A Subjective Evaluation of Multimodal Notifications

David Warnock
Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
Email: warnockd@dcs.gla.ac.uk

Abstract—**The primary users of home care technology often have significant sensory impairments. Multimodal interaction can make home care technology more accessible and appropriate, but most research in the field of multimodal notifications is aimed at office or high-pressure environments instead of the home. Two experiments were carried out that evaluated the subjective workload of responding to visual, auditory, tactile and olfactory notifications (simulating home care reminders) while carrying out a primary task (a card matching memory game). The subjective measurements and observations revealed that participants were open-minded about the possibilities and applications of these modalities, suggesting that home care technology should embrace a much wider range of interaction methods than are currently used.**

Index Terms—**Multi-modal interfaces, accessibility and usability, technology in healthcare**

I. Introduction

Advances in medicine and modern lifestyle choices have endowed us with a greatly increased life expectancy [1]. However, our extended lifespan poses an economic problem; with a larger elderly population and a reduced working population, it will become difficult to afford a high quality care. One way to combat this is to provide assistive technology that can help to maintain a high standard of living and provide independence to those who require care.

A home reminder system is a form of assistive technology designed to help people manage their lifestyle and environment by providing helpful notifications to the user. Notifications could remind users about appointments, meals and medicine times. It could also help users to manage their home by informing them of running taps, unlocked windows and grocery needs. It could even be used to promote healthy social behaviour, such as phoning friends and family members, helping to ensure that the user does not become isolated from the outside world. Such notifications can help to maintain a high standard of living while increasing the self-reliance of the user.

With an increased lifespan we are more likely to develop natural health problems associated with ageing, such as visual and hearing impairments; 80% of people over 60 having a visual impairment and 75% have a hearing impairment [2]. Multimodal interaction could be used to compensate for impairment by delivering notifications via non-impaired channels.

An ideal home reminder system would consider multiple factors such as user activity, social context, user preferences, device availability, reminder urgency and reminder sensitivity when delivering reminders. For example, it should be able to automatically switch from a mobile device to a static device (such as a television) when the user is at home. In a social context, notifications could be delivered discretely via tactile or abstract notifications. If there is lot of background noise, then a user-preferred audio notification could be swapped for a visual one. In addition, older users may find their requirements will change over time; new impairments may develop, or they may begin to suffer from reduced mobility. If a reminder system has a wide range of multimodal delivery methods at its disposal, it will be much better equipped to deal with both evolving requirements and the dynamic, complex environment of the home.

This paper presents a person-centric evaluation of a number of multimodal notification methods using subjective measures, in order to guide the design of multimodal notifications to help make them more effective and appropriate for their environment and their users.

II. Related Work

Wang & Turner [3] identified four 'key properties' that a home care system should always consider: adaptability, personalisation, customisation and dependability. Adaptability refers to a system being able to cope with changing requirements, such as evolving to cope with a user developing a visual impairment. Personalisation refers to creating a system that meets the user's needs, such as compensating for impairments. Customisation means the user can manage and modify the system on their own, such as disabling a notification method that has started to annoy them. Dependability means that the system should not fail the user, which is vital in a system responsible for health and well being.

Wang & Turner's key properties are integral to creating an effective home care system; however, user acceptance is not fully incorporated into their model. User acceptance is vital in a home care system; if the users reject the system, it is likely to significantly reduce its effectiveness. McBryan & Gray [4] describe a hypothetical scenario where notifications are delivered to a user, Fred, via his mobile phone. Televised messages serve as a backup when the phone is unavailable. The phone-based messages irritate Fred when he is at home, so he deliberately turns off his phone to force the notifications to be delivered via the television. The notifications have irritated the user, and as a result he has effectively sabotaged the system and reduced its effectiveness. As a side effect, by turning off his phone the user has become more socially isolated.

It is important to understand what makes notifications acceptable in the home in order to avoid such scenarios. Vastenburg, Keyson & Ridder [5] found that the acceptability of a notification in the home is dependant on the urgency of the information, arguing that user state and context are secondary factors. Nagel, Hudson & Abowd [6] disagree, arguing that the user's activity and social context are important factors in determining the acceptability of an interruption. As a home reminder system might be delivering sensitive information, identifying social context and user activity will be vital in choosing how to deliver a notification in an acceptable manner. For example, switching to a private notification in a social context if the notification is urgent or personal is likely to increase user acceptance of the system.

In order to provide such functionality, a home care system would need access to various types of notifications. Such variety might include a choice between notifications that are directed at a target or broadcast into an environment, delivered privately and discretely or in a publicly accessible way, and notifications that abstract their meaning or explicitly state it. Multimodal interaction can provide these options to a system, and many researchers have advocated using a range of modalities to provide more appropriate and acceptable interactions [4], [7], [8], [9], [11].

