Will it Stick? Visual Representations of Push Notifications for Content Retention

Srikanth Aravamuthan
University of Wisconsin – Madison
Madison, WI, USA
aravamuthan@wisc.edu

Brahma S. Pavse
University of Wisconsin – Madison
Madison, WI, USA
pavse@wisc.edu

Zach Potter
University of Wisconsin – Madison
Madison, WI, USA
zmpotter@wisc.edu

ABSTRACT

The purpose of push notifications is to deliver content, while performing a primary task. However, reading and dismissing push notifications can be distracting and interfere with a primary task. Given the ubiquity of push notifications, we hope to better understand push notifications in context of user interruption, user attention, and user perception. We study push notifications from a perspective that has largely been overlooked and under-studied: *content retention*. Specifically, we conducted an experiment to better understand how various visual designs impact content retention of push notifications while performing a primary task. We designed, developed, and deployed a mobile application called Will It Stick Experiment (WISE) [1] to generate custom-built push notifications while the user is performing the primary task of text transcription. We collected quantitative data on ($N = 32$) participants to evaluate performance on a primary task and content retention. We also performed a limited qualitative analysis to understand user perception of the different notification attributes on user interruption and attention. The results suggested that while not all visual designs are equal, some improve the user's ability to retain content without sacrificing performance on the primary task.

Author Keywords

push notifications; visual designs; content retention; mobile applications; user perception

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI);** *Haptic devices;* User studies; Please use the 2012 Classifiers and see this link to embed them in the text: https://dl.acm.org/ccs/ccs_flat.cfm

[1] https://github.com/aravamu2/cs-770-project

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04...$15.00

DOI: https://doi.org/10.1145/3313831.XXXXXXX

INTRODUCTION

Push notifications are important to user experience on mobile devices. The average user is likely to encounter over 50 push notifications on any given day [20]. Push notifications are proficient at keeping users informed on what matters most at a moment's notice, but these timely interactions come at a cost. Typically, push notifications take only a few seconds to read, but they have been shown to negatively affect both cognitive function and concentration [11]. Consequently, push notifications have been shown to hinder the general efficiency of the user and the overall efficacy of the task. While it is not practical to eliminate push notifications [22], remedial measures for such interruptions exist to reduce the impact. Auda et al. studied user interaction with various deferment strategies and Li et al. explored system presentation of notifications immediately or suppress for later [2, 13]. The temporal nature of push notifications about *when* it is presented to users has been well-studied, whereas *how* to present notifications to a user is relatively under-studied. In particular, the question of how visual mappings (e.g. shape, size, and color) could potentially have a significant effect on how efficiently a user retains information provided in a push notification has been rarely studied. Content retention is an important area of push notifications that should be studied for various reasons. First, the sole purpose of push notifications is to convey content. If a notification fails at doing this task, it is failing at its job. Second, it is common for the content of push notifications to go unnoticed while performing some primary task. This issue results in the user having to switch from their primary task to the application that delivered the notification, read the notification, and switch back to the primary task. Naturally, this interruption is distracting and is a waste of time. Motivated by these problems, we believe that studying content retention is a worthwhile pursuit. Moreover, certain visual mappings may also alleviate the negative effects that push notifications tend to induce. We hope to explore whether the modality of push notifications has a noticeable effect on the user retention of information while also mitigating the negative effects of push notifications such as distraction.

RELATED WORK

Consequences of Poor Notification Design

It is imperative to understand how to present notifications to users optimally. Much work has shown that poorly presented notifications can be distracting and lead to workflow inefficiency [7, 9, 10, 19, 12]. Horvitz et al. describe user attention as a valuable resource, but as the amount of online content and interaction increases, then the difficulty utilizing this resource effectively increases as well [7]. They present principles and criteria to follow when deciding when and how to notify a user. They presented a decision-theoretic computational model that automatically determines how critical a message is, and then notify users based on this model. Iqbal et al. examined user behavior as they multi-task on the computer and received notifications, emails, and text messages [9]. They found that users on average spent nearly 10 minutes on switches caused by alerts and took another 10 to 15 minutes to return to a fully focused state on their initial task. Iqbal et al. conducted a field study where they studied the absence of notifications can lead to self-interrupt during the primary task to look for notifications [10]. Users tended to incur the cost of disrupting their work in order to be informed about some event. Okoshi et al. built a system called Attelia to identify breakpoints (recess) while a user is performing some task and delivers notifications only during breakpoints instead of during the task [19]. In general, notifications influence user attentiveness and performance on a primary task. Thus, users are inundated with content and notifications, and it is critical to find the optimal way to balance user attention with notification delivery.

Push Notifications

Generally, understanding *user interruption* by push notifications [13, 23, 16, 1, 4], *user perception* of notification design [18, 14, 5, 3], and *user attention* of notification content [21] has been studied. In practice, these factors were studied and evaluated on specific devices such as desktops [8], smartphones [24, 17, 15], smart TVs [27], and virtual reality systems [6, 4].

