

Effect of Distinct Ambient Noise Types on Mobile Interaction

ZHANNA SARSENBAYEVA, The University of Melbourne, Australia

NIELS VAN BERKEL, The University of Melbourne, Australia

EDUARDO VELLOSO, The University of Melbourne, Australia

VASSILIS KOSTAKOS, The University of Melbourne, Australia

JORGE GONCALVES, The University of Melbourne, Australia

The adverse effect of ambient noise on humans has been extensively studied in fields like cognitive science, indicating a significant impact on cognitive performance, behaviour, and emotional state. Surprisingly, the effect of ambient noise has not been studied in the context of mobile interaction. As smartphones are ubiquitous by design, smartphone users are exposed to a wide variety of ambient noises while interacting with their devices. In this paper, we present a structured analysis of the effect of six distinct ambient noise types on typical smartphone usage tasks. The evaluated ambient noise types include variants of music, urban noise and speech. We analyse task completion time and errors, and find that different ambient noises affect users differently. For example, while speech and urban noise slow down text entry, being exposed to music reduces completion time in target acquisition tasks. Our study contributes to the growing research area on situational impairments, and we compare our results to previous work on the effect of cold-induced situational impairments. Our results can be used to support smartphone users through adaptive interfaces which respond to the ongoing context of the user.

CCS Concepts: • **Human-centered computing** → Empirical studies in HCI; • **Human-centered computing** → Ubiquitous and mobile computing; • **Human-centered computing** → Smartphones;

Additional Key Words and Phrases: Smartphones, ambient noise, situational impairments, mobile interaction, performance.

ACM Reference Format:

Zhanna Sarsenbayeva, Niels van Berkel, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2018. Effect of Distinct Ambient Noise Types on Mobile Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 82 (June 2018), 23 pages. <https://doi.org/10.1145/3214285>

1 INTRODUCTION

Research in Ubiquitous Computing has been largely driven by the realisation that the context of use has a substantial influence on the interaction with a system. When non-conventional contextual and environmental factors negatively affect the use of the system, they lead to *situational impairments* [13, 16, 17, 45, 53]. As constant companions for most users, smartphones are frequently used in situations where the user is situationally impaired, e.g., due to cold environments [17], motion [14], or encumbrance [37, 38]. Therefore, one of the key challenges in our community is to enhance the capability of mobile devices to detect situational impairments and to adapt the interaction accordingly.

Authors' addresses: Zhanna Sarsenbayeva, The University of Melbourne, Melbourne, Australia, z.sarsenbayeva@student.unimelb.edu.au; Niels van Berkel, The University of Melbourne, Melbourne, Australia, n.vanberkel@student.unimelb.edu.au; Eduardo Velloso, The University of Melbourne, Melbourne, Australia, eduardo.velloso@unimelb.edu.au; Vassilis Kostakos, The University of Melbourne, Melbourne, Australia, vassilis.kostakos@unimelb.edu.au; Jorge Goncalves, The University of Melbourne, Melbourne, Australia, jorge.goncalves@unimelb.edu.au.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, or to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

2474-9567/2018/6-ART82 \$15.00

<https://doi.org/10.1145/3214285>

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 2, No. 2, Article 82. Publication date: June 2018.

In this paper we investigate the effects of one such situational impairment on mobile interaction that remains relatively underexplored [46]—*ambient noise*. Ambient noise is defined as any noise the user is exposed to while their attention is directed at some task or activity [25]. This includes disruptive sounds like the hammering of construction work, the barely perceptible humming of a fridge, the pleasant music playing on headphones, and the disruptive conversation of co-workers. Even though people commonly use their mobile devices in environments where they are exposed to some level of ambient noise (e.g., public places, cafeterias) [25], little is known about their effect on mobile interaction. This paper aims to fill this gap.

Research in Cognitive Science and related fields has highlighted the negative impact that noise has on human behaviour [22, 39], cognitive performance [5, 44, 54], and emotional state [25]. Banbury *et al.* suggest that this effect is mainly caused by the spread of attention while completing a primary task [6]. They state that even if the attention is directed elsewhere, sound is perceived and processed by the brain and, hence, diverts the attention from the main task, causing degradation in performance. Therefore, it is likely that ambient noise can also have an adverse effect on mobile interaction in certain situations. However, due to limited research, it is unclear what is the magnitude of this effect, if there are any differences between different types of noises, and how ambient noise compares to other situational impairments in the context of mobile interaction.

To explore these issues, we investigate how different categories of common ambient noises — *music*, *urban noise*, and *speech* — affect mobile interaction while performing typical smartphone activities: target acquisition, visual search and text entry. To further unpack the effects of the different types of noise, we explore two variants of each. We presented users with *slow* and *fast* tempo music, as previous work has shown that different music tempos have varying effects on cognitive performance [22, 39]. We also presented users with two types of urban noise: *outdoor* and *indoor*. Outdoor urban noise has been shown to have a significant adverse effect on memory tasks. However, there is no consensus regarding the effect of indoor urban noise: some studies report its distracting effect during memory tasks [5, 54], while others did not observe any effect on cognitive performance [44]. Finally, we presented users with two types of speech: *meaningful* and *meaningless* speech. Meaningful speech is speech spoken in a familiar language, whereas meaningless speech is defined as the speech presented in a language unfamiliar to the listener [32]. To contextualise our study within the body of research on situational impairments, we compare our findings to previous work on cold-induced situational impairments [45], which collected similar data to the one described in our study.

In summary, this paper advances the state-of-the-art on situational impairment research in Ubiquitous Computing through three main contributions. First, we investigate how different types of ambient noise affect the interaction with mobile devices during three typical activities: target acquisition, visual search, and text entry. Second, we contribute towards the situational impairments research agenda by comparing our findings to the effects of cold-induced situational impairments. Third, we discuss and provide recommendations on the detection of certain types of ambient noise that can affect mobile interaction and how to accommodate a situationally impaired user in such cases.

2 RELATED WORK

2.1 Impact of Situational Impairments on Mobile Interaction

Previous research has shown that interaction with mobile devices can be adversely affected by implicit environmental and contextual factors, subjecting the user to what is known as a *situational impairment* [46, 49]. A number of causes for situational impairment have been studied within the HCI/UbiComp community, including the effects of ambient temperature [17, 45], motion [14], and encumbrance [37, 38].

Regarding cold-induced situational impairments, Gonçalves *et al.* and Sarsenbayeva *et al.* showed a negative effect of cold temperatures on smartphone input performance [17, 45]. The authors found that colder temperatures are associated with lower throughput and accuracy in tapping tasks. Other studies investigated the effect of

user motion on interaction with mobile phones. For example, walking has been found to adversely affect mobile interaction in completing typing tasks [14, 36], and target acquisition tasks [48]. Furthermore, Ng *et al.* [37, 38] demonstrated the negative impact of encumbrance on target acquisition tasks during mobile interaction. In particular, encumbrance has been shown to decrease accuracy, while increasing error rate and target selection time [38]. Given the effect situational impairments have on mobile interaction, previous work was conducted to explore ways of detecting situational impairments and develop design solutions to accommodate them. For example, Sarsenbayeva *et al.* [47] suggest using smartphone's battery temperature to detect ambient temperature drop. Similarly, Goel *et al.* [14] successfully utilise a smartphone's accelerometer sensor to detect walking. In terms of design solutions, researchers have shown that increasing target size [48], providing audio guidance [52] and adaptive text entry [14] can compensate the negative effect of situational impairments.

In comparison to other causes of situational impairment, the effect of ambient noise on mobile interaction lacks appropriate investigation [46]. Previous work emphasises the importance of studying unexplored causes of situational impairment in order to broaden the research scope in this area [46]. There is little prior work that has studied the effect of ambient noise on mobile interaction. A study by Hoggan *et al.* [21] showed that loud noise levels had an adverse effect on participants' performance when completing text entry tasks on mobile phones. However, the aim of their study was simply to identify the noise threshold at which the audio and tactile feedback become ineffective. Harvey & Morgan [18] demonstrated that noisy environments have a negative effect on user performance during web search tasks. They showed an effect of noise on task performance using questionnaire data only and did not provide any quantitative metrics regarding performance. To the best of our knowledge, this is the first study to investigate and quantify the effect of ambient noise as a potential cause of situational impairment during mobile interaction.

2.2 Effect of Ambient Noise on Human Behaviour and Cognitive Performance

Previous work has shown that ambient noise has an effect on human daily activities in terms of behavioural [29, 33, 39], emotional [25], and cognitive [5, 54] performance. As mobile interaction is now an essential part of people's everyday activities, it is important to investigate its impact on mobile interaction. Next, we discuss related work on the impact of music, urban noise and speech on human behaviour and cognitive performance.

2.2.1 Effect of Background Music on Human Behaviour and Cognitive Performance. Previous works have investigated the effect of music on people's physical activities [33], cognitive performance [26], and mood [25]. Cassidy *et al.* showed a significant negative effect of heavy metal music on immediate recall, free recall, delayed recall and performance in Stroop tasks [15] compared to a silent condition [10]. Similarly, Wen *et al.* demonstrated that performance in recall is significantly higher under classical music in contrast to rock music [55]. However, there is no consensus regarding the effect of background music on reading [25]. Whereas some studies report an increase on reading performance [41], other studies failed to replicate this effect [32]. Another study showed that slow tempo music resulted in longer reading time and poorer reading efficiency in contrast to fast tempo music [25]. Henderson *et al.*'s results suggest that what is detrimental is not the music itself, but the speech contained in it [19]. The authors found that participant performance in completing reading tasks significantly decreased when they were exposed to popular music containing lyrics, but found no significant decrease with classical music or silence [19]. Given these results, we evaluate the effects of music and speech in separate study conditions.