A. *Multimodal Notifications*

Technological developments have prompted product developers to begin exploring alternative interaction methods. Modern smartphones are equipped with touch-screens and tactile feedback, and gestural interaction devices like the Nintendo Wii or Microsoft Kinect have been enormously successful. Multimodal hardware is cheaper and more available than it has ever been before, providing home reminder systems access to interaction mediums that were previously only available through expensive bespoke hardware.

Visual communication has been the primary interaction method of technology for many years. The most common techniques are Text, icons, photographs and diagrams. Ambient visual interaction has recently been applied in the Philips Ambilight series of televisions, which project coloured lights on the adjascant wall in order to increase immersion in movies and games.

Audio information is usually delivered via speech (either pre-recorded or synthetic), earcons (structured non-speech sound), auditory icons or simple pager style beeps. McGee-Lennon, Wolters & McBryan [9] compared the performance of speech, earcons and simple 'beep' notifications. They found that while speech was the least disruptive, it was not the most popular: speech and earcons were equally popular, while simple beeps were slightly less popular.

Tactile notifications usually take the form of a simple vibration, although other methods exist including temperature and pressure. Hoggan *et al.* [10] compared the addition of audio and tactile feedback to a visual display. While both led to performance improvements, they found that background vibration reduced this effect for tactile feedback as did noise for audio feedback. Interference differs depending on the modalities of the task and interference, as shown by Latorella [11].

Smell-based technology is often dismissed as impractical; however air fresheners are commonplace in houses and cars, while the perfume industry is hugely profitable. Smells can invoke strong emotive responses in a person, such as hunger or lust: their potential as a notification medium has been overlooked. Brewster, McGookin & Miller [13] found that participants were able to use olfactory tags to recall photos, but their performance was not as high as with textual tags. In addition, they reported practical problems working with smells, such as difficulty containing them when they were not wanted. Warnock, McGee-Lennon & Brewster [7] also reported practical problems with smells, yet found them to be no more disruptive than the other modalities tested and comparable to the performance of tactile notifications.

Warnock, McGee-Lennon & Brewster [7] have carried out the most comprehensive comparison of multimodal interaction techniques and compared the performance of visual, auditory, tactile and olfactory notifications. They concluded that the differences between the modalities were due to their inherent properties, *e.g.* that tactile notifications were difficult to miss. None of the modalities tested were found to be ineffective as notifications, although the authors asserted that some of the notifications would be suited to a particular purpose, and not broadly applicable in every situation.

Arroyo, Selker & Stouffs [8] investigated the disruptive properties of heat, smell, sound, vibration and light. Their primary finding was that there were pronounced differences between participants; they believed that these differences were dependent on prior exposure to similar sensory stimuli, *e.g.* that a musician would be less disrupted by audio notifications. Other research has shown that training can reduce the disruptiveness of notifications [14], as can familiarity with the primary task [15], which may support their conclusion.

If so, then common methods of notification (such as an audio beep) should be less disruptive than uncommon methods (such as an olfactory notification). Warnock, McGee-Lennon & Brewster [7] did not find increased disruption for less common notification modalities, and Arroyo, Selker & Stouffs [8] based their opinion on subjective measures. McGee-Lennon, Wolters & McBryan [9] found that for audio notifications, a user's preferences seemed unrelated to their performance.

Designing home reminder systems requires a greater understanding of multimodal notification properties than we currently have. Most of this research is based on objective measures, but it is also important to evaluate subjective measures. A loud alarm might be very effective at alerting the user, but that doesn't make it an appropriate notification. Presented here is a subjective evaluation of a variety of unimodal notifications.

III. Study Design

A study was carried out over two experiments to subjectively evaluate a range of unimodal notifications. Both

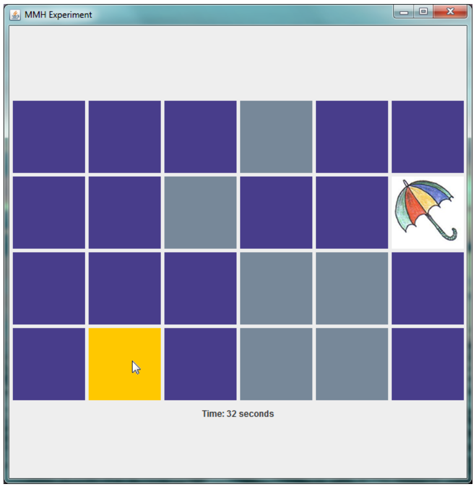


Fig. 1. A card-matching game in progress.

experiments were within-subjects designs consisting of a primary task and notifications which instructed the participant to carry out a brief secondary task. The primary task was a card-matching game as described in Sect. III-A, and the secondary task was to receive and acknowledge notifications. The notifications used in both experiments are discussed in Sect. III-B.