Li et al. studied multiple facets of notifications: 1) how user's responded to undesired notifications (based on their notification preferences), 2) what causes a user to interact with a notification, and 3) how to predict a user's notification preferences based on their past actions [13]. Notification preferences were determined based on the following content factors: importance (valuable, good to know, or irrelevant), notification type (text message, ongoing conversation, Facebook, etc.), time sensitivity, and people mentioned. Exler et al. investigated the effect of different notification types on user perception depending on the location of the smartphone (on the table, in the pant pocket, or in the backpack) [5]. Notifications on vibrate and with tones were best perceived regardless of the smartphone position. Notifications on vibrate and with tones were perceived more than notifications where only the LED screen lit up. From the perspective of perceptibility, they suggest that important notifications should use a ringtone and unimportant notifications should use LED lighting. Pielot et al. investigated the effect of content such as new messages, emails, social network updates, and other events on user attention [21]. However, the work used predefined notifications such as the default design of WhatsApp notifications to study user attention. Moreover, they defined user attention using a metric defined by WhatsApp (double ticks on a WhatsApp message). Our work, like these works, focuses on studying notification attributes and how they relate to a user's behavior on a mobile device. However, it is distinct in that we focus on how the *visual* attributes of a notification affect *content retention*.

Visual Design of Push Notifications

In the following sub-section, we expand on the previous work and compare them to our study. Specifically, we are interested in understanding how various visual mappings/attributes of a notification design such as the style, size, color, duration as well as icon reinforcement and haptic feedback affect the user's ability to retain content of the notification. Muller et al. investigated the impact of desktop background on user perception of desktop notifications with respect to size, placement, or opacity [18]. They designed a publicly-available software tool to generate realistic desktop backgrounds. User perception of the notification was measured while the user performs a primary task. They showed that the visual importance of the background at the notification location significantly impacts whether users detect notifications i.e. if the notification and desktop background where the notification was presented do not "agree" the notification would not be detected. Finally, they introduced the concept of a noticeability map on the desktop screen to generate a 2D heat map encoding the high-noticeability regions of the desktop space compared to the low-noticeability regions. Mairena et al. studied the effect of visual features such as color and motion on user perception of desktop notifications at the periphery of desktop screens [14]. Users were presented with various visual targets with different visual features for a short period of time (240ms) and user attention was measured by the user's ability to detect the visual target. Notification distraction was measured by the user's performance on a primary task (in this case, playing a video game). They showed that the noticeability of feature combinations was approximately equal to the better of the individual features. Like this work, we also study how different combinations of the tested attributes impact content retention. Our study is distinct from the previous works such that we extend the design and analysis of push notifications from noticeability to content retention during task performance on mobile devices. Bahir et al. showed visual elements of a notification can improve the receptivity of the notification and user response [3]. In particular, they found that users were more likely to react to notifications that include an icon or image compared to standard notifications. They also found that users were more likely to react to notifications that include an action button with the notification compared to those excluding such a button. Like this work, we also study the impact of icon reinforcement, but in the context of content retention. Moreover, we study how icon reinforcement interacts with other tested visual attributes. In general, our work is different from past work in that we focus on understanding the effect of notification attributes such as style, size, color, duration, icon reinforcement, and haptic feedback on content retention while performing a task on a mobile device. While the space of notification design is extensive, the implications of different notification attributes on content retention have been often overlooked.

METHODS

Design/System

We designed, developed, and deployed a mobile application to generate custom push notifications that pop up on the mobile device using React Native while the user transcribes a prompt in a text-entry box to simulate the task of composing an email or sending a text message. We provided the participant with a printout of text to transcribe in the text-entry box of the application. This serves as the primary task. The participant transcribes the prompt in the text-entry box as push notifications are periodically presented on the mobile device. We instructed the participant to pay attention to the push notifications while performing the primary task.

Two prompts were selected corresponding to the two phases of the experiment. Both prompts were chosen from a library of English typing test paragraphs such that each prompt was a similar level of difficulty and the same number of words [26]. The level of difficulty depends on the average word length and special characters such as capitals, numbers, and symbols.

The app was implemented on an iPhone 13 to handle both the control and experimental cases. For the control case, we trigger default iOS push notifications. For our experimental cases, we trigger custom-built push notifications that differ from the default iOS push notifications on five key attributes: Font Family, Background Color, Duration of Visibility, Icon Reinforcement, and Haptic Feedback. We test each of these push notification attributes, each with two levels, for a total of 32 combinations.

Font Family: iOS push notifications are presented with a plain Arial font, so we experiment with an informal font (Chalkduster) and a formal font (Times New Roman).

Background Color: The background colors of iOS push notifications are either light or dark depending on system preferences, so we experiment with yellow and red (typically associated with warnings or errors) backgrounds.

Duration of Visibility: Traditional iOS push notifications, when left alone, stay visible for approximately 7 seconds. We experiment with notifications that disappear faster (after 4 seconds) and slower (after 12 seconds).

Icon Reinforcement: Alongside the content of iOS push notifications is the icon for the app from which the push notification originated. Instead of traditional icons, we experiment with contextually relevant icons (e.g. a food icon for a notification concerning lunch or a corporate logo for a notification concerning an upcoming meeting) and no icon at all.

Haptic Feedback: iOS push notifications are commonly accompanied by vibration. We take the default iOS vibration to be one single pulse, and we experiment specifically with device vibration: repeated vibrations throughout the duration of visibility and no vibration at all.