The most widely studied musical element in terms of the effect of music on human behaviour and cognitive performance has been the *tempo*, *i.e.*, the speed of the underlying beat of the music. Related work suggests that the faster the tempo, the faster people complete tasks. Milliman showed that faster tempo increased the walking speed of customers, decreasing the average time customers spent in a store [34]. Research in restaurants and cafeterias has shown that faster music also caused people to eat and drink faster [9, 33, 35]. Conversely, slower

music was shown to slow down people's eating and drinking [9, 33, 35] as well as to prolong the perception of elapsed time spent in casinos [40].

Despite accelerating task completion, previous research suggests that music with a faster tempo is also more cognitively demanding. For instance, Holbrook showed that fast tempo music demanded more cognitive resources to be processed compared to slow tempo music [22]. According to North *et al.*, this phenomenon occurs because more information is perceived and processed by the listener's brain when fast tempo music is played than when slow tempo music is played [39]. Given the different effects of the tempo of the music, we evaluated the effect of music on mobile interaction with fast and slow tempo music.

An interaction effect between the tempo of the music and the gender of the listener has also been identified. For example, Kallinen showed that female and male participants perceived information differently under slow and fast music conditions [25]. In this study, the author found an inverse effect of the tempo of the music on the perception of news articles — a slow tempo led male participants to evaluate the news as negative, whereas it led female participants to evaluate the same set as positive. Further, men read more slowly when slow tempo music was playing, whereas women read more slowly when no music was playing. This suggests that it is important to ensure appropriate gender balance in the participant sample to avoid any gender bias in the results.

Taking into consideration the effect of music tempo on other aspects of human behaviour and cognitive performance, we chose to investigate its effect on mobile interaction. We hypothesise that fast tempo music will result in a faster completion of tasks when compared to the silent condition, particularly those that require fine-motor movements (*e.g.*, target acquisition tasks).

2.2.2 Effect of Urban Noise on Human Behaviour and Cognitive Performance. As with background music, urban ambient noise is also known to influence human behaviour and cognitive performance. In this study, we focus on two types of urban noise: outdoor and indoor. We define outdoor urban noise as noise containing sound as sampled in an urban area: a combination of street, construction, and traffic sounds. We define indoor urban noise as the noise coming from an office or cafeteria: working sounds, people's murmur, and chatter.

A negative effect of urban outdoor noise on human performance has been shown in several studies. For example, Cassidy *et al.* showed a significant negative effect of outdoor urban ambient noise on free, immediate, and delayed recall and Stroop tasks [15] when compared to no noise [10]. Furthermore, Stansfeld *et al.* showed that communities exposed to lower traffic noise had a lower rate of psychiatric hospitalisation [50]. Previous work has also shown that an exposure to a broadband noise at 100 dB for 30 minutes can lead to significant impairments when performing memory tasks [2].

Regarding urban indoor noise, Banbury and Berry demonstrated that students' performance in completing mathematical and recall tasks was significantly worsened with its presence [5]. Furthermore, a two-year longitudinal study by Cohen and Weinstein showed that children in noisy schools with sound levels of 60-80 dB performed worse in solving puzzles and mathematical tasks [11], as reported by Holmberg and Coon [23].

This literature coupled with the fact that it is common practice for mobile device users to interact with their devices while being exposed to urban outdoor (*e.g.*, streets) and indoor (*e.g.*, office, cafeteria) noise, suggests that these types of noises can have an effect on mobile interaction, which we investigate in this paper.

2.2.3 Effect of Speech on Human Behaviour and Cognitive Performance. Previous research has made a distinction between the effects of two types of speech on human behaviour and cognitive performance [32]: *meaningful* — speech that the listener can understand — and *meaningless* — speech that the listener does not understand. For instance, Martin *et al.* found a detrimental effect of continuous meaningful speech on reading performance when compared to silence [32]. The authors compared the effect of meaningful and meaningless speech on reading comprehension and showed that reading performance is significantly worse under both speech conditions as compared to silence, with a stronger effect for meaningful speech [32].

Previous work has also shown that both meaningful and meaningless speech have an equally distracting effect on memory recall [51]. Tremblay *et al.* found that natural speech was more disruptive than sine wave speech (*i.e.*, a form of artificially degraded speech) on tasks requiring memory recall [51]. Moreover, in a study reported in [43] meaningless speech was shown to have a negative effect on memory tasks. Salame *et al.* showed that immediate memory was disrupted the most by the the unattended foreign language speech, when compared to other conditions such as instrumental music, urban noise and silence [44]. These results contradict previous findings that did not find any effect of meaningless speech on cognitive performance in phonological judgement tasks [3]. Nevertheless, the authors argue that the phenomenon needs further investigation [3]. Finally, speech is also known to reduce performance in completing arithmetic tasks and memory for prose tasks [4, 54].

In this paper we explore the effect of both meaningful (English) and meaningless (Kazakh) speech on mobile interaction. Based on the literature, we hypothesise that the performance in tasks with higher cognitive requirements (*e.g.*, text entry) is likely to deteriorate while hearing background speech.

3 STUDY

We operationalise mobile interaction in terms of three common activities conducted on smartphones: target acquisition, visual search, and text entry. For this study we used two software applications as experimental tasks, *TapCircle* and *FindIcon*, previously used to investigate the effect cold-induced situational impairments on mobile interaction [45]. This allows for direct comparison of our results to those previous findings. Additionally, we developed a new custom software called *TypeMe*. The tasks presented by *TapCircle* measures users' fine-motor performance during target acquisition, the task in *FindIcon* measures users' cognitive performance during visual search [45], and the task in *TypeMe* measures users' text entry performance. The three tasks were combined into one Android application and were presented in a random order. Details for each task are presented below, the application interface is shown in Figure 1.

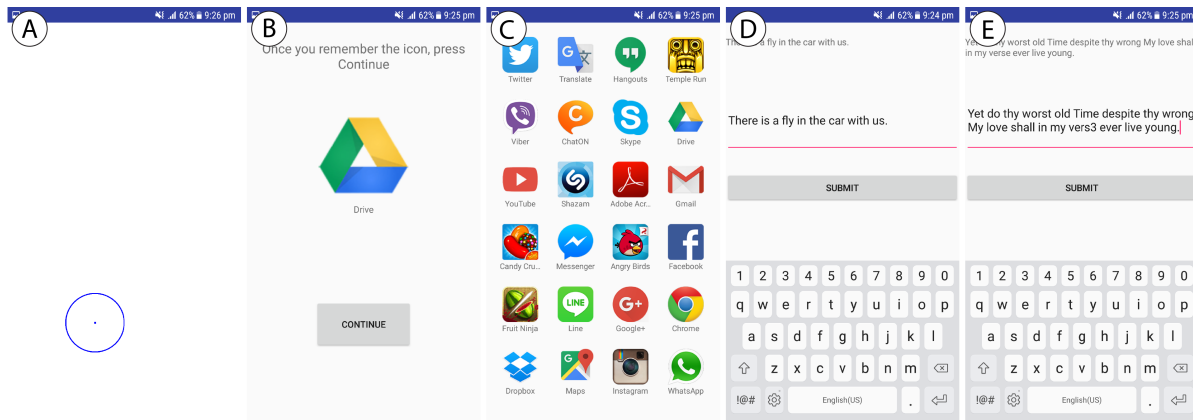


Fig. 1. Interface of the application with TapCircle task (A), FindIcon task (B-C), and TypeMe task with user's input for easy and difficult texts (D-E)

3.1 Tasks

3.1.1 Task 1: Target Acquisition. Circular targets with a radius of 135 pixels randomly appear one at a time in each position of a 4×6 grid at least once. A similar task was used by Henze *et al.* [20]. Every target has an indicated centre and participants were instructed to tap these circles as quickly and as precisely as possible. The

application logged the coordinates of the target's centre and participant's touch point, the elapsed time, and the position of the circle on the grid. Figure 1-A shows a screen of the TapCircle task interface.

3.1.2 Task 2: Visual Search. In the FindIcon task, participants were first presented with an icon which they subsequently had to locate within a grid of other icons. Participants were free to look at the target icon for as long as they wanted to memorise it. In the following screen, participants were required to locate and tap the target icon amongst a set of 24 icons (100×100px) in a 4×6 grid. To minimise any possible learning effects, each of the 24 icons was designated as a possible target icon in random order in each condition. The application also ensured that every grid position hosted a target icon in a random order. To make our results comparable, we used the same set of application icons as described by Sarsenbayeva *et al.* [45]. Figures 1-B-C show the interface of the task. The application recorded time spent on memorising an icon, time spent on locating and tapping the target icon, grid position of the icon, and the coordinates of the centre of the icon and of the participant's touch point.