In Experiment 1 the notifications were grouped by sensory apparatus, *i.e.* visual, auditory, tactile and olfactory. Every notification required a response, which meant every notification interrupted the user. The *independent variable* in Experiment 1 was the sensory apparatus that received the notification. This experiment is presented in Sect. IV.

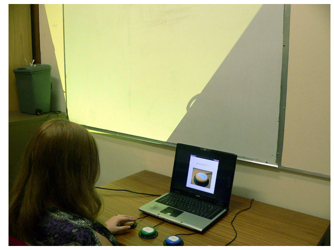
In Experiment 2, the notifications were not grouped by sensory apparatus, but by the modality itself, *i.e.* speech, text, tactile, *etc*. The majority of the notifications delivered did not require a response, making them distractions. The *independent variable* in Experiment 2 was the notification modality. This experiment is presented in Sect. V.

In both experiments the *dependent variable* was the subjective workload, measured by the NASA Task Load Index (NASA-TLX) subjective workload assessment form [16]. NASA-TLX measures subjective workload on six 21-point Likert scales representing Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. These are then combined to generate an overall workload score. The NASA-TLX also provides 15 pair-wise questions to generate a 'weighting' for the six scales, although research has shown that the assessment is just as powerful without weightings [17].

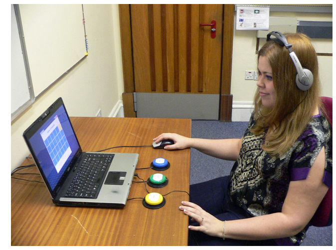
The hypotheses for the study were as follows:

- H1 With *interrupting* notifications, the unimodal notifications (grouped by sensory apparatus) will have different subjective workload ratings, measured by NASA TLX.
- H2 With *distracting* notifications, the unimodal notifications will have different subjective workload ratings, measured by NASA TLX.

The subjective evaluation was carried out by evaluating subjective workload with the NASA TLX form, observing participants as they carry out the experiment, and through informal discussions with the participants. Participants were encouraged to discuss their thoughts and opinions on the notification between experimental conditions. This would allow for the identification of modalities unsuitable for home use. The observations from the experiment will be presented in Sect. VI and discussed in Sect. VII.

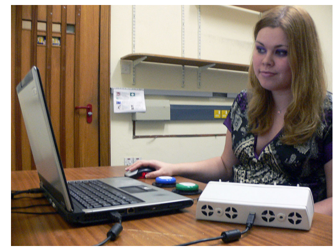


(a) Abstract Visual



(b) Audio

(c) Tactile



(d) Olfactory

Fig. 2. Hardware configurations.

A. Primary Task

A primary task was desired that would encourage the type of cognitive workload experienced by a person at home. This would provide a more realistic picture of how multimodal notifications might interfere with home life. McGee-Lennon *et al.* [9] used a digit span test to evaluate serial recall, and Arroyo & Selker [8] used a proof-reading test for their primary task. It was felt that such reading and writing tasks would encourage a cognitive workload that is more likely to be encountered in an office environment than at home.

The task chosen was a simple card-matching game called 'Concentration' (also known as Memory or Pairs), as used by Warnock, McGee-Lennon & Brewster [7]. In concentration, pairs of cards are presented face-down to the player. The player can then turn over two cards per turn in an attempt to find the pairs and remove them from the game (see Fig. 1). Concentration is a simple leisure activity that might well be carried out at home, it is a well-known game with simple rules and it can quickly build a mental workload. The game was configured in the same way as by Warnock *et al.* ; the cards showed simple alphabet caricatures[1], and each game comprised 24 cards with a 60 second time.

TABLE I
MODAL SPECIFICATION

Modality	Message: Heating	Lights	Telephone	Justification
Visual Modalities				
Text	"Heating"	"Lights"	"Phone"	A simple one-word message displayed in a large bold font above the game.
Pictogram	IEC-60878 Thermometer	IEC-60878 Light	ISO-7001 Telephone	Taken from two international standards; IEC-60878 and ISO-7001.
Abstract Visual	Yellow Light	Green Light	Blue Light	Projector used to shine a coloured light against the wall. The colour of the light corresponds to the button.
Auditory Modalities				
Voice	Spoken "Heating"	Spoken "Lights"	Spoken "Phone"	Created using the same synthetic voice that was used by McGee-Lennon, Wolters & McBryan [9].
Earcon	Acoustic Grand Piano	Clarinet	Marimba	The earcons had the same rhythm and varied in the sound of the instrument; taken from an experiment by McGee-Lennon, Wolters & McBryan [9].
Auditory Icon	Gas Ignition	Light Switch Click (x2)	Phone Dialing Beeps	Auditory icons at 1 second each taken from an online sound effect archive.[2]
Tactile Modality				
Tacton	multiLP	textLP	voiceLP	Tactons varied in rhythm, and were taken from an experiment by Brewster & Brown [18].
Olfactory Modality				
Aromacon	Dale Air "Dark Chocolate"	Dale Air "Riverside"	Dale Air "Raspberry"	Smells were selected based on information from an experiment by Brewster *et al.* [13].