While the attributes of the push notifications are randomized from participant to participant, the content is standardized for everyone. The content is simple and widely applicable (e.g. reminders and calendar updates). At the conclusion of the testing window, we test the participant on content retention of the push notifications to explore how well they recalled the information.

Quantitative Analysis

Research question

Based on background and related work regarding push notifications and user experience, the study aims to understand how visual effects affect the user's ability to retain content while also performing a task on a mobile device. We ask the following research questions:

RQ1: What main visual mappings are associated with content retention?

RQ2: What interactions of visual mappings are associated with content retention?

RQ3: What are the consequences of visual mappings on the success of a primary task?

Hypotheses

Based on results described by Muller et al. and Mairena et al., background affects the user's ability to distinguish desktop notifications whereas visual effects such as color and motion affect the user's ability to detect notifications [18, 14]. Consequently, we make the following hypotheses:

H1: Content retention will be higher for visual mappings with high contrast such as font family, background color, or duration of visibility compared to the baseline while performing a task on a mobile device. Visual mappings with low contrast such as icon reinforcement or haptic feedback will have no effect on content retention.

H2: Content retention will be higher for visual mappings with high contrast in the presence of other visual mappings or haptic feedback compared to visual mappings or haptic feedback individually.

H3: Primary task performance remains unchanged when an alternative design is presented compared to when a baseline design is presented.

Data Collection

For the experiment, data were collected via a questionnaire by means of Google Forms provided to the participant beforehand for demographic information to allocate individuals into study groups and afterward to evaluate content retention. Additionally, the transcript for each participant and phase was saved and then compared to the prompt to calculate the similarity score between the two documents. For qualitative feedback, open-ended, free-response questions or Likert rating scale were included to elucidate the perception of visual mappings by understanding their experiences with notifications.

Participants

We recruited 32 participants whose ages range from 18 to 30 through convenience sampling via colleagues and classmates. When a potential participant was admitted to the study, we conducted a brief survey to record demographic information and allocated participants accordingly by texting acuity in the appropriate study group.

(a) WISE app (b) Custom-designed notifications

Figure 1. (a) Snapshots of different screens from the WISE app. Left: home screen: toggles control vs experiment and also provide a code for a given session which influences the notifications a participant receives; center: test screen: user is actively typing while receiving push notifications (includes example of a custom-designed push notification); right: example of automation, specifically in calling time. (b) Examples of custom-designed notifications with various visual attributes.

Specifically, participants were assigned to the two different study groups using block randomization for groups of four participants with pair matching based on mobile phone experience from the pre-study questionnaire, namely the use of iOS compared to Android operating system.

The target population of the study is ideally all individuals whose ages range from 18 to 30. The sample population of the study is the attainable undergraduate and graduate students enrolled at the University of Wisconsin-Madison to serve as a proxy for the study population. Additionally, exclusion criteria include individuals without visual impairment or dexterity impediments since the study focuses on visual notifications and given the primary task of text transcription at hand.

Study Design

We conducted 5-way factorial mixed model design with main fixed effects terms for five notification attributes and random effect term for participant, and interaction terms for the five notification attributes where the participant is the level of analysis. The five factors of font family (formal or informal), background color (red or yellow), duration of visibility (short or long), icon reinforcement (broad or specific), and haptic feedback (strong or weak) are within-participant variables and between-participant variables.

The factorial study design includes five notification attributes with two levels each or 32 custom push notification combinations. Given that five push notifications are present for each phase of the study, a minimum of 32 participants was evaluated for 160 custom push notifications or five replicates for each custom push notification combination. Custom push notifications are randomly generated such that all 32 custom push notification combinations are exhausted before a custom push notification is repeated. Moreover, precautions are taken to prevent the same participant from receiving the same custom push notification more than once.

The primary outcome of user attention is content retention where the dependent variable of the retention score is the proxy. Additionally, the secondary outcome of user interruption is success or completion of the primary task where the dependent variable of similarity score between the text entry and prompt is the proxy. The independent variables are the five notification attributes while all other notification attributes are controlled and kept constant. Similarly, the transcription text and notification text are controlled for the given order. Counterbalancing is performed to mitigate carryover effects and minimize practice effects between different orderings of levels. Lastly, the study is single-blinded such that participants are not aware of group assignment until after the experiment.

Study Task

The tasks were conducted within our custom-designed mobile application. The primary task is to transcribe the provided text in the mobile application within a 3-minute time window. The secondary task is to recall information presented via push notifications during the course of the experiment. The primary task is a reasonable representation of the real-world context of interest, namely composing and replying to emails; sending texts or messages to contacts or group conversations; or commenting on social media post such as social networking, video, and music/podcasts. Email continues to retain a place of prominence in the smartphone era; whereas text messaging is the most widely-used smartphone feature among young smartphone owners; while social networking, video consumption, and music/podcasts are especially popular with younger smartphone owners [25]. The custom-designed mobile application is intuitive to use as a participant and easy to control other confounding variables. The measure of content retention via and completion of the primary task is easy to interpret in the context of the investigation. Lastly, the minimal instructions provided to the participant before the experiment and the questionnaire provided to the participant after the experiment is appropriate to perform the secondary task without prior information to focus on notification content instead of completing the primary task.