3.1.3 Task 3: Text Entry. In the text entry task, the application presented some text at the top of the screen and participants were asked to type it verbatim in the text-box below it. In order to have varying complexity, we distinguished between easy and difficult texts. Easy texts consisted of only one sentence and contained common words that are used on a daily basis. Difficult texts were selected from Shakespeare's sonnets and consisted of more than 1 sentence. We quantified the difficulty of the sentences using the Flesch-Kincaid readability test [28]. The easy texts had an average Flesch-Kincaid grade level of 1.6, whereas the difficult texts had an average Flesch-Kincaid grade of 5.1. Participants completed both an easy and a difficult text during the task. Figures 1-D-E show the task as presented to the participants, including examples of an easy and a difficult text.

3.2 Hardware

Participants completed all tasks using a Samsung Galaxy S7 smartphone running Google's Android 7.0 (Nougat). The smartphone has a 5.1-inch screen with a resolution of 1080×1920px. The smartphone was selected due to identical screen size and resolution to the smartphone's screen parameters used in Sarsenbayeva *et al.*'s evaluation of cold-induced situational impairments [45]. This allowed us to directly compare the effects of noise-induced situational impairments to cold-induced situational impairments on mobile interaction.

3.3 Experimental Conditions

We selected the following experimental ambient noise conditions for this study given their demonstrated effect on human behaviour and cognitive performance in the literature:

- music (fast and slow tempo)
- urban ambient noise (indoor and outdoor)
- speech (meaningful - English, meaningless - Kazakh)
- silence, which acted as a control condition.

In the music conditions we used the same composition sampled at fast and slow rates to avoid other characteristics of the music, other than tempo (*e.g.*, pitch, timbre), to affect the participants performance. This is because tempo is considered to be the main factor in music to affect human behaviour and performance [25]. We chose music that did not contain any lyrics in order to avoid overlap with the speech condition. Given these guidelines, we selected Bach's "Brandenburg Concerto No. 2". As the original tempo of the composition is considered to be fast (92 beats per minute) [25], we used it as is for the fast tempo music condition. For the slow tempo music condition we sampled the same composition at 60 beats per minute, below the 70 beats per minute considered in the literature to be slow [25], without any noticeable decrease in sound quality.

As for the urban outdoor ambient noise condition, we chose a clip of street noise containing road traffic, vehicles' motor sounds, honking, and indistinguishable crowd speech. For the urban indoor condition we chose a

cafeteria noise composed of coffee machine, cutlery sounds, and indistinguishable people's murmur. Both of the sounds were obtained from YouTube (urban outdoor ¹, urban indoor ²).

For the meaningful speech condition, we selected a weather forecast presented in English. For the meaningless speech condition we selected TV news presented in a language that participants did not understand (Kazakh). Both clips were narrated by a female voice and did not contain any other sounds (e.g., background noise, music, speech of different people).

The volume level for all of the ambient noise conditions was kept in the range of 55–60 dB following Iwanaga and Ito's design guidelines, to avoid discomfort to our participants [24]. All audio files were long enough for participants to complete all tasks before the audio clip ended.

3.4 Participants and Procedure

We recruited twenty-four participants (12 male, 12 female) through our university's mailing lists. We balanced gender as the literature suggests men and women can react differently to the same noise conditions [25]. Participants were aged between 19 and 54 years ($M = 31.67$, $SD = 8.95$) and had a diverse range of educational backgrounds (e.g., Accounting, Geology, Linguistics, Biochemistry, Computer Science, Elderly Care). All participants were fluent in English and used it as their main language of communication at work. The foreign language presented in this study ('meaningless speech') was not known to any of the participants. Each participant was assigned a unique anonymous ID (participant ID) in our study.

Our study had a within-subjects experimental design. Condition acted as an independent variable and the order of conditions presented to the participants was counter-balanced. This way, we minimised the impact of any potential fatigue or learning effects. Unlike the study protocol presented in [45], participants completed all tasks in only one smartphone holding posture as opposed to two different holding postures. We instructed our participants to complete all tasks in two-handed interaction mode while standing (i.e., interacting with the phone with index finger of the dominant hand while holding the phone in the non-dominant hand [45]). Figure 2 shows the participant completing a task during the experiment. As each participant was required to complete three tasks in each of seven conditions, including a second smartphone holding posture (and essentially doubling study duration) would have likely introduced considerable fatigue and irritation among our participants. Each participant was rewarded with a \$10 voucher for participation.

We collected data for several dependent variables to measure performance while conducting the tasks. In the target acquisition task, we recorded the time taken to tap a circular target and the size of the tap offset. In the visual search task, we recorded the time taken to find the correct icon, the time taken to memorise an icon, and the number of incorrectly selected icons. Finally, in the text entry task, we recorded the time taken to complete typing a text, the number of characters per text, and the number of errors within the text that were not corrected by participants. We measured the typing performance with the time taken per character entry by dividing the total typing time by the number of characters per text.

Upon arrival at our usability lab, participants were briefed about the purpose of the study and asked to sign a consent form. We then collected their personal details (age, gender, background, native language, dominant hand) and performed a training procedure in order for the participants to get acquainted with the tasks. During the training, participants completed all three study tasks in a random order, until they were comfortable with each one, in order to minimise any potential learning effects.

Once the training was completed, participants started the experiment. As the study contained four main conditions (silence, music, urban noise, speech), we created 24 combinations of the conditions to ensure proper counter-balancing. This means that no participant had a repeated order of conditions presented to them. Three

¹<https://www.youtube.com/watch?v=cDWZkXjDYsc>

²<https://www.youtube.com/watch?v=ISlOXT3AKBo>

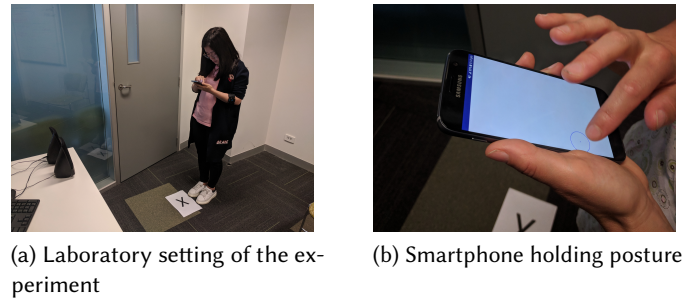


Fig. 2. Participant completing tapping circles task

of the conditions included sub-conditions (music fast/slow tempo, urban noise outdoor/indoor, speech meaningful/meaningless). We randomly allocated the participants into two groups. Within each of these conditions, the first group was exposed to fast music tempo, indoor urban noise, and meaningful speech conditions first. The second group, opposite to the first group, was exposed to slow music tempo, outdoor urban noise, and meaningless speech first. Both groups experienced both sub-conditions.

During the experiment, participants had to complete the same set of tasks in each of seven conditions. Each round of tasks consisted of: target acquisition, visual search, and text entry. The order of the tasks was randomised by the application, and the participants had to complete each task one after another without taking any breaks. Once the sound of a condition started playing, the researcher started a timer. After being exposed to the sound for one minute, the participant was then instructed to begin that round of tasks. Listening to the sound for one minute ensured participants were more accustomed to that particular ambient noise condition [17, 45].

After the experiment was completed, we conducted a short semi-structured interview. Participants were asked about their personal perception of their performance during the conditions, whether any of the sounds were experienced as particularly distracting, and whether they usually listen to music or any other sounds while working, reading, or performing tasks that demand concentration. Finally, we enquired if they listen to music or other type of audio when they interact with their smartphones. The experiment lasted approximately 70 minutes per participant, including briefing, training, data collection, and final interview.

4 RESULTS

From our 24 participants, we collected 15,556 and 4,094 target hits for the tapping the circles and finding icons task respectively, and 336 typed sentences for the typing task. Data collected from left-handed participants ($N = 2$) was mirrored relative to the X-axis of the screen for the tapping a circle and finding an icon tasks.

To investigate whether mobile interaction was affected by ambient noise, we built generalised linear mixed-effect models to describe participant performance in the three tasks (target acquisition, visual search, and text entry). Apart from the ambient noise conditions (discussed in Subsection 3.3), we also considered other factors that may affect participants' performance during mobile interaction. We provide a list of these variables below. Unless otherwise stated, the variable was included for all conditions (silence, music, urban noise, and speech).

- **Gender** - A binary variable indicating gender of the participant.
- **Age** - A numeric variable indicating the age of the participant.
- **Music listener** - A binary variable indicating whether the participant typically listens to music or other sounds while working, reading, or performing tasks that require cognitive demand. We only used this predictor in models predicting the effect of music on mobile interaction.

- **Native English speaker** - A binary variable indicating if the participant is a native English speaker. We only used this predictor in models predicting the effect of speech on mobile interaction.
- **X centre coordinate** - X-axis coordinate of the centre of target (circle or icon for tapping a circle and finding an icon task respectively).
- **Y centre coordinate** - Y-axis coordinate of the centre of target (circle or icon for tapping a circle and finding an icon task respectively).
- **Participant ID** - Participant ID was treated as a random effect in order to control for individual differences in our models.

We applied a backfitting algorithm using AIC to all our models for predictor selection. AIC penalises the inclusion of additional predictors, which could lead to overfitting of a model. Finally, to ensure the validity of the models we checked for the presence of multicollinearity. All of the predictors for each of the presented models had a variance inflation factor between 1 and 1.38, well below the often used threshold of 5 to 10 to detect multicollinearity.