B. Notifications

To evaluate the differences between notifications in different modalities, a wide range of unimodal notifications were designed for the experiments. These included common notification techniques such as text and speech along with less common notification modalities such as olfaction and abstract visual display. In total, 8 unimodal notifications were developed for the experiments, ensuring a thorough evaluation of the different techniques. Table I specifies the exact configuration of each notification and also shows how they were grouped by sensory apparatus in Experiment 1.

The text and pictogram notifications are delivered directly into the game window to the top of the play area. No additional hardware was required for this. The abstract visual display was created with a short-throw projector positioned to project a coloured light against the wall adjacent to the participant. The projector was deliberately aligned so that the projection lies in the peripheral vision of the participant, as shown in Fig. 2a.

In all the audio conditions, the notifications were delivered through a pair of Sennheiser HD 25-1 II closed-back headphones, as shown in Fig. 2b. These headphones helped to prevent background noise from causing interference, such as reported by Latorella [11].

Tactile notifications were delivered via an Engineering Acoustics Inc. C2 vibrotactile actuator[3] powered by a small amplifier. This was secured to the top of the wrist on the participant's non-dominant hand with a stretchable bandage, as shown in Fig. 2c. The device has a very low latency and was able to create precise tactile messages.

The olfactory notifications were delivered using a Dale Air[4] Vortex Active smell device, which has the capacity for 4 different scents. Scents are stored on 1-inch disks, which are blown on by a fan to deliver the smell. Delivery times are much longer than other devices; to ensure the smells were delivered in a reasonable time frame, the smell device was placed directly in front of the participants as shown in Fig. 2d.

The notifications used in the study were all powered by off-the-shelf technology; any commercial product providing interaction in these modalities would be likely to offer them at a similar quality. The notifications are explicitly defined in Table I, and were used in Experiment 1 exactly as shown. Minor modifications were made for Experiment 2, which are detailed in Sect. V.

[1]Caricatures by Speech Teach UK, http://www.speechteach.co.uk
[2]PacDV Sound Effects, http://www.pacdv.com/sounds/
[3]Engineering Acoustics, http://www.eaiinfo.com
[4]Dale Air, http://www.daleair.com

IV. Experiment 1

Experiment 1 asked participants to respond to three different notifications per condition, matching them up with 3 physical buttons. There were four experimental conditions and one control condition. The experimental conditions corresponded to a sensory apparatus: visual, audio, tactile and olfactory.

The secondary task is to press one of three buttons in response to a notification. These buttons are large, physical, coloured buttons placed directly in front of the participant and fixed to the desk. Similar to the experiment carried out by McGee-Lennon, Wolters & McBryan [9], the buttons were labelled with the terms "Heating", "Lights" and "Telephone" in order to provide home-related context, and also to provide additional semantic links between the notifications and buttons where the notification modality allowed for it.

A. Procedure

Experiment 1 was carried out with 27 participants (14 male and 13 female). The participants included 20 people in the 18-30 age group, 4 people aged 31-45, and 3 aged 46-60. All experiments were carried individually under the same conditions.

At the beginning of each trial, participants were given an information sheet, consent form and a short demographic survey to collect gender and age information. Participants were also asked to self-assess their sensory abilities on a 21-point Likert scale.

Each participant then completed a control condition and 4 experimental conditions, one for each sensory apparatus. During the visual and audio conditions, each participant would only recieve *one* of the modalities described in Sect. III-B; *e.g.* in the Visual condition, a participant would receive either textual, pictographic or abstract-visual notifications. The 27 participants in the visual condition were counter-balanced: 9 received textual notifications, 9 received pictograms and 9 received abstract visual notifications (and similarly for audio). This data was grouped by sensory apparatus and the experiment was treated as a within-subjects design.

The control condition had no notifications and always came first, helping to ease participants into the experiment. The experimental conditions were then delivered in a random order. At the start of each condition, a screen described the type of notification, and the hardware was configured as required (see Sect. III-B). The participant was provided the opportunity to make minor comfort adjustments (such as volume) before each condition began.

Participants were trained by introducing each notification in turn and associating them with the correct button. Notifications were then delivered randomly; in order to proceed the participant had to correctly acknowledge 6 notifications in a row. If a mistake was made, the participant was informed and corrected. This continued until the participant had correctly acknowledged 6 sequential notifications. This training helped to ensure that each participant had fully understood the links between notifications and buttons at the start of the game.