Study Procedure

The study will follow the following steps:

1. The participant will be given a pre-study questionnaire for demographic information to allocate participants into study groups: first control group and second experimental group or first experimental group and second control group.
2. The participant will be given an iPhone 13 with the custom-built mobile application.
3. Additionally, they will then be given a prompt to type in the mobile application.
4. They will be instructed to transcribe the text within a 3-minute time window and pay attention to the push notifications.
5. While the participant is typing, they will be presented with various push notifications. If the participant is in the control group, they will receive the default iOS push notifications. If they are in the experimental group, they will receive the custom-built push notifications.
6. After completion of the first phase, the participant will be given a new prompt to type in the mobile application.
7. They will be instructed to transcribe the text within a 3-minute time window and pay attention to the push notifications.
8. While the participant is typing, they will be presented with various push notifications under a different set of conditions. If the participant is in the first control group, then they are in the second experimental group and receive the custom-built push notifications. If they are in the first experimental group, then they are in the second control group and receive the default iOS push notifications.
9. We will measure the number of characters for each set of conditions.
10. The participant will be given a post-study questionnaire to evaluate content retention. The questionnaire will be specific to the content. For example, if the notification said “Dinner at Fresh Fin on Tuesday at 5pm", then a prospective question can be “What was the day and time for dinner?" or “What place was dinner?". Lastly, the post-study questionnaire will include questions for qualitative feedback on content retention via self-reflection by means of Likert scale scores and open-ended, free-response questions.

Measures and Analyses

The dependent variable of retention score is used as a proxy for content retention. Specifically, success is determined if the participant is able to recall the push notification during the post-experiment questionnaire given a set of conditions for the push notification. Therefore, logistic regression was used to analyze the notification attributes. Post-hoc analysis using Wald’s test was performed to calculate pairwise comparisons between group levels of the notification attributes with corrections for multiple testing.

Moreover, the dependent variable of similarity score between the text-entry and prompt is used as a proxy for success or completion of the primary task. Thus, linear regression was used to determine the non-inferiority of custom notifications compared to the default notifications. Post-hoc analysis using t-test was performed to calculate pairwise comparisons between group levels of the notification attributes with corrections for multiple testing. Lastly, age, sex, and texting acuity in the pre-experiment questionnaire was used as control variables when pair matching was performed for group assignment and reduce confounding effects.

Furthermore, qualitative analysis was conducted on the Likert score ratings in the post-study questionnaire. The dependent variable of agreement is used as a proxy for user perception of the notification attributes on either task. The Likert rating score was converted to a binomial variable for agreement. An exact, binomial test was used to determine the association of the perceived effect.

Lastly, qualitative analysis was conducted on the open-ended free response questions in the post-study questionnaire using grounded theory via open-coding. The first rater independently developed two types of codes: six general codes and seventeen specific subcodes. The code described high-level themes and the subcode described low-level details observed in the user responses to the open-ended free response questions. The codes and subcodes were provided to a second rater, who then independently performed their open-coding. Finally, Cohen’s kappa was used to measure agreement between the two raters and a z-test was performed to determine significance.

RESULTS

User interruption and attention were measured under different push notifications. Additionally, user perceptions of user interruption and attention were reported with respect to different push notifications.

User Interruption: Text Transcription

User interruption was studied using the text transcription of a provided prompt via similarity score between the two texts. The effect of default iOS push notifications (control) was compared to custom-built push notifications (experiment) on performing the primary task of text transcription. The response variable is the Levenshtein distance (LD) as a measure of the similarity between two strings. The distance is the number of deletions, insertions, or substitutions required to transform the source string or text-entry to the target string or provided prompt. A one-sided, paired t-test for mean difference in similarity scores was used to determine non-inferiority of custom-built push notifications with respect to default iOS push notifications (Figure 2).

There was no significant difference between the control and experiment ($\bar{x}_d = 0.0153$; $t = 0.901$; $p = 0.188$). The difference between the control and the experiment does not exceed the margin of error. Non-inferiority can be inferred where custom-built push notifications perform as well as default iOS push notifications. Note that there exists a pattern or trend between order and performance such that the user performed better on the second attempt than on the first attempt. However, block randomization was used to control for the confounding effect of the order.

User Attention: Information Recall

User attention was explored using information recall of push notifications via retention score on the post-study questionnaire. The effect of the five notification attributes was examined with reference to retaining content from the push notification. The response variable is a binomial outcome where a success is a correct response and a failure is an incorrect response on the post-study questionnaire. A mixed-effects logistic regression model was fit with five fixed-effect terms for each of the five notification attributes and a random-effect term for participant (Figure 3). The alternative level was contrasted with the reference level for each of the five notification attributes and a Wald test for odds ratio was used to determine the significance between the two levels.

There was a significant decrease between formal and informal font family ($\hat{OR} = 0.444$; $\chi^2 = -2.524$; $p = 0.012$) as well as a significant increase between long and short duration of visibility ($\hat{OR} = 2.058$; $\chi^2 = 2.256$; $p = 0.024$). All other odds ratios were not significant. Interestingly, the absence of haptic feedback increased odds of success, but the effect size was not large where the odds ratios was not significant.