4.1 Results: Background music condition

The tested background music conditions are fast tempo music, slow tempo music, and silence. The conditions are tested across all three defined tasks. We discuss the results per task below.

4.1.1 Target Acquisition. Our first model describes the time taken to hit a circle. The final prediction model retained 3 of the 6 predictors, as shown in Table 1. The model is statistically significant ($\chi^2(4) = 56.05$, $p < 0.01$) and describes 7% of variance of the time taken to hit the circle ($Marginal R^2 = 0.07$, $Conditional R^2 = 0.28$). The results indicate that participants were significantly quicker to tap circles in the fast and slow music conditions as compared to the silent condition. However, we did not find a significant difference between the fast and slow music condition; Wilcoxon signed-rank test ($V = 1.17e+06$, $p = 0.21$). A boxplot of the time taken to complete the tapping circles task across the different conditions is shown in Figure 3 (a).

Table 1. Effects of model factors on predicting time taken to hit a circle in music condition

	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	4.63e+02	4.55e+01	10.19	7.68e-10 ***
Condition (Fast music)	-1.51e+01	3.41	-4.43	9.61e-06 ***
Condition (Slow music)	-2.01e+01	3.41	-5.91	3.71e-09 ***
Y centre coordinate	8.76e-03	2.62e-03	3.35	0.001 ***
Age	3.98	1.38	2.89	0.008 **

Significance: '***' <0.001, '**' <0.01, '*' <0.05

Mean values for the offset size per condition for the tapping task are visualised in Figure 3 (b). We built a model to describe the mean offset size when tapping a circle. The final model contains both conditions and the X centre coordinate as predictors, and is summarised in Table 2. The model was significant ($\chi^2(4) = 63.43$, $p < 0.01$, $Marginal R^2 = 0.01$, $Conditional R^2 = 0.10$). Participants were significantly less accurate when tapping the circles in the slow music condition as compared to the silent condition. A Wilcoxon signed-rank test did not show a significant difference between the offset size in the fast tempo music condition and the slow tempo music conditions ($V = 1183600$, $p = 0.21$). From Table 2, we observe that the offset is smaller for the circular targets located closer to the edge of the screen on which participants held their phone.

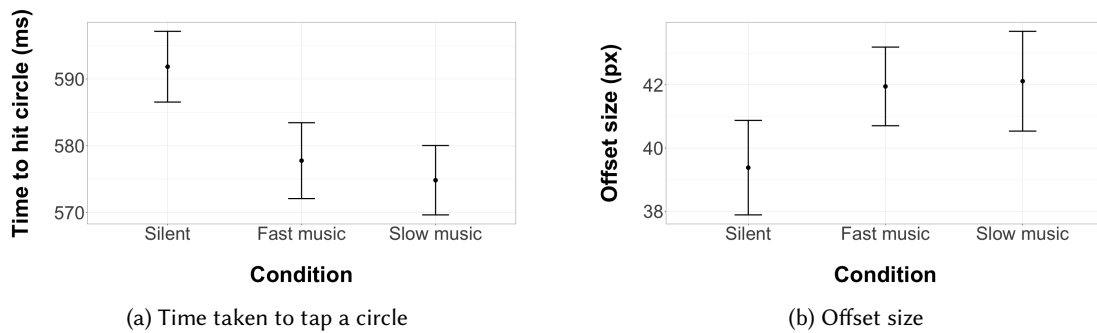


Fig. 3. Mean values for time taken to tap a circle and offset size per condition

Table 2. Effects of model factors on predicting the offset size when tapping a circle in music condition

	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	4.47e+01	2.22	20.17	<2e-16 ***
Condition (Fast music)	1.58	1.01	1.57	0.12
Condition (Slow music)	2.74	1.01	2.72	0.007 **
X centre coordinate	-1.01e-02	1.34e-03	-7.52	6.22e-14 ***

Significance: '***' <0.001, '**' <0.01, '*' <0.05

4.1.2 Visual Search. Next, we built a model to describe the time taken to memorise an icon. The final prediction model contained only age as a predictor. The direction of age was negative, indicating that younger participants took less time to memorise an icon. The model was statistically significant ($\chi^2(1) = 12.26$, $p < 0.01$) and described 9% of variance of time taken to find an icon ($Marginal R^2 = 0.09$, $Conditional R^2 = 0.22$).

We built another model to describe the time taken to find an icon. The final model again contained only age as a predictor, and direction was again negative (indicating that younger participants took less time). The model was statistically significant ($\chi^2(1) = 18.22$, $p < 0.01$) and described 2% of variance of the time taken to find an icon ($Marginal R^2 = 0.02$, $Conditional R^2 = 0.08$).

Finally, we built a model to describe the number of errors made by participants when finding an icon. However, none of the predictors sufficiently described this dependent variable. We were thus unable to create a model for this variable.

4.1.3 Text Entry. The last task for the background music condition is the typing task. We built a model to predict the time per character entry. However, we were unable to create a model for this variable as none of the predictors sufficiently described it.

Following this, we build a model to describe the number of errors made by participants when typing a message. The final model contained one predictor (participant age). Participants of younger age made fewer errors. The model was statistically significant ($\chi^2(1) = 8.55$, $p < 0.01$) and described 12% of variance of the number of uncorrected errors ($Marginal R^2 = 0.12$, $Conditional R^2 = 0.31$).

4.2 Results: Urban noise condition

The tested urban noise conditions are outdoor, indoor, and silence. The conditions are tested across all three defined tasks. We discuss the results per task below.

4.2.1 Target Acquisition. We first built a model to describe the time taken to tap a circle. The final model was described with three predictors: condition, Y centre coordinate of the circle, and participant age. We provide a summary of the factors in Table 3. The model was statistically significant ($\chi^2(4) = 54.73$, $p < 0.01$) and described 8% of variance of the time taken to tap a circle ($Marginal R^2 = 0.08$, $Conditional R^2 = 0.27$). Participants took significantly less time to tap a circle in both indoor and outdoor noise conditions, with the stronger effect of urban outdoor noise. Wilcoxon signed-rank test did not show a significant difference between the indoor and outdoor noise conditions for time taken to tap circle ($V = 1.26e+06$, $p = 0.81$). We visualised mean values for the time taken to tap a circle per condition in Figure 4. Table 3 shows that older participants required more time to tap a circle, similar to our observation in the music conditions. Furthermore, we show that the further from the top left corner the target is, the longer the target acquisition time is.

Table 3. Effects of model factors on predicting time taken to tap a circle in urban noise condition

	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	4.68e+02	4.03e+01	11.61	6.42e-11 ***
Condition (Indoor noise)	-7.89	3.09	-2.55	0.011 *
Condition (Outdoor noise)	-19.4	3.12	-6.21	5.65e-10 ***
Y centre coordinate	6.33e-03	2.37e-03	2.67	0.007 **
Age	3.85	1.22	3.14	0.005 **

Significance: '***' <0.001, '**' <0.01, '*' <0.05

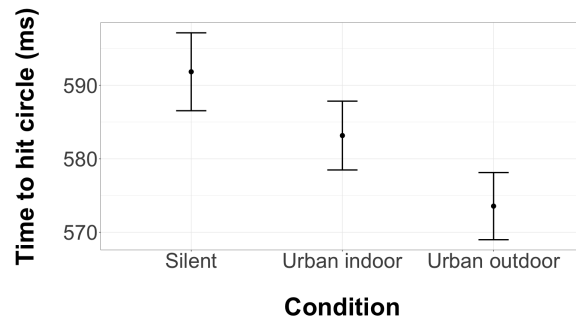


Fig. 4. Mean values for time taken to hit a circle per condition

We then built a model to predict the offset size. The final model contained the X centre coordinate of a circle as its only predictor. The direction of the predictor is negative, indicating that the closer to the right edge of the screen the circle is located, the smaller the offset of the touch is. The model was statistically significant ($\chi^2(1) = 33.77$, $p < 0.01$) describing 0.5% of variance of the offset size ($Marginal R^2 = 0.005$, $Conditional R^2 = 0.07$).

4.2.2 Visual Search. To check the effect of the urban noise conditions on performance during icon search task, we built a model to predict the time taken to memorise an icon. The final model contained two predictors: condition and participant age. The model was significant ($\chi^2(3) = 21.33, p < 0.01$) and described 8% of variance of time taken to memorise an icon (*Marginal* $R^2 = 0.08$, *Conditional* $R^2 = 0.22$). The coefficients of the predictors are summarised in Table 4. From the table, we observe that participants spent significantly less time memorising an icon in the urban indoor noise condition compared to the silent condition. However, the effect of urban outdoor condition did not significantly affect the time taken to memorise an icon. Mean values for time taken to memorise an icon per condition can be found in Figure 5.