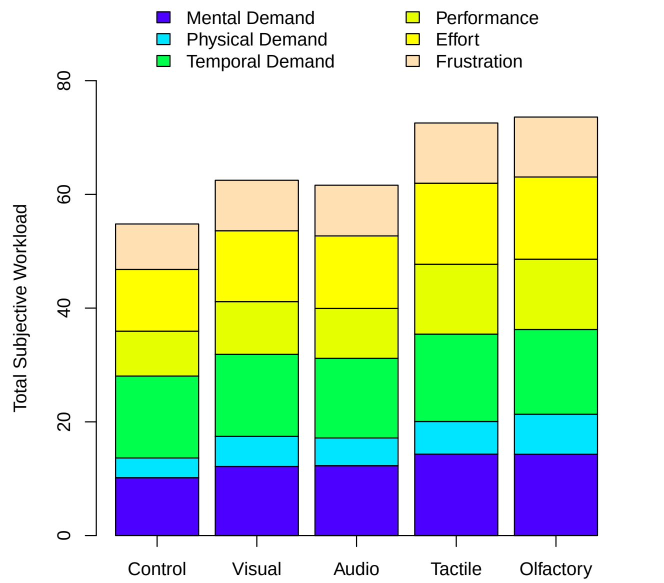


Fig. 3. Total subjective workload by modality (grouped by sensory apparatus). The total subjective workload ranges from 0 to 126.

Once the training had been completed, the participant played five games of concentration. Three notifications were delivered in each game, with buffers to prevent overlap and stop quick players missing the final notification. At the end of each game the participants clicked a button to start the next game, providing an opportunity to rest.

When all the games were completed, participants were asked to stop and fill in a paper-based NASA-TLX form [16]. Once all the conditions had been completed, participants were paid and offered the opportunity to ask questions. The experiment required around 50 minutes per participant.

B. Results

A Kruskall-Wallis test found a main effect of modality on the overall workload rating $(\chi^2(4) = 22.15, p < 0.001)$. *Post-hoc* pair-wise comparisons with Bonferroni correction found significant differences between the Control and the Olfactory $(p < 0.001)$ and Control and Tactile $(p < 0.01)$ conditions. The total workload scores are shown in Fig. 3.

Breaking the workload down into its component parts showed that there was no statistical difference between the workloads of the Physical, Temporal, and Frustration components of the test. Kruskal-Wallis tests found significant differences for Mental Demand $(\chi^2(4) = 20.05, p < 0.001)$, Performance $(\chi^2(4) = 18.66, p < 0.001)$ and Effort $(\chi^2(4) = 16.16, p < 0.01)$. In all three, *post-hoc* pair-wise comparisons with Bonferroni corrections found significant differences between Control and Olfactory (Mental Demand $p < 0.01$, Performance $p < 0.01$, Effort $p < 0.01$) and Control and Tactile (Mental Demand $p < 0.01$, Performance $p < 0.05$, Effort $p < 0.05$) conditions.

The experiment found no significant differences between the workload ratings associated with each sensory apparatus; differences only existed between the control and experimental conditions. There is no evidence to support the first hypotheses that notification modality will have an effect on the workload. The workload ratings and observations are explored further in Sect. VII.

V. EXPERIMENT 2

While Experiment 1 grouped the notifications by sensory apparatus, Experiment 2 considered the notifications individually. Participants were asked to respond to one notification per modality and ignore two 'distractor' notifications. There were eight experimental conditions and one control condition. The experimental conditions corresponded to the modalities discussed in Sect. III-B. This experiment used only one button, which was unlabelled and a different colour from any colours used in the experiment. This prevented any accidental semantic links being created between the notifications and the button.

A number of changes were made after Experiment 1 to improve the design of the experiment, detailed as follows;

- New earcons were used that varied in both rhythm and instrument, to help ensure that the earcons were sufficiently different from each other.
- Two of the tactons were replaced with alternatives from the same source (Brewster & Brown [18]), so that they varied in both rhythm and roughness. This helped to ensure they were sufficiently different from each other.
- Practice games were provided at the start, allowing participants to familiarise themselves with the game.
- The control condition was randomly delivered with the other conditions.

A. Procedure

Experiment 2 was carried out with 20 participants (14 male and 6 female). 19 of the participants were aged 18-30, and 1 participant was aged 31-45. All experiments were carried out under the same conditions with the same hardware used in Experiment 1.

Most of the experimental procedure was identical to Experiment 1. The experiment began by giving participants an information sheet, consent form, demographic survey and sensory ability self-assessment. Conditions were randomly delivered and the opportunity was given to make comfort adjustments (such as volume) and take breaks between conditions. The only difference in the general format was the addition of practice games at the start of the experiment.