Another mixed-effects logistic regression model was fit with the previous five fixed-effect terms for each of the five notification attributes and a random-effect term for participant as well as ten two-way interaction fixed-effect terms for each combination of the five notification attributes (Figure 4). The alternative level was contrasted with the reference level for each of the five notification attributes and the interaction of the two alternative levels was contrasted with the converse for each combination of the five notification attributes.

The main effects term for the odd ratio between red and yellow background color ($\hat{OR} = 9.301$; $\chi^2 = 2.671$; $p = 0.008$) and between long and short duration of visibility ($\hat{OR} = 12.403$; $\chi^2 = 2.930$; $p = 0.003$) were significant. The interaction effects term for the odd ratio between red background color and long duration of visibility ($\hat{OR} = 0.250$; $\chi^2 = -1.970$; $p = 0.049$) and between red background color and presence of haptic feedback ($\hat{OR} = 0.229$; $\chi^2 = -2.096$; $p = 0.036$) were significant. All other odds ratios were not significant. Conversely, the main effects term for the odds ratio of background color and duration of visibility were greater than 1, but the interaction effects term for the odds ratio of background color and duration of visibility was less than 1 implying an antagonistic effect between the notification attributes. Additionally, the odd ratio between formal and informal font family was not significant implying multicolinearity between the notification attributes.

User Perception

User perception was evaluated jointly with user interruption and user attention. First, user perception was investigated with

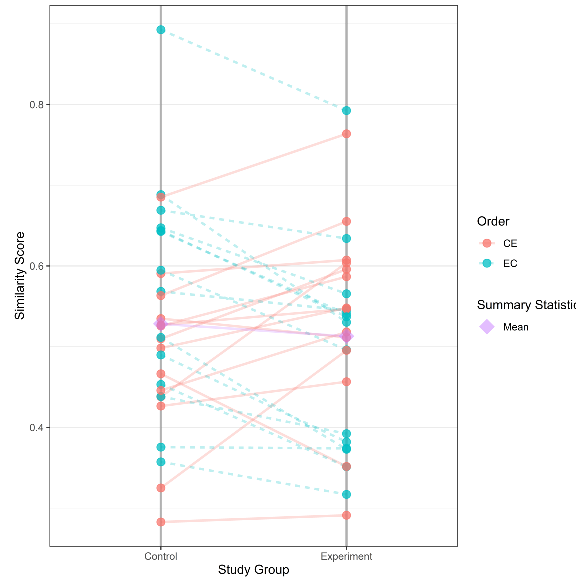

Figure 2. Parallel coordinates chart of similarity score for text transcription between study groups of control and experiment. Red and blue color mapping encode for the order of the paired data and purple color mapping encodes for the mean of the study group.

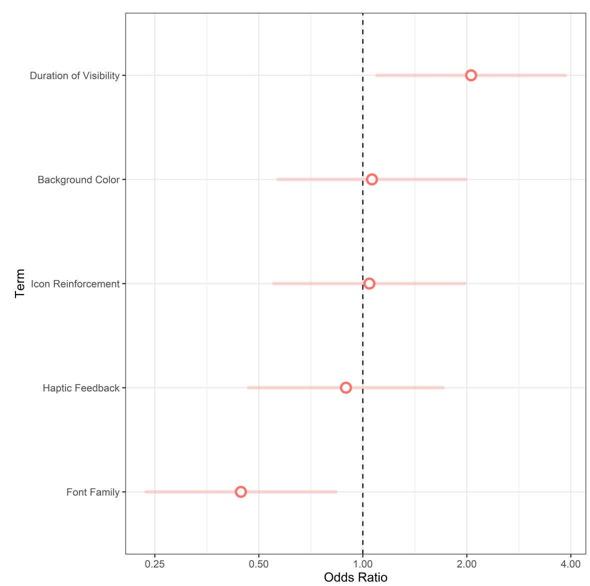

Figure 3. Forest plot of the fixed effects for the reduced mixed-effects logistic regression model. A dashed line at $OR = 1.00$ is used to indicate significance if the 95% confidence interval of the odds ratio for the term contains $OR = 1.00$.

regard to the perceived effects of the five notification attributes on the tasks of text transcription and information recall. The response variable is a Likert rating with five levels converted to a binomial variable for agreement regarding the perceived effect. A one-sided, exact binomial test for agreement was used to determine the association between the five notification attributes and the task of text transcription or information recall (Figure 5).

There was a significant positive association between font family ($\hat{p} = 0.793$; $X = 23$; $p = 0.002$), background color ($\hat{p} = 0.900$; $X = 26$; $p < 0.001$), and haptic feedback ($\hat{p} = 0.828$; $X = 24$; $p < 0.001$) versus the task of text transcription. Moreover, there was a significant positive association between background color ($\hat{p} = 0.724$; $X = 21$; $p = 0.021$) and duration of visibility ($\hat{p} = 0.828$; $X = 24$; $p < 0.001$) versus the task of information recall. Note that only background color was significant for both tasks whereas only icon reinforcement was not significant for both tasks.

Lastly, user perception was reported using feedback via the open-ended free response question with respect to content retention. Open coding was performed by two independent raters to code that represent significant outcomes or attributes in the data (Figure 6) and subcode of the specific concepts (Figure 7) as well as positive, negative, or neutral sentiment. Reliability analysis was performed to measure the extent to which independent coders reach a consensus.