Table 4. Effects of model factors on predicting time taken to memorise an icon in urban noise condition

	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	510.38	69.36	7.36	2.01e-07 ***
Condition (Indoor noise)	-36.48	12.33	-2.96	0.003 **
Condition (Outdoor noise)	-5.18	12.32	-0.42	0.67
Age	7.55	2.10	3.593	0.002 **

Significance: ‘***’ <0.001, ‘**’ <0.01, ‘*’ <0.05

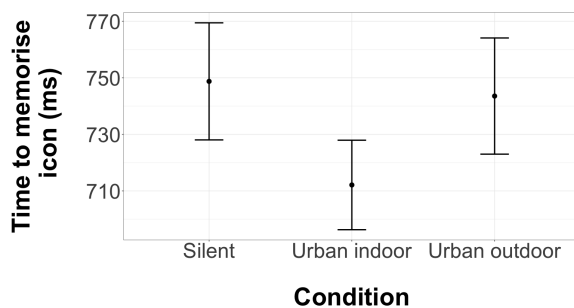


Fig. 5. Mean values for time taken to memorise an icon per condition

Next, we built a model to describe the total time taken to find an icon in the urban noise conditions. The final model contained the X centre coordinate of the icon as its only predictor. The predictor was again negative, indicating that the icons were quicker to find the closer they were located to the right edge of the screen. The model was statistically significant ($\chi^2(1) = 9.32, p < 0.01$) and described 0.5% of variance of time taken to find an icon (*Marginal* $R^2 = 0.005$, *Conditional* $R^2 = 0.09$).

Finally, we built a model to describe the number of errors made by participants in the finding an icon task. The model was described with three predictors; condition and both the X and Y centre coordinates of the icons. The model was statistically significant ($\chi^2(4) = 16.45, p < 0.01$) and explained 1% of variance of the number of errors participants performed during icon search task (*Marginal* $R^2 = 0.01$, *Conditional* $R^2 = 0.04$). The model is summarised in Table 5.

Table 5. Effects of model factors on predicting errors in finding an icon in urban noise condition

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.89e-02	1.63e-02	1.16	0.25
Condition (Indoor noise)	6.44e-03	1.19e-02	0.54	0.59
Condition (Outdoor noise)	2.60e-02	1.19e-02	2.18	0.03 *
X centre coordinate	-4.42e-05	1.71e-05	-2.59	<0.01 **
Y centre coordinate	2.08e-05	9.90e-06	2.10	0.04 *

Significance: '***' <0.001, '**' <0.01, '*' <0.05

The mean values for the number of errors in the icon search task are visualised in Figure 6. According to our findings, participants were significantly less accurate in the urban outdoor noise condition when compared to silent condition. The effect of X and Y centre coordinates is again in line with our previous findings.

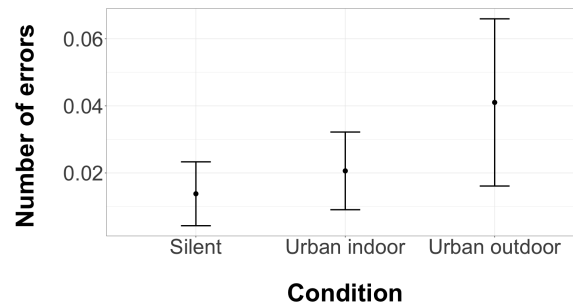


Fig. 6. Mean values for errors in find an icon task per condition

4.2.3 Text Entry. We then built a model describing the time required per character entry. The final prediction model contained both condition and gender as predictive variables. The model was statistically significant ($\chi^2(3) = 10.79$, $p = 0.01$) and explained 14% of variance of the time per character entry participants spent during the typing task ($Marginal R^2 = 0.14$, $Conditional R^2 = 0.49$). We presented the summarised model in Table 6. We then visualised the values of the mean time per character entry in a boxplot in Figure 7. Finally, we built a model to predict the number of errors made by participants during the typing task. None of the included predictors sufficiently describe this variable.

Table 6. Effects of model factors on predicting time per character entry in a typing task in urban noise condition

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	376.59	47.16	7.99	5.17e-09 ***
Condition (Indoor noise)	26.57	32.67	0.81	0.42
Condition (Outdoor noise)	69.50	32.67	2.13	0.04 *
Gender (M)	155.59	61.13	2.55	0.02 *

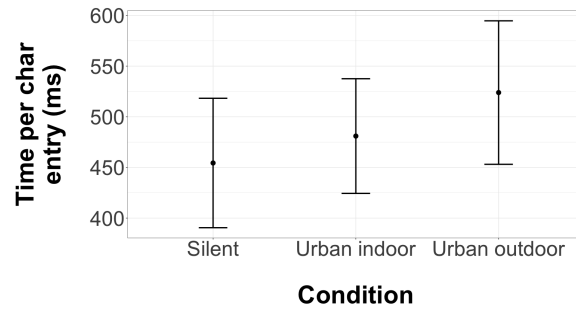


Fig. 7. Mean values for time per character entry in urban noise condition

4.3 Results: Speech condition

4.3.1 Target Acquisition. First, we built a model to describe the time taken to tap a circle. The final model contained the Y centre coordinate of the circle and participant age as its predictors. The relationship of the Y centre coordinate was positive, indicating that the lower the circle was from the top left corner of the screen, the quicker it was to tap. For age, we see again the same trend: young participants were quicker in tapping the circles. The model was statistically significant ($\chi^2(1) = 105.39, p < 0.01$) describing 7.6% of variance of time taken to tap a circle (*Marginal R*² = 0.08, *Conditional R*² = 0.26).

We then built a model to describe the offset size of the touch. The final model contained only the X centre coordinate of the circle as a factor. As before, the predictor was negative, indicating that the icons were quicker to find the closer they were located to the right edge of the screen. The model was statistically significant ($\chi^2(1) = 41.85, p < 0.01$) and described 0.6% of the variance of the offset size (*Marginal R*² = 0.006, *Conditional R*² = 0.06).

4.3.2 Visual Search. We created a model to describe the time taken to memorise an icon. The final model contained only participant age as a predictor, again in the same direction - indicating that young participants were quicker to find the icons. The model was statistically significant ($\chi^2(6) = 12.31, p < 0.01$) and explained 7% of variance of time taken to memorise an icon (*Marginal R*² = 0.07, *Conditional R*² = 0.17).

Finally, we construct a model to describe the time taken to find an icon and the corresponding number of errors. However, none of the predictors sufficiently described this dependent variable.

4.3.3 Text Entry. To investigate the effect of speech condition on a typing task, we built a model describing the time needed per character entry. The model was statistically significant ($\chi^2(3) = 9.30, p < 0.03$) and described 9% of variance of time taken to type a character (*Marginal R*² = 0.09, *Conditional R*² = 0.40). Coefficients of the model are summarised in Table 7. As can be seen from Table 7, participants were significantly slower when typing a character under meaningful speech condition, compared to silent condition. We drew a boxplot to visualise the values of the mean time per character entry for each condition in Figure 8.

We then built a model describing the number of uncorrected errors in a typing task. However, none of the predictors were descriptive enough to predict the number of uncorrected errors during the typing task under the speech condition.

Table 7. Effects of model factors on predicting time per character entry in speech condition

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	396.01	47.56	8.33	1.12e-09 ***
Condition (Speech meaningful)	85.44	36.53	2.34	0.02 *
Condition (Speech meaningless)	53.09	36.76.20	1.44	0.15
Gender (M)	116.75	60.31	1.94	0.07

Significance: '***' <0.001, '**' <0.01, '*' <0.05

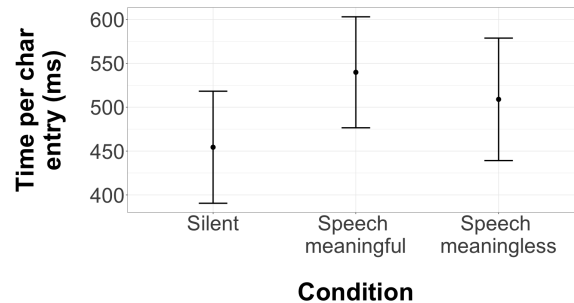


Fig. 8. Time per character entry in speech condition

4.4 Results: Interviews

The qualitative analysis presented in this section reflects on participants' answers to the short interview session that we held after they finished the experiment. Several participants (4 out of 24) mentioned that they felt their "overall performance was worse towards the end of the experiment". In contrast to this opinion, 3 out of 24 participants mentioned that their performance was better towards the end of the experiment as they were "more used to the tasks".

4.4.1 Perceived task performance in music condition. Participants felt the effect of music was most prominent in the tapping task. 12 out of 24 participants claimed that they followed the tempo of the music when tapping circles. The most popular comment we received from the participants was "I was tapping circles in rhythm with the music tempo". Some of the participants thought that the music improved their performance: "Classical music improved my performance" (P24), "I was performing better in finding icons under both music conditions" (P21). Interestingly, participants mentioned that in both of the music conditions they tapped on circles faster when compared to the silent condition: "I was fast in both music conditions. But even with the slow music I was performing faster than in silence" (P14, P20, P24). The participants also mentioned that the music made their task completion experience more enjoyable: "Music made the task easier psychologically, I was not stressed or worried" (P01), "Slow music was relaxing" (P06), "Clicking the circles was fun with the music" (P14), "I was enjoying the music and focused on the task better" (P24). Overall, interview remarks demonstrate that the participants noticed the influence of music tempo on completing target acquisition task. However, there were exclusive remarks from our participants who believed that music was annoying and distracting (P08, P10, P17) and, hence, might have deteriorated their performance: "Slow music was annoying and distracting. I could not concentrate on finding an icon and was slower when typing a message" (P10). However, upon checking the quantitative data we found that mean time taken to

find an icon on the visual search task during the slow music condition for participants P08 ($M = 1166.21$ ms, $SD = 257.99$) and P17 ($M = 1359.12$ ms, $SD = 462.27$) was less than the mean time for all participants ($M = 1753.19$ ms, $SD = 1166.21$). Only participant P10 took longer time finding an icon under the slow music condition ($M = 1860.92$ ms, $SD = 1098.26$) compared to the total mean time for the visual search task. We also found that all three participants took less time per character entry in the text entry task ($M = 431.56$ ms, $SD = 59.26$; $M = 379.44$ ms, $SD = 67.38$; and $M = 424.65$ ms, $SD = 72.02$ for participants P08, P10, P17 respectively) compared to the mean time of all participants ($M = 485.14$ ms, $SD = 192.48$). This shows that participants' perception does not always align with their actual performance.