Training in Experiment 2 was simpler and quicker as participants were only required to memorise one notification per condition. Training began by introducing all three notifications, then randomly allocating one to the participant. Participants were shown all three notifications again, and the training concluded when they were able to correctly respond to their notification and ignore the distractors.

With the training complete, each participant played four games of concentration per condition. Each game had a distinct configuration, as follows;

1) No notifications requiring a response.
2) 1 notification requiring a response.
3) 2 notifications requiring a response.
4) 0-3 notifications requiring a response, determined at random to prevent participants identifying a pattern of notification delivery.

The rest of the procedure was identical to Experiment 1. The experiment required around 60 minutes per participant.

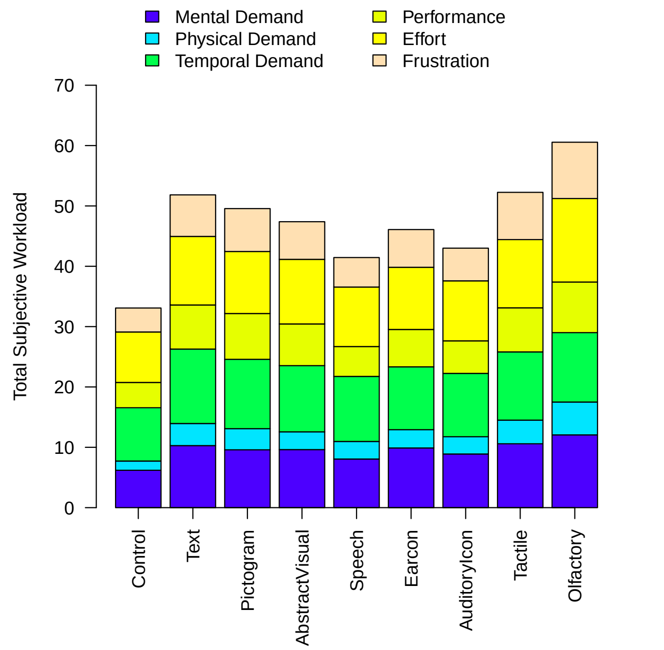


Fig. 4. Total subjective workload by modality. The total subjective workload can range from 0 to 126.

B. Results

A Kruskall-Wallis test found a main effect of delivery method on the overall workload ($\chi^2(8) = 20.62, p < 0.01$). *Post-hoc* pairwise comparisons with Bonferonni correction found significant differences between the Control and the Olfactory conditions with $p < 0.01$. The overall workloads are shown in Fig. 4.

Considering the individual workload components found no effect of modality on the Mental, Physical and Temporal workload scores. Kruskal-Wallis tests did find an effect of modality on Performance ($\chi^2(8) = 16.67, p < 0.05$), Effort ($\chi^2(8) = 15.6, p < 0.05$) and Frustration ($\chi^2(8) = 19.89, p < 0.05$). *Post-hoc* pairwise comparisons with Bonferroni corrections found no statistical differences within Performance, and found significant differences only between Control and Olfactory in Effort ($p < 0.05$) and Frustration ($p < 0.01$).

As in Experiment 1, the evidence here does not support the hypothesis that notification modality will affect workload ratings. Significant differences only exist between the control and experimental conditions. These results are shown in Fig. 4 and will be discussed in Sect. VII.

VI. OBSERVATIONS

In Experiment 2 some participants reported that the textual notifications were quite distracting, and the average workload in the textual condition is slightly higher than the other visual conditions. Schumann-Hengsteler [20] theorised that when playing concentration adults can re-encode the card contents in order to spread the workload across available memory. The textual notifications may be introducing some interference that makes this task more difficult.

The speech condition was well received, with some participants impressed when informed that the speech was synthetic. The Auditory Icons also received a favourable reception, with many participants stating that they liked the notifications.

The earcons used in Experiment 1 had the same rhythm, and many participants found it difficult to learn the mapping of notifications to buttons. In Experiment 2 different earcons were used which participants responded to more favourably. Despite the abstract nature of earcons, in both experiments the subjective workload ratings of earcons were comparable to notifications that have semantic links to the buttons.

In Experiment 1, many participants in the Tactile condition reported 'forgetting' the mapping between the notifications and the buttons. This usually happened at the first in-game notification (immediately after responding to six notifications in a row to complete the training). It was theorised that this might happen because most people are unused to remembering and processing this type of tactile information. This issue was not reported in Experiment 2; this could be because the tactons were changed or because the participants had less to remember.