The Cohen's kappa for the general code was 0.717 ($z = 19.1$; $p < 0.001$) interpreted as good to substantial agreement. Similarly, the Cohen's kappa for the subcode was 0.677 ($z = 26.1$; $p < 0.001$) interpreted as good to substantial agreement as well.

Most codes were positive for font family, background color, duration of visibility, and icon reinforcement with similar counts of attributes and outcomes for each of the notification attributes. Most subcodes were labeled as Better Retention followed by the Appropriate Attribute for each of the notification attributes respectively. Note that icon reinforcement demonstrated Better Association and background color displayed High-Contrast. Conversely, most codes were negative for haptic feedback. Furthermore, most subcodes were labeled as Distracting for haptic feedback followed by Inappropriate Vibration and Poor Retention. Interestingly, font family also were labeled as Hard to Read implying both negative effects of inappropriate fonts and positive effects of appropriate fonts.

DISCUSSION

The study was able to validate all three hypotheses to varying levels. Users were able to better retain content for font family and duration of visibility while performing a task on a mobile device (H1). Visual mappings with low contrast such as icon reinforcement or haptic feedback did not have an effect on content retention. Additionally, background color did not have an effect on content retention. Informal font family performed significantly better than formal font family with respect to content retention. Formal font family with sans serif typeface has been shown to better convey information compared to informal font family. This result implies that the novelty of

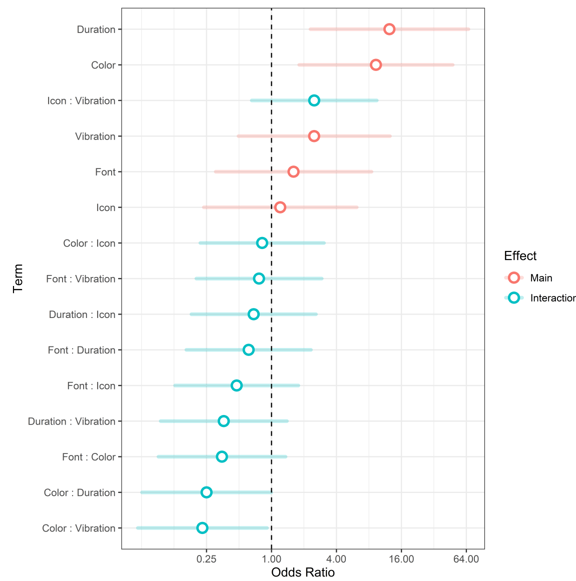

Figure 4. Forest plot of the fixed effects for the extended mixed-effects logistic regression model. A dashed line at $OR = 1.00$ is used to indicate significance if the 95% confidence interval of the odds ratio for the term contains $OR = 1.00$. Red and blue color mapping encode for the main and interaction effects respectively.

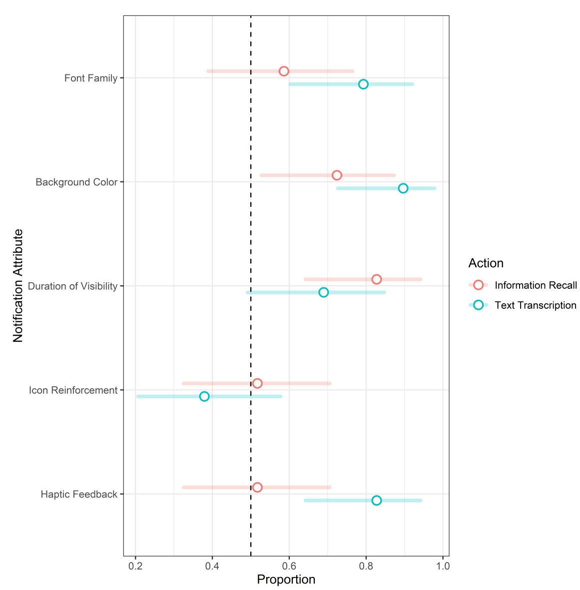

Figure 5. Forest plot of the agreement to the perceived effect. A dashed line at $p = 0.5$ is used to indicate significance if the 95% confidence interval of the proportion for the notification attribute contains $p = 0.5$. Red and blue color mapping encode for the information recall and text transcription respectively.

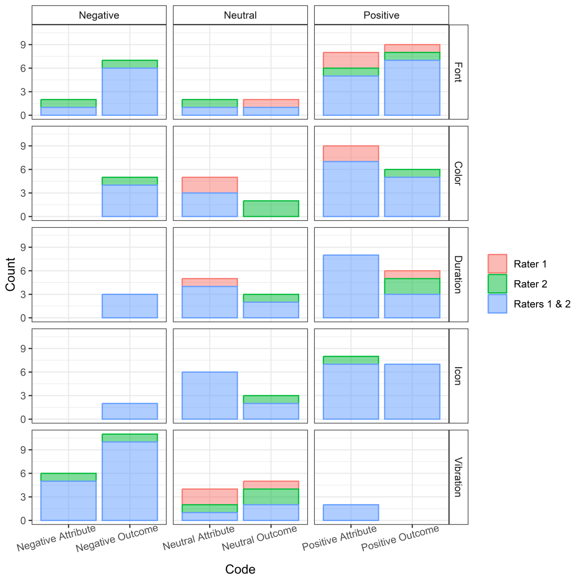

Figure 6. Faceted bar plot of counts for general code of significant outcomes or attributes. Rows encode for notification attributes and columns encode for the sentiment. Red, green, and blue color mapping encode for open coding by rater 1, rater 2, or both raters 1 & 2 respectively.