4.4.2 Perceived task performance in urban noise condition. There was a lack of consensus in the remarks regarding the urban noise condition. More than half of the participants agreed that they were annoyed and distracted with the urban outdoor noise (13 out of 24). Several participants mentioned that both urban noise conditions were distracting. Five participants complained about the outdoor urban noise condition: *"In urban outdoor noise I was stressed and alert. I had to correct the texts a lot under outdoor urban noise condition"* (P01), *"Urban outdoor noise was distracting"* (P11, P14, P19), *"I made several mistakes when finding icons in outdoor noise condition"* (P19). One participant mentioned that they performed quicker in all of the tasks during the outdoor noise as they tried to *"tune out the noise"* (P06). Meanwhile other participants complained about urban indoor noise condition: *"Ambient indoor noise was distracting, as it made feel like I was in a cafe and I felt as if I needed to be involved in a conversation with a friend"* (P06), *"Indoor noise put me almost asleep"* (P13). Surprisingly, one participant claimed that they liked the urban outdoor noise condition and it did not affect their performance: *"Traffic noise was good, I am used to it, I was quicker in tapping circles and typing tasks."* (P13). These comments are mostly in line with our quantitative findings, as our results show that participants took less time in tapping circles and memorising an icon under both indoor and outdoor urban conditions. However, time per character entry was higher for these conditions.

4.4.3 Perceived task performance in speech condition. Regarding the speech condition, 13 out of 24 participants found the speech condition to be distracting and claimed their overall task performance was worse in both of the speech conditions. The participants emphasised that speech condition deteriorated their performance particularly in visual search and text entry tasks. Some of the insights from the participants are: *"English speech was distracting. When listening to it I made more mistakes while typing"* (P13, P14, P17, P22, P23), *"I made several mistakes when finding icons in English speech condition"* (P19). Regarding the target acquisition task only a few participants felt that speech affected their performance. For example, one participant said that they were slower in tapping circles under English speech condition but could not explain the reason behind (P01). However, another participant mentioned that they tapped on circles quicker to avoid "listening to the foreign language" (P06). When we examined their data, we found that P01 did in fact take less time ($M = 574.42$ ms, $SD = 112.69$) tapping circles in English speech condition compared to the mean time of all participants ($M = 592.81$ ms, $SD = 119.60$). The quantitative data of the participant P06 is in line with their comment, as they took less time tapping circles ($M = 542.75$ ms, $SD = 117.19$) compared to the total mean time taken to tap a circle ($M = 609.38$, $SD = 132.66$). Overall, the participants claimed that the speech condition affected their performance in visual search and text entry tasks.

4.5 Summary of Results

We summarise the effects of each condition on the different dependent variables and present them in Table 8. As can be seen from the table, the target acquisition time decreased in both music conditions, as well as in both urban noise conditions. Offset size was significantly larger in the music with slow tempo condition. Further, participants spent significantly less time on memorising an icon in the urban indoor noise condition. In addition,

participants produced significantly more errors in urban outdoor noise condition when searching for an icon. Finally, the time per character entry was significantly longer in the urban outdoor noise and meaningful speech conditions when compared to the silent condition.

Table 8. Table summarising the effect of ambient noise conditions on predicted variables compared to silent condition

	Music		Urban noise		Speech	
	Slow Tempo	Fast Tempo	Indoor	Outdoor	Meaningful	Meaningless
Time to tap a circle	↓*	↓*	↓*	↓*	-	-
Offset size	↑*	↑	-	-	-	-
Time to memorise an icon	-	-	↓*	↓	-	-
Time to find an icon	-	-	-	-		
Errors in icon search			↑	↑*		
Time per character entry			↑	↑*	↑*	↑
Errors in typing	-	-				

‘↓’ – decreased, ‘↑’ – increased (relative to baseline)

‘*’ – the effect was statistically significant, ‘-’ – no effect was observed

Empty cells indicate a failed attempt to describe the variable with provided factors

4.6 Comparison of cold- and ambient noise-induced situational impairments

In this section we compare our findings to the effect of cold ambience on mobile interaction. We compare our mean values for time taken to tap a circle, offset size of the tap, time taken to memorise an icon, and the time taken to find an icon with the values presented in Sarsenbayeva *et al.* [45] and summarise them in Table 9. The comparison suggests that the effect of cold was more pronounced than all our ambient noise conditions in terms of the time taken to memorise an icon and time taken to find an icon. The effect of cold on offset size was similar for the fast tempo music, slow tempo music and meaningful speech conditions (approx. 43 pixels). In terms of time to tap a circle, the effect of cold was larger than the conditions of fast tempo music, slow tempo music, urban indoor noise, urban outdoor noise, and meaningful speech conditions, except for the meaningless speech condition. The mean values for the silent condition are slightly smaller for offset size, time to memorise an icon, and time to find an icon compared to the warm condition. Meanwhile, the time to tap a circle is approximately equal for both of the baseline conditions.

Table 9. Comparison of the effect of situational impairments against the baseline

Mean (SD)	Baseline		Situational impairments						Speech (Meaning- ful)	Speech (Meaning- less)
	Warm	Silent	Cold	Music Fast	Music Slow	Urban Indoor	Urban Outdoor			
Time to tap a circle, ms	593 (137.89)	591.84 (125.07)	603 (144.62)	577.76 (135.85)	574.15 (123.45)	583.16 (117.38)	573.56 (110.39)	592.81 (119.60)	609.38 (132.66)	
Offset size, px	41.34 (40.89)	39.38 (35.26)	42.66 (33.04)	42.95 (29.72)	42.11 (37.48)	41.06 (32.70)	40.91 (38.47)	42.90 (49.86)	39.47 (30.55)	
Time to memorise an icon, ms	815 (150.15)	748.76 (253.89)	854 (196.47)	737.71 (233.54)	745.15 (294.29)	712.11 (194.40)	743.56 (253.03)	753.20 (319.57)	738.90 (254.39)	
Time to find an icon, ms	1632.24 (1235.27)	1587.74 (871.95)	1942.46 (2750.85)	1564.70 (830.81)	1753.19 (2263.27)	1543.12 (700.54)	1633.15 (1191.98)	1637.54 (1072.61)	1520.99 (767.07)	

5 DISCUSSION

5.1 Effect of ambient noise on smartphone interaction

A large body of scientific work highlights the effect of ambient noise on human behaviour [22, 39], cognitive performance [5, 44, 54], and emotional state [25]. Our findings show that participants were quicker in the target acquisition task in both of the fast and slow music conditions. Our results are partly in line with the literature, as it has been shown that fast tempo music accelerates human performance in drinking [33], eating [9], or walking [34]. However, previous work has not identified a positive effect of slow tempo music on performance. In our qualitative data, participants mentioned that music in general helped them get into the rhythm of the task, which can explain why both music conditions led to a decrease in time taken to tap a circle. However, a faster completion of the task also resulted in a larger offset size. This agrees with previous work that demonstrated that while music increases the task performance speed, it also reduces overall accuracy [12]. As for the remaining tasks (visual search and text entry), we did not observe an effect of music on performance.

Another important factor to consider is that, even though literature has identified tempo as the main factor influencing human performance when listening to music [34], different types of music (*e.g.*, vocal music, rock) can also have a different effect on mobile interaction. For instance, Wen *et al.* demonstrated that performance in recall is significantly higher under classical music in contrast to rock music [55]. In our study, we manipulated the tempo of a classical music piece. While both slow and fast tempo classical music affected performance in a similar way, further research is needed to investigate the effect of a multitude of different music genres, a wider tempo range, and other musical elements such as pitch and timbre on mobile interaction.

Our results demonstrate an overall positive trend regarding the effect of urban noise on the target acquisition time and the time taken to memorise an icon. However, urban noise was negatively perceived by participants, with many participants commenting on its distracting nature. As such, there was an incentive for participants to perform quicker in these two tasks in order to reduce time spent in this unpleasant condition. However, in a real-world scenario a user would not be able to “escape” these unpleasant background noises by simply completing a task on their mobile device. Moreover, under the urban outdoor noise condition, participants made significantly more errors when finding an icon and took longer to type each character. These results are in line

with the findings presented in previous work [5, 22] and show the negative effect of urban noise as reported by participants in the interviews.