Participants showed very different attitudes towards olfactory notifications. While many participants were sceptical, there were few problems during the Olfactory condition. While some participants believed that the notifications were highly disruptive, evidence suggests that the olfactory notifications are no more disruptive than any of the other notifications [7]. It is believed that this reaction is due to the unconventional nature of olfactory interaction devices. Many participants were also afraid of 'unpleasant smells'; one participant threatened to leave if the smells would be unpleasant, but found the smells quite agreeable when they were introduced. Participants expressed a generally positive attitude towards the smells used in the experiment.

The olfactory condition presented some practical issues. The smells were difficult to contain, and would disperse naturally while loaded into the Vortex device. The Vortex device was very slow, which meant that all participants had to be positioned close to the device. Some participants would sit in awkward positions in order to be close to the device and carry out the primary task. The smell disks themselves had a limited lifespan, and four disks per scent were required to carry out both experiments. The disks remained potent for most of their lifespan, then quickly became flat. It was very difficult to predict the remaining lifespan of the disks, as they did not seem to degrade gently as expected.

In general, observation and informal discussions did not reveal any serious issues; most of the misgivings expressed by participants were entirely subjective with other participants expressing opposing views. The exception was the Olfactory condition, which presented a number of practical issues. However, the scope for smell based notifications is quite small; an example application is the smell of food to remind a person to eat. This novel application for smell demonstrates that even with the observed issues, is could still have a place in the home.

VII. Discussion

While both experiments found a main effect of Modality on Workload, *post-hoc* pairwise comparisons revealed that these differences only existed between the control and experimental conditions. In addition, Fig. 3 and Fig. 4 show that all the conditions have a similar distribution of workload components. This is quite encouraging, as it suggests that the notifications used in the experiments all required a similar level of work to respond to.

An interesting result was the lack of significant differences between the control conditions (which had no notifications) and the experimental conditions. This shows that multi-tasking and identifying distractions did not require a significant increase in work for most of of the notifications, (olfactory notifications being the main exception).

In Experiment 1 participants were able to judge their performance accurately, with their subjective performance ratings correlating strongly with the number of cards matched ($\rho(-0.67) = 132, p < 0.001$). Similar correlations exist for each individual modality with the exception of olfactory. The participants were unable to accurately assess their own performance in the olfactory condition, perhaps due to a lack of confidence when dealing with an uncommon interaction modality.

Notification response accuracy did not correlate with subjective performance in both experiments, with the exception of the Audio ($\rho(-0.42) = 135, p < 0.001$) and Tactile ($\rho(-0.24) = 130, p < 0.01$) conditions of Experiment 1. This strongly suggests that participants primarily based their subjective performance ratings on their success in the primary task.

Comparing the results of the two experiments is difficult; while the subjective workload scores in Experiment 2 are lower than those in Experiment 1, the control condition also has a lower workload rating. Theoretically, the workload ratings for both control conditions should be very similar. These differences are likely due to the addition of the practice games and the randomisation of the control condition in Experiment 2. Regardless, direct comparison between the two experiments would be difficult as the NASA-TLX measure does not provide separate measures for the primary and secondary tasks.

The NASA-TLX subjective workload was chosen because it could be administered quickly and is a validated tool for subjective workload measurement [17]. However, the assessment does not allow for the separating of primary and secondary task data, making it difficult to interpret the results. An alternative assessment would have been preferable for this experiment that additionally allowed for the measurement of factors such as 'confidence' or 'pleasantness' which could have provided a greater insight into the subjective ratings. However, the TLX data gathered is able to successfully demonstrate that the modalities used in these experiments can all be considered acceptable for use as simple notifications.

VIII. Conclusion & Future Work

Participants had very different opinions on the various modalities they encountered, which highlights the need to consider the user when developing a multimodal system. As noted by McGee-Lennon, Wolters & McBryan [9], a user's modal preferences will not always match the modalities that are most effective for them. In order to ensure that multimodal notifications are acceptable and effective, both subjective and objective data should be used to customise the system for the end user as defined by Wang & Turner's key properties [3].

Fulfilling Wang & Turner's key properties [3] will not guarantee that the notification system will be acceptable in the home environment; in order to achieve this goal, the system would need to need to dynamically select the most appropriate notification modality based a number of factors including user activity and social context [6], environmental factors [11] and message urgency [5]. Accomplishing this would require a wide range of modalities; this study has shown a range of modalities that have performed well under subjective and objective studies [7], and which can be practically implemented with off-the-shelf technology.

The lab-based study presented here shows that people have an open mind regarding multimodal interaction, and the recent success of multimodal devices such as the Microsoft Kinect reveals that many people will accept this technology into their home. Yet this study cannot tell us how people will react to living with such technology over a longer period of time. Home trials with end users need to be carried out in order to evaluate how compatible this technology is with life at home.