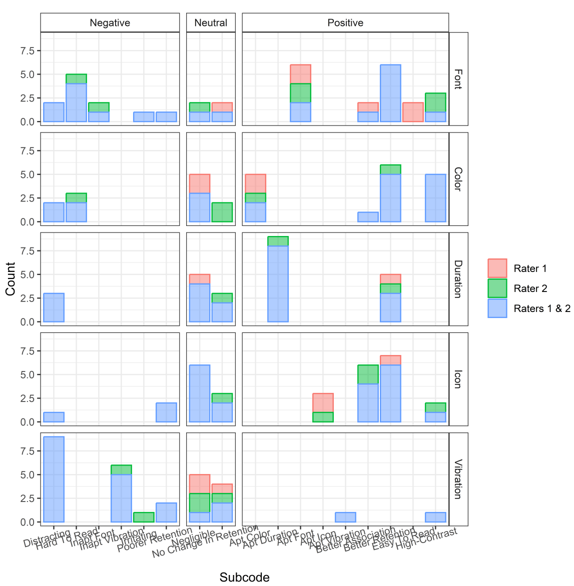

Figure 7. Faceted bar plot of counts for subcodes of the specific concepts. Rows encode for notification attributes and columns encode for the sentiment. Red, green, and blue color mapping encode for open coding by rater 1, rater 2, or both raters 1 & 2 respectively.

the notification attribute can affect content retention as much as the notification attribute itself.

Users were able to better retain content for background color in the presence of duration of visibility or haptic feedback compared to mappings individually (H2). Moreover, font family did not have an effect on content retention in the two-way interaction effects model compared to the main effects model. This result can be explained by multicolinearity where the main effects and interaction effects are correlated in the extended model. Furthermore, the interaction effect for background color and duration of visibility as well as background color and haptic feedback decreased the odds of success compared to individual mappings. This result can be described by an antagonistic effect between the two notification attributes where the overall effect is less than would be expected if the individual effects were added.

Finally, the primary task performance remained unchanged when the custom-built notification design is presented compared to when the default iOS notification design is presented (H3). The experiment performed negligibly worse than the control with respect to a primary task. Therefore, non-inferiority can be concluded where custom-built push notifications perform as well as default iOS push notifications.

The results show that mobile push notifications can be designed to improve content retention. This implies that push notifications can be much more effective in their ability to deliver content to the user. Additionally, no difference in the performance of the primary task implies that notification design can improve content retention without sacrificing performance on the primary task. Collectively, the results suggest the user can better recall information conveyed in the push notifications without negatively impacting their performance on some primary task. In general, we believe that future research should understand the effect of content retention when studying any question related to push notifications.

The study proposes that improved content retention is possible without sacrificing primary task performance. However, there are alternative approaches for further studies. First, we analyzed only binary options for each attribute given the time to design, develop, and deploy the application and recruit participants. Intuitively, users would respond differently to a broad spectrum of notification attributes with a wide array of levels. The study can be expanded to different options and examined on a continuous scale or finer resolution (for example, font color, icon color, and duration of haptic feedback as well as notification size, shape, and location.) The study can be extended to different platforms and explored for scalability and flexibility. Moreover, the test and target population were largely students between the ages of 18-30. It is unknown if the results generalize to other populations (for example, adolescents or senior citizens where user attention would be distinct as well as first responders where content retention would be critical.) Furthermore, the sample size of 32 participants was used to ensure all 2^5 combinations of notification attributes were evaluated for five replicates. For a comprehensive analysis, the sample size for n attributes each with m level would be m^n participants for appropriate power with other conditions remaining the same. Ultimately, user perception of user interruption and user attention should be directly compared to determine if user perception of notification attributes corresponds to differences in retention scores for information recall and the similarity score for text transcription.

CONCLUSION

We explored the design of push notifications on user interruption, user attention, and user perception of notifications attributes by comparing the retention score for information recall and the similarity score for text transcription. We compared 32 participants to determine the effect of font family, background color, duration of visibility, icon reinforcement, and haptic feedback on user performance and perception. Ultimately, the study is the first exploration to provide an all-inclusive approach to user interruption, user attention, and user perception and contribute design insights for future notification delivery and further notification development.