The effect of speech was limited to the text entry task. Our findings show that participants took significantly longer to type a text when listening to meaningful speech. These results correspond with previous work where meaningful speech was shown to have a negative effect on cognitive performance [4, 27, 50, 54]. Participants mentioned that they were listening to the English speech (meaningful), which resulted in longer completion times when typing a text. The effect of meaningless speech was much smaller. As participants did not understand the Kazakh speech (meaningless), it was likely easier for participants to ignore the spoken text.

It is important to note that factors outside the ambient noise conditions also affected participants' performance during the tasks. For example, in the target acquisition task, the Y centre coordinate of the circle played a significant role for predicting the time to tap a circle. Furthermore, the X centre coordinate of the circle significantly affected the offset size of the tap. The offset size of touch was smaller for the targets located closer to the right edge of the screen. These findings are in line with previous work which showed that screen coordinates have a significant effect on time to tap as well as tap accuracy [20, 45]. Another common factor influencing results was age, influencing performance for time taken to tap a circle, to memorise an icon, and to find an icon. This is in line with previous work which shows that age is an important factor affecting both memory and errors during time based tasks [42].

5.2 Contrasting situational impairments in mobile interaction

In addition to ambient noise, cold ambience has been identified as a situational impairment that can affect mobile interaction. By comparing these situational impairments to each other, we were able to establish a benchmark measurement and determine what strategy should be prioritised and in which situations. To allow for a fair comparison between previously reported situational impairments, we used the same target acquisition and visual search tasks as presented in Sarsenbayeva *et al.* [45]. We then compared the results of our participants against the results obtained by those authors, in which participants were exposed to cold as a source of situational impairment.

Our results show that the mean values for time taken to tap a circle, memorise an icon and find an icon were larger in a cold environment compared to the ambient noise conditions. Cold ambience affects people physiologically [7], with a decrease in body temperature when exposed to cold — albeit at different rates per individual. As a result, the cold affects the dexterity of the fingers and leads to deteriorated fine-motor performance. Our comparison shows that the effect of ambient noise is more nuanced. Some people are more accustomed to certain types of noise than others, and the effect of these noises can be in opposite directions between people. For instance, people with a certain preference in music might be less tolerant to other types of music. These preferential differences between people cause the effect of noise to be less homogeneous than situational impairments that affect everyone similarly (cold reduces our body temperature).

Previous work has highlighted the need for accumulated knowledge in the HCI/UbiComp community [30, 31]. Liu *et al.* [31] argue that accumulated knowledge contributes to the formation of important research themes in the field. In our work we have obtained new results that are directly comparable to a previous study, and therefore allow us to benchmark mobile interaction under different situational impairments.

5.3 Detecting and accounting for noise-induced situational impairments

Our findings demonstrate that mobile interaction can be affected by ambient noise in certain situations. Previous work has argued that detecting situational impairments is a fundamental step towards the successful adaptation of mobile interfaces [46]. The automatic detection of situational impairments during mobile interaction opens the way to appropriate interface adaptations; thus, enabling the interaction with mobile devices to be more

appropriate to the user's context. Furthermore, solutions should ideally leverage the built-in sensors of the mobile device [47] as opposed to requiring additional instrumentation (e.g., external temperature and humidity sensors to receive climatic parameters [1]). Following this suggestion, we argue that the built-in microphone of mobile devices can be used to detect noise-induced situational impairments, as previously suggested by Kanjo *et al.* [27]. In addition, a classifier could be used to distinguish between different noise types, as our results show that they can have different effects. Smartphone interfaces would then accommodate accordingly based on the condition detected. For example, if exposure to urban noise or speech is detected (shown to adversely affect text entry) while a user is typing, the phone can present an alternative interface to mitigate the effect of potential noise-induced situational impairments (e.g., "WalkType" interface by Goel *et al.* [14], "Fat thumb" technique by Boring *et al.* [8]).

5.4 Limitations

Our study had several limitations. First, the study settings were strictly controlled. It is possible that in a naturalistic environment ambient noise types would be more diverse, and multiple noise types could occur simultaneously (e.g., urban outdoor noise might contain music in a touristic part of the city). However, controlling for individual noise types was necessary as our goal was to systematically compare the effect of specific ambient noises on performance. Second, we were limited in the number of sounds included and did not include other types of music (e.g., varying music genres, music with lyrics), urban noise (e.g., markets, public performances), or speech (e.g., second language of the participant). These are potential research directions that can be explored in future studies. We also did not run our study under different volume levels, but kept the noise level constant at between 55-60 dB to avoid participant discomfort.

Finally, we restricted our participants to use only their index finger to interact with the smartphone. We argue that this restriction was necessary to draw a fair comparison between the effect of ambient noise and cold ambience [45]. Moreover, by restricting the interaction technique we created a more comparable setting between participants.

6 CONCLUSION

In this study we investigate the effect of ambient noise on mobile interaction performance in target acquisition, visual search, and text entry tasks. We found that participants were significantly quicker in completing the target acquisition task in music conditions (both fast and slow) compared to the silent condition. However, they were significantly less accurate while listening to slow music. During the visual search task, participants took significantly less time to memorise an icon while listening to urban noise, but made more errors when finding an icon under urban outdoor noise. Participant performance during the text entry task was significantly affected by the urban outdoor noise and meaningful speech conditions. The comparison of cold-induced and noise-induced situational impairments on mobile interaction showed that the effect of cold ambience was more prominent on tasks requiring fine-motor movements, whereas the effect of ambient noise was more prominent on tasks requiring cognitive skills. Our findings enhance the understanding of noise-induced situational impairments on mobile interaction and contribute towards accumulating knowledge in situational impairments research. Furthermore, detection of ambient noise and noise-induced situational impairments may be used for sensing user context and adapting the interface accordingly to mitigate their effect on mobile interaction.

ACKNOWLEDGMENTS

This work is partially funded by the Academy of Finland (Grants 285459-iSCIENCE), the European Commission (Grant 6AIKA-A71143-AKAI), and Marie Skłodowska-Curie Actions (645706-GRAGE).

REFERENCES

- [1] S. Aram, A. Troiano, and E. Pasero. 2012. Environment sensing using smartphone. In *2012 IEEE Sensors Applications Symposium Proceedings*. 1–4. <https://doi.org/10.1109/SAS.2012.6166275>
- [2] Amy FT Arnsten and Patricia S Goldman-Rakic. 1998. Noise stress impairs prefrontal cortical cognitive function in monkeys: evidence for a hyperdopaminergic mechanism. *Archives of general psychiatry* 55, 4 (1998), 362–368.
- [3] Alan Baddeley and Pierre Salame. 1986. The unattended speech effect: perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12, 4 (1986), 525.
- [4] Simon Banbury and Dianne C Berry. 1997. Habituation and dishabituation to speech and office noise. *Journal of Experimental Psychology: Applied* 3, 3 (1997), 181.
- [5] Simon Banbury and Dianne C Berry. 1998. Disruption of office-related tasks by speech and office noise. *British journal of psychology* 89, 3 (1998), 499–517.
- [6] Simon P. Banbury, William J. Macken, Sébastien Tremblay, and Dylan M. Jones. 2001. Auditory Distraction and Short-Term Memory: Phenomena and Practical Implications. *Human Factors* 43, 1 (2001), 12–29. <https://doi.org/10.1518/001872001775992462> arXiv:<https://doi.org/10.1518/001872001775992462> PMID: 11474757.
- [7] ML Bartelink, A De Wit, H Wollersheim, A Theeuwes, and Th Thien. 1993. Skin vascular reactivity in healthy subjects: influence of hormonal status. *Journal of Applied Physiology* 74, 2 (1993), 727–732.
- [8] Sebastian Boring, David Ledo, Xiang ‘Anthony’ Chen, Nicolai Marquardt, Anthony Tang, and Saul Greenberg. 2012. The fat thumb: using the thumb’s contact size for single-handed mobile interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. ACM, 39–48.
- [9] Clare Caldwell and Sally A Hibbert. 2002. The influence of music tempo and musical preference on restaurant patrons’ behavior. *Psychology & Marketing* 19, 11 (2002), 895–917.
- [10] Gianna Cassidy and Raymond AR MacDonald. 2007. The effect of background music and background noise on the task performance of introverts and extraverts. *Psychology of Music* 35, 3 (2007), 517–537.
- [11] Sheldon Cohen and Neil Weinstein. 1981. Nonauditory effects of noise on behavior and health. *Journal of Social Issues* 37, 1 (1981), 36–70.
- [12] Brian H Dalton and David G Behm. 2007. Effects of noise and music on human and task performance: A systematic review. *Occupational Ergonomics* 7, 3 (2007), 143–152.
- [13] Anind K. Dey. 2001. Understanding and Using Context. *Personal Ubiquitous Comput.* 5, 1 (Jan. 2001), 4–7. <https://doi.org/10.1007/s007790170019>
- [14] Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: Using Accelerometer Data to Accomodate Situational Impairments in Mobile Touch Screen Text Entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’12)*. ACM, New York, NY, USA, 2687–2696. <https://doi.org/10.1145/2207676.2208662>
- [15] Charles J Golden and Shawna M Freshwater. 1978. Stroop color and word test. (1978).
- [16] Jorge Goncalves, Vassilis Kostakos, Simo Hosio, Evangelos Karapanos, and Olga Lyra. 2013. IncluCity: Using Contextual Cues to Raise Awareness on Environmental Accessibility. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS ’13)*. ACM, New York, NY, USA, Article 17, 8 pages. <https://doi.org/10.1145/2513383.2517030>
- [17] Jorge Goncalves, Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Simo Hosio, Sirkka Risanen, Hannu Rintamäki, and Vassilis Kostakos. 2017. Tapping Task Performance on Smartphones in Cold Temperature. *Interacting with Computers* 29, 3 (2017), 355–367. <https://doi.org/10.1093/iwc/iww029>
- [18] Morgan Harvey and Matthew Pointon. 2018. Noisy Signals: Understanding the Impact of Auditory Distraction on Web Search Tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. ACM, 241–244.
- [19] Mack T Henderson, Anne Crews, and Joan Barlow. 1945. A study of the effect of music distraction on reading efficiency. *Journal of Applied Psychology* 29, 4 (1945), 313.
- [20] Niels Henze, Enrico Rukzio, and Susanne Boll. 2011. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI ’11)*. ACM, New York, NY, USA, 133–142. <https://doi.org/10.1145/2037373.2037395>
- [21] Eve Hoggan, Andrew Crossan, Stephen A Brewster, and Topi Kaaresoja. 2009. Audio or tactile feedback: which modality when?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2253–2256.
- [22] Morris B Holbrook. 1981. Integrating compositional and decompositional analyses to represent the intervening role of perceptions in evaluative judgments. *Journal of Marketing Research* (1981), 13–28.
- [23] Sharon K Holmberg and Sharon Coon. 1999. Ambient sound levels in a state psychiatric hospital. *Archives of psychiatric nursing* 13, 3 (1999), 117–126.
- [24] Makoto Iwanaga and Takako Ito. 2002. Disturbance effect of music on processing of verbal and spatial memories. *Perceptual and motor skills* 94, 3_suppl (2002), 1251–1258.

