://doi.org/10.4103/1463-1741.121249

[20] Martin Pielot, Karen Church, and Rodrigo De Oliveira. 2014. An in-situ study of mobile phone notifications. In *MobileHCI 2014 - Proceedings of the 16th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services*, 233–242. https://doi.org/10.1145/2628363.2628364

[21] Stuart J. McFarlane, Jair E. Garcia, Darrin S. Verhagen, and Adrian G. Dyer. 2020. Alarm tones, music and their elements: Analysis of reported waking sounds to counteract sleep inertia. *PLOS ONE* 15, 1: e0215788. https://doi.org/10.1371/journal.pone.0215788

[22] Gary A. Smith, Mark Splaingard, John R. Hayes, and Huiyun Xiang. 2006. Comparison of a personalized parent voice smoke alarm with a conventional residential tone smoke alarm for awakening children. *Pediatrics* 118, 4: 1623–1632. https://doi.org/10.1542/peds.2006-0125

[23] Judy Edworthy, Kathryn Walters, Elizabeth Hellier, and Ben Weedon. 2000. Comparing Speech and Nonspeech Warnings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44, 22: 746–749. https://doi.org/10.1177/154193120004402265

[24] Provincial Emergency Operations Centre. *Radiological Hazard Issued Jan 12 at 7:23 AM,* Retrieved March 26, 2020 from https://alertable.ca/#/details/2020/119430 .

[25] Provincial Emergency Operations Centre. *Radiological Hazard Issued Jan 12 at 9:11 AM.* Retrieved March 26, 2020 from https://alertable.ca/#/details/2020/119476 .

[26] Ontario Provincial Police. *Amber Alert Issued Feb 14 at 7:23 AM,* Retrieved March 26, 2020 from https://alertable.ca/#/details/2019/74176

[27] Steam. *Zuma's Revenge!,* Retrieved March 26, 2020 from https://store.steampowered.com/app/3620/Zumas_Revenge/

[28] Truls Gjestland. *Standardized general-purpose noise reaction questions*. Zurich, Switzerland. Retrieved March 26, 2020 from http://www.icben.org/2017/ICBEN 2017 Papers/SubjectArea06_Gjestland_0611_2449.pdf.

[29] Andrew Mackinnon, Anthony F. Jorm, Helen Christensen, Ailsa E. Korten, Patricia A. Jacomb, and Bryan Rodgers. 1999. A short form of the Positive and Negative Affect Schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences* 27, 3: 405–416. https://doi.org/10.1016/S0191-8869(98)00251-7

[30] Penny Bergman, Anders Sköld, Daniel Västfjäll, and Niklas Fransson. 2009. Perceptual and emotional categorization of sound. *The Journal of the Acoustical Society of America* 126, 6: 3156–3167. https://doi.org/10.1121/1.3243297