REFERENCES

[1] Christoph Anderson, Judith S. Heinisch, Sandra Ohly, Klaus David, and Veljko Pejovic. 2019. The Impact of Private and Work-Related Smartphone Usage on Interruptibility. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1058–1063. DOI: **http://dx.doi.org/10.1145/3341162.3344845**

[2] Jonas Auda, Dominik Weber, Alexandra Voit, and Stefan Schneegass. 2018. Understanding User Preferences towards Rule-Based Notification Deferral. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. DOI: **http://dx.doi.org/10.1145/3170427.3188688**

[3] Ravit Avraham Bahir, Yisrael Parmet, and Noam Tractinsky. 2019. Effects of Visual Enhancements and Delivery Time on Receptivity of Mobile Push Notifications. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. DOI: **http://dx.doi.org/10.1145/3290607.3312993**

[4] Kuan-Wen Chen, Yung-Ju Chang, and Liwei Chan. 2022. Predicting Opportune Moments to Deliver Notifications in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 186, 18 pages. DOI:**http://dx.doi.org/10.1145/3491102.3517529**

[5] Anja Exler, Christian Dinse, Zeynep Günes, Nadim Hammoud, Steffen Mattes, and Michael Beigl. 2017. Investigating the Perceptibility Different Notification Types on Smartphones Depending on the Smartphone Position. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 970–976. DOI: **http://dx.doi.org/10.1145/3123024.3124560**

[6] Sarthak Ghosh, Lauren Winston, Nishant Panchal, Philippe Kimura-Thollander, Jeff Hotnog, Douglas Cheong, Gabriel Reyes, and Gregory D. Abowd. 2018. NotifiVR: Exploring Interruptions and Notifications in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1447–1456. DOI: **http://dx.doi.org/10.1109/TVCG.2018.2793698**

[7] Eric Horvitz, Andy Jacobs, and David Hovel. 1999. Attention-Sensitive Alerting. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 305–313.

[8] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 93–102. DOI: **http://dx.doi.org/10.1145/1357054.1357070**

[9] Shamsi T. Iqbal and Eric Horvitz. 2007. Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 677–686. DOI: **http://dx.doi.org/10.1145/1240624.1240730**

[10] Shamsi T. Iqbal and Eric Horvitz. 2010. Notifications and Awareness: A Field Study of Alert Usage and Preferences. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 27–30. DOI: **http://dx.doi.org/10.1145/1718918.1718926**

[11] Seul-Kee Kim, So-Yeong Kim, and Hang-Bong Kang. 2016. An Analysis of the Effects of Smartphone Push Notifications on Task Performance with regard to Smartphone Overuse Using ERP. *Computational Intelligence and Neuroscience* 2016 (2016), 1–8. DOI: **http://dx.doi.org/10.1155/2016/5718580**

[12] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. 2016. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1011–1020. DOI: **http://dx.doi.org/10.1145/2858036.2858359**

[13] Tianshi Li, Julia Katherine Haines, Miguel Flores Ruiz de Eguino, Jason I. Hong, and Jeffrey Nichols. 2021. Alert Now or Never: Understanding and Predicting Notification Preferences of Smartphone Users. *ACM Trans. Comput.-Hum. Interact.* (Jul 2021). DOI: http://dx.doi.org/10.1145/3478868 Just Accepted.

[14] Aristides Mairena, Carl Gutwin, and Andy Cockburn. 2019. Peripheral Notifications in Large Displays: Effects of Feature Combination and Task Interference. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. DOI: http://dx.doi.org/10.1145/3290605.3300870

[15] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. PrefMiner: Mining User's Preferences for Intelligent Mobile Notification Management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 1223–1234. DOI: http://dx.doi.org/10.1145/2971648.2971747

[16] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-Driven Intelligent Notification Mechanisms for Mobile Applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 813–824. DOI: http://dx.doi.org/10.1145/2750858.2807544

[17] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. DOI: http://dx.doi.org/10.1145/2858036.2858566

[18] Philipp Müller, Sander Staal, Mihai Bâce, and Andreas Bulling. 2022. Designing for Noticeability: Understanding the Impact of Visual Importance on Desktop Notifications. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 472, 13 pages. DOI: http://dx.doi.org/10.1145/3491102.3501954

[19] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 96–104. DOI: http://dx.doi.org/10.1109/PERCOM.2015.7146515

[20] Martin Pielot, Karen Church, and Rodrigo de Oliveira. 2014a. An In-Situ Study of Mobile Phone Notifications. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 233–242. DOI: http://dx.doi.org/10.1145/2628363.2628364

[21] Martin Pielot, Rodrigo de Oliveira, Haewoon Kwak, and Nuria Oliver. 2014b. Didn't You See My Message? Predicting Attentiveness to Mobile Instant Messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3319–3328. DOI: http://dx.doi.org/10.1145/2556288.2556973

[22] Martin Pielot and Luz Rello. 2017. Productive, Anxious, Lonely: 24 Hours without Push Notifications *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. DOI: http://dx.doi.org/10.1145/3098279.3098526

[23] Benjamin Poppinga, Wilko Heuten, and Susanne Boll. 2014. Sensor-Based Identification of Opportune Moments for Triggering Notifications. *IEEE Pervasive Computing* 13, 1 (2014), 22–29. DOI: http://dx.doi.org/10.1109/MPRV.2014.15

[24] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-Scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. 3055–3064.

[25] Aaron Smith. 2015. U.S. Smartphone Use in 2015. (April 2015). https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/

[26] The Practice Test. 2022. Typing Test Paragraphs. (2022). https://thepracticetest.com/typing/tests/practice-paragraphs/

[27] Dominik Weber, Sven Mayer, Alexandra Voit, Rodrigo Ventura Fierro, and Niels Henze. 2016. Design Guidelines for Notifications on Smart TVs. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '16)*. Association for Computing Machinery, New York, NY, USA, 13–24. DOI: http://dx.doi.org/10.1145/2932206.2932212