- [25] Kari Kallinen. 2002. Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. *Computers in Human Behavior* 18, 5 (2002), 537 – 551. [https://doi.org/10.1016/S0747-5632\(02\)00005-5](https://doi.org/10.1016/S0747-5632(02)00005-5)
- [26] Juliane Kämpfe, Peter Sedlmeier, and Frank Renkewitz. 2011. The impact of background music on adult listeners: A meta-analysis. *Psychology of Music* 39, 4 (2011), 424–448.
- [27] Eiman Kanjo. 2010. Noisespy: A real-time mobile phone platform for urban noise monitoring and mapping. *Mobile Networks and Applications* 15, 4 (2010), 562–574.
- [28] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Technical Report. Naval Technical Training Command Millington TN Research Branch.
- [29] Hannu Kukka, Jorge Goncalves, Kai Wang, Tommi Puolamaa, Julien Louis, Mounib Mazouzi, and Leire Roa Barco. 2016. Utilizing Audio Cues to Raise Awareness and Entice Interaction on Public Displays. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 807–811. <https://doi.org/10.1145/2901790.2901856>
- [30] Yong Liu, Jorge Goncalves, Denzil Ferreira, Simo Hosio, and Vassilis Kostakos. 2014. Identity Crisis of Ubicomp?: Mapping 15 Years of the Field's Development and Paradigm Change. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 75–86. <https://doi.org/10.1145/2632048.2632086>
- [31] Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. CHI 1994-2013: Mapping Two Decades of Intellectual Progress Through Co-word Analysis. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3553–3562. <https://doi.org/10.1145/2556288.2556969>
- [32] Randi C Martin, Michael S Wogalter, and Janice G Forlano. 1988. Reading comprehension in the presence of unattended speech and music. *Journal of memory and language* 27, 4 (1988), 382–398.
- [33] Heather McElrea and Lionel Standing. 1992. Fast music causes fast drinking. *Perceptual and Motor skills* 75, 2 (1992), 362–362.
- [34] Ronald E Milliman. 1982. Using background music to affect the behavior of supermarket shoppers. *The journal of Marketing* (1982), 86–91.
- [35] Ronald E Milliman. 1986. The influence of background music on the behavior of restaurant patrons. *Journal of consumer research* 13, 2 (1986), 286–289.
- [36] Sachi Mizobuchi, Mark Chignell, and David Newton. 2005. Mobile Text Entry: Relationship Between Walking Speed and Text Input Task Difficulty. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (MobileHCI '05)*. ACM, New York, NY, USA, 122–128. <https://doi.org/10.1145/1085777.1085798>
- [37] Alexander Ng, Stephen A. Brewster, and John H. Williamson. 2014. Investigating the Effects of Encumbrance on One- and Two- Handed Interactions with Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1981–1990. <https://doi.org/10.1145/2556288.2557312>
- [38] Alexander Ng, John H. Williamson, and Stephen A. Brewster. 2014. Comparing Evaluation Methods for Encumbrance and Walking on Interaction with Touchscreen Mobile Devices. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI '14)*. ACM, New York, NY, USA, 23–32. <https://doi.org/10.1145/2628363.2628382>
- [39] Adrian C North, David J Hargreaves, and Sarah J Heath. 1998. Musical tempo and time perception in a gymnasium. *Psychology of Music* 26, 1 (1998), 78–88.
- [40] Theodore J. Noseworthy and Karen Finlay. 2009. A Comparison of Ambient Casino Sound and Music: Effects on Dissociation and on Perceptions of Elapsed Time While Playing Slot Machines. *Journal of Gambling Studies* 25, 3 (01 Sep 2009), 331–342. <https://doi.org/10.1007/s10899-009-9136-x>
- [41] Michael D Oliver. 1996. *The effect of background music on mood and reading comprehension performance of at-risk college freshmen*. Ph.D. Dissertation. Wayne State University.
- [42] Denise C Park, Christopher Hertzog, Daniel P Kidder, Roger W Morrell, and Christopher B Mayhorn. 1997. Effect of age on event-based and time-based prospective memory. *Psychology and aging* 12, 2 (1997), 314.
- [43] Pierre Salame and Alan Baddeley. 1982. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior* 21, 2 (1982), 150 – 164. [https://doi.org/10.1016/S0022-5371\(82\)90521-7](https://doi.org/10.1016/S0022-5371(82)90521-7)
- [44] Pierre Salame and Alan Baddeley. 1989. Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology* 41, 1 (1989), 107–122.
- [45] Zhanna Sarsenbayeva, Jorge Goncalves, Juan García, Simon Klakegg, Sirkka Rissanen, Hannu Rintamäki, Jari Hannu, and Vassilis Kostakos. 2016. Situational Impairments to Mobile Interaction in Cold Environments. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 85–96. <https://doi.org/10.1145/2971648.2971734>
- [46] Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Vassilis Kostakos, and Jorge Goncalves. 2017. Challenges of Situational Impairments During Interaction with Mobile Devices. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (OZCHI '17)*. ACM, New York, NY, USA, 477–481. <https://doi.org/10.1145/3152771.3156161>
- [47] Zhanna Sarsenbayeva, Niels van Berkel, Aku Visuri, Sirkka Rissanen, Hannu Rintamäki, Vassilis Kostakos, and Jorge Goncalves. 2017. Sensing Cold-Induced Situational Impairments in Mobile Interaction Using Battery Temperature. *Proc. ACM Interact. Mob. Wearable*

- Ubiquitous Technol.* 1, 3, Article 98 (Sept. 2017), 9 pages. <https://doi.org/10.1145/3130963>
- [48] Bastian Schildbach and Enrico Rukzio. 2010. Investigating Selection and Reading Performance on a Mobile Phone While Walking. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, New York, NY, USA, 93–102. <https://doi.org/10.1145/1851600.1851619>
- [49] Andrew Sears, Min Lin, Julie Jacko, and Yan Xiao. 2003. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. In *HCI International*, Vol. 2. 1298–1302.
- [50] Stephen Stansfeld, John Gallacher, Wolfgang Babisch, and Martin Shipley. 1996. Road traffic noise and psychiatric disorder: prospective findings from the Caerphilly Study. *Bmj* 313, 7052 (1996), 266–267.
- [51] Sébastien Tremblay, Alastair P Nicholls, David Alford, and Dylan M Jones. 2000. The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 6 (2000), 1750.
- [52] Kristin Vadas, Nirmal Patel, Kent Lyons, Thad Starner, and Julie Jacko. 2006. Reading On-the-go: A Comparison of Audio and Hand-held Displays. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '06)*. ACM, New York, NY, USA, 219–226. <https://doi.org/10.1145/1152215.1152262>
- [53] Niels van Berkel, Chu Luo, Theodoros Anagnostopoulos, Denzil Ferreira, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2016. A Systematic Assessment of Smartphone Usage Gaps. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4711–4721. <https://doi.org/10.1145/2858036.2858348>
- [54] N Venetjoki, A Kaarlela-Tuomaala, E Keskinen, and V Hongisto. 2006. The effect of speech and speech intelligibility on task performance. *Ergonomics* 49, 11 (2006), 1068–1091.
- [55] We Wen and Kanachi Michihiko. 2006. The effects of music type and volume on short-term memory. *Tohoku Psychol Folia* 64 (2006), 68–76.