Furthermore, additional experiments will be needed in order to ascertain the suitability of these notifications for different applications. For example, which notification modalities are most appropriate when the user has guests or is watching television? Vastenberg & Ridder [21] have shown that urgency and delivery method are key to the acceptability of a notification, but further research is needed to evaluate the relationship between acceptability of a notification and modality, message urgency, social context and the user's activity.

In conclusion, notification systems should make use of a wider range of multimodal interaction techniques. The work presented here and other work [7], [8], [9], [21] have shown that a range of modalities can provide effective reminders, and that switching modalities allows for adaptation to changing requirements, environmental factors, user preferences, social context and user activity. More work is needed to help developers include these modalities, in particular by identifying guidelines for the design of home reminder systems that can satisfy Wang & Turner's four key requirements [3] while remaining acceptable to the user.

Acknowledgment

I'd like to thank the EPSRC for funding this research (grant number EP/G069387/1), the experiment participants, Miss M. Vernigor for modelling the equipment and Dr M. McGee-Lennon for her feedback and insights.

References

[1] Office for National Statistics, "National Population Projections," 2008.

[2] Department of Health, "National Service Framework for Older People," 2007.

[3] F. Wang and K. Turner, "Towards personalised home care systems," *PETRA*, vol. 282, 2008.

[4] T. McBryan, M. R. McGee-Lennon, and P. Gray, "An integrated approach to supporting interaction evolution in home care systems," *PETRA*, vol. 282, 2008.

[5] M. H. Vastenburg, D. V. Keyson, and H. Ridder, "Considerate home notification systems: a field study of acceptability of notifications in the home," *Personal and Ubiquitous Computing*, vol. 12, no. 8, pp. 555–566, Jun. 2007.

[6] K. S. Nagel, J. M. Hudson, and G. D. Abowd, "Predictors of availability in home life context-mediated communication," *Computer Supported Cooperative Work*, vol. 6, no. 3, p. 497, 2004.

[7] D. Warnock, M. McGee-Lennon, and S. Brewster, "The role of modality in notification performance," School of Computing Science, University of Glasgow, Glasgow, UK, Technical Report TR-2011-325, Jan. 2011.

[8] E. Arroyo, T. Selker, and A. Stouffs, "Interruptions as multimodal outputs: which are the less disruptive?" in *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*. IEEE Comput. Soc, 2002, pp. 479–482.

[9] M. R. McGee-Lennon, M. Wolters, and T. McBryan, "Audio Reminders in the Home Environment," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, Canada, 2007, pp. 437–444.

[10] E. Hoggan, A. Crossan, S. Brewster, and T. Kaaresoja, "Audio or tactile feedback: which modality when?" in *Proceedings of the 27th international conference on Human factors in computing systems*, ser. CHI '09. Boston, MA, USA, 2009, pp. 2253–2256.

[11] K. A. Latorella, "Effects of Modality on Interrupted Flight Deck Performance: Implications for Data Link," 1998.

[12] T. Matthews, A. K. Dey, J. Mankoff, S. Carter, and T. Rattenbury, "A toolkit for managing user attention in peripheral displays," *Symposium on User Interface Software and Technology*, 2004.

[13] S. Brewster, D. McGookin, and C. Miller, "Olfoto: designing a smell-based interaction," in *Conference on Human Factors in Computing Systems*, 2006, pp. 653–662.

[14] D. Cades, J. Trafton, and D. Boehm-Davis, "Mitigating disruptions: Can resuming an interrupted task be trained," in *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 50, no. 3. Human Factors and Ergonomics Society, 2006, pp. 368–371.

[15] J.-M. Cellier and H. Eyrolle, "Interference between switched tasks," *Ergonomics*, vol. 35, no. 1, pp. 25–36, Jan. 1992.

[16] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Human mental workload*, vol. 1, pp. 139–183, 1988.

[17] S. Hart, "NASA-task load index (NASA-TLX); 20 years later," *Human Factors and Ergonomics Society Annual*, 2006.

[18] S. Brewster and L. M. Brown, "Tactons: structured tactile messages for non-visual information display," in *ACM International Conference Proceeding Series; Vol. 53*, Darlinghurst, Australia, 2004, pp. 15–23.

[19] E. Arroyo and T. Selker, "Self-adaptive multimodal-interruption interfaces," in *International Conference on Intelligent User Interfaces*, 2003, pp. 6–11.

[20] R. Schumann-Hengsteler, "Children's and adults' visuospatial memory: The game concentration," *Journal of Genetic Psychology*, vol. 157, no. 1, pp. 77–92, 1996.

[21] M. H. Vastenburg, D. V. Keyson, and H. Ridder, "Considerate home notification systems: A user study of acceptability of notifications in a living-room laboratory," *International Journal of Human-Computer Studies*, vol. 67, no. 9, pp. 814–826, Sep. 2009.