AUDIO REMINDERS IN THE HOME ENVIRONMENT

Marilyn Rose McGee-Lennon, Maria Wolters, Tony McBryan

Computing Science, University of Glasgow
17 Lilybank Gardens, Glasgow G12 8QQ, Scotland
(mcgeemr,mcbryan)@dcs.gla.ac.uk

School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland
mwolters@inf.ed.ac.uk

ABSTRACT

In this paper we report an experimental comparison between three different types of audio reminders in the home setting: speech, earcons, and a simple pager sound. We examine how quickly and accurately participants were able to interpret the reminders, and to what extent presentation of the reminders interfered with a digit span background task. In addition, a questionnaire was used to gather user preferences and attitudes towards the different types of reminders. Although participants perform best with speech reminders, there are large inter-subject differences in performance, and over 50% prefer non-speech audio reminders. The implications for the design and application of auditory interfaces for home-based reminder systems are discussed.

Keywords: audio reminders, earcons, speech synthesis, home care

1. INTRODUCTION

Auditory output can be very useful if used appropriately, yet it can be an annoyance if used carelessly. Furthermore, differences in setting, context, personality, user preference, and output devices all affect the potential success or failure of auditory reminders. In order to ensure user acceptance, it is crucial that the user can switch between different types of audio reminders, whether these are speech or non-speech audio.

Since there have been very few studies comparing speech and non-speech audio presentation, it is very difficult to posit general design guidelines. This paper presents a pilot study designed to address this gap in the literature. We examine three types of audio reminders, a simple pager-style chime, earcons (sequences of sounds with meaning), and speech. In our comparison, we not only compare the effectiveness of the three audio options, we also investigate their acceptability and the degree to which they interfere with the user's current activities.

1.1. Audio Reminders in the Home Environment

Delivering reminders is a key task in many domains, ranging from appointment reminders to memory aids for people with cognitive impairment [1]. Since reminders can be delivered in many different ways, both within and across modalities, reminder systems are ideal for comparing different strategies for auralising the same content. In a home environment, auditory reminders complement visual reminders well. They can be attended to while the user is doing something else, and their effectiveness does not rely on a user being near a visual display.

Previous work in assistive technology has explored different ways of presenting audio reminders: sound alerts (eg [1, 2]), earcons [3, 4], or spoken dialogue systems [5, 6]. Simple sound alerts tend to be used in pager systems which alert people with minor cognitive impairment to scheduled tasks and events. Such systems have proved to be very effective in field trials [2]. Earcons [7, 8] have been used less widely. Sainz de Salces, England, and Vickers designed and tested a range of earcons for alerting older people to events in the home [3]. Their earcons consisted of two motifs, one denoting an appliance, one denoting appliance status. They did not use timbre to differentiate between earcons. Their older participants found earcons difficult to remember and suggested using relevant familiar melodies instead. A related suggestion would be to replace earcons by auditory icons [9]. Research shows however that users report auditory icons to be annoying after prolonged use [10, 11]. Although earcons do not possess the same intuitive mapping as speech or auditory icons, they can be learned and users have found them appropriate for general applications [10].

Lines and Hone [12, 13] have investigated the use of speech in alarm and alert systems. Their work was implemented in the Millenium Home home care system [14]. In this system, critical alerts were always given by broadcasting speech over loudspeakers, because speech is a quick and reliable alert mechanism that is relatively independent of the user's position. When comparing natural speech and computer-generated speech, Lines and Hone found that natural speech was more pleasant to listen to than computer-generated speech, while intelligibility was the same for both types of speech.

Spoken reminder systems have also been implemented as part of the functionality of a nursing robot [5]. The Autominder system developed for the Carnegie Mellon Nursebot [15] determines the best time for giving a reminder by using a sophisticated planning system that takes the user's daily routine and current activity into account.

1.2. Speech versus Non-Speech Audio in HCI

There has been very little research comparing speech and non-speech audio presentation of the same content. Bronstad, Lewis, and Slatin [16] compared two ways of indicating the presence of a hyperlink in a screen reader, a tone and the spoken word “link”. They found that participants made fewer errors when the link was indicated by a simple tone. Fröhlich compared different audio cues used to indicate waiting time in a dialogue system, including speech, natural sounds, and musical pieces. Speech was rated the most appropriate option, closely followed by musical indicators. These results show that the design choice depends very much not just on the task, but also on the audio options available. Furthermore, it is important to compare the intrusiveness of different types of audio cues—a key problem in sonification [17].

1.3. Cognitive Processing of Audio Reminders

Audio reminders should not only be pleasant to listen to, they also need to be easy to decode and minimally disruptive. Users may be engaged in a cognitively demanding task when the reminder is played, and may not be able to resume this task easily once they have attended to the audio signal. Disruptiveness is a particularly serious problem for home care reminder systems. A large number of conditions that require people to rely on automated reminder systems are associated with deficits in performing dual tasks, such as Parkinson's Disease (PD) [18] or head injuries [2, 19, 20]. Ashburn, Stack, Pickering, and Ward [18] showed that PD patients who were more prone to falling also had lower dual-task performance. This demonstrates why it is crucial that the reminder system doesn't make more cognitive demands than necessary.

Fortunately, there exists a large body of work that examines the disruptiveness of different types of audio stimuli [21]. The key phenomenon here is the Irrelevant Sound Effect: Attending to a spoken utterance makes it more difficult to recall a series of items [22]. Various hypotheses have been advanced to explain this effect. Salamé and Baddeley [23] argue that the effect arises because all speech, be it relevant to the current task or not, is processed in a phonological store of limited capacity. Items in this store are kept in memory through subvocal rehearsal. If other material has previously been stored in the phonological store, the recall of this material is potentially disrupted; if that material has been stored in the visuo-spatial component of working memory, no disruption occurs. Jones, Madden and Miles [24] proposed that the interference occurs because of parallel processes of seriation, one process which maintains the order of the material to be recalled, and another that parses incoming auditory percepts for serial order. The first hypothesis implies that spoken reminders will necessarily be more disruptive than earcons, while according to the changing state hypothesis, auditory sequences with a strong serial order will be just as disruptive as speech of equal length.

Various studies [25, 26] have shown that the irrelevant sound effect also occurs when non-speech sounds are used. However, this only occurs when the acoustic variation in non-speech sounds is the same as the acoustic variation in speech, with acoustic variation encompassing changes in pitch, tempo, or timbre. Therefore, very simple earcons should be less disruptive than speech.

When designing speech reminders, keeping messages short and to the point can mitigate some of the deleterious effects of using speech. Vilimek and Hempel [27] found that long spoken messages disrupted serial recall more than short keywords, which were as disruptive as auditory icons and earcons.

Another important strand of research concerns general memory capacity. As we have already seen, memory may be affected in users of home care systems. Although memory tends to decline with age [28, 29], there is great variability [30]. Hence, for users with severe memory problems, explicit spoken messages may be better than earcons, whose meaning needs to be remembered. This concern is expressed by Sainz de Salces, England, and Vickers, who report that their older participants found earcons difficult to remember [3]. Vilimek and Hempel [27] found that reaction times were longer for earcons, where the mapping from audio to meaning may not be intuituve, than for auditory icons and keywords. This shows that good design of non-speech auditory cues is vital.

1.4. Hypotheses

In this experiment, we compare three different types of audio reminders in a situation that simulates a real life home situation: adjusting the setting of common household appliances. Since people are often engaged in other tasks when a reminder is received, a simple background task was running concurrently with the presentation of the audio reminders. This task is digit span, which is highly sensitive to Irrelevant Speech effects because it is a serial recall task. The three reminder types were compared along two dimensions:

User Performance: The best type of reminder is one that distracts the user the least while still enabling him/her to successfully perform the required action.

User Preference: Regardless of performance, users will show clear individual preferences for reminder types which will need to be squared with their performance. Preferences may also depend on contexts of use.

More specifically, we make the following predictions:

H1: (Performance) Participants will make more errors attending to reminders when presented with earcons, because speech gives explicit instructions and earcons do not, and the pager sound forces participants to check the full reminder instructions textually.

H2: (Performance) Speech will result in more errors in participants' performance in the digit span task than earcons or a simple pager sound due to irrelevant speech effects.

H3: (Preference) Participant will report a preference for shorter reminders (earcons, pager sound) rather than longer ones (speech).

H4: (Preference) Participants will report a preference for the reminder type that interferes least with their performance in the digit span task.

2. METHOD

2.1. Design

The experimental task (the primary task) was to attend to an auditory reminder when participants heard the reminder play via the handheld computer. There were three types of audio reminder (pager, earcon, speech), three different types of household appliance to select (heating, TV, fan), and two operations to perform on each appliance (up +, down -). This resulted in 18 reminder trials ($3 \times 3 \times 2$). An equal number (18) of blank (no reminder) trials was included in order that a reminder was not present on every trial. These reminders were played randomly throughout the digit span trials in order to reduce the expectation that a reminder would sound at regular intervals. On trials with reminders, the reminder was played after the digit sequence had concluded. In order to avoid tiring participants, the duration of the experiment was limited to 36 trials. The complete experiment, including questionnaires, lasted an hour.

We measured user performance for both the background task (digit span) and the primary task (adjusting an appliance). The main independent variable was **Reminder Type** (Speech, Earcon, Pager). The dependent variables were:

Digit Span Correct: This was scored 1 if a subject successfully repeated a digit span with all numbers in the correct order at the correct position, 0 otherwise.

Reminder Correct: This was scored 1 if a subject selected the correct appliance (TV, heating, fan) and the correct action (up vs. down), 0 otherwise.

2.2. Participants

11 native speakers of English were recruited. Three participants were older (age 62 $\pm$ 2 years), eight participants were younger (age 27 $\pm$ 5 years). 6 participants were male, 5 were female. We deliberately included both younger and older participants because both groups are potential users of reminder systems in the home—be it in the context of smart homes or in the context of home care. A power analysis showed that his sample size is sufficient for detecting large effects in user preferences and user performance with power > 0.8.

Participants were screened for hearing problems using a simple questionnaire. Only two of the older participants reported slight problems. Participants were also asked to fill in the Prospective and Retrospective Memory Questionnaire (PRMQ), a well-validated instrument for self-reporting memory problems [31].

PRMQ scores were converted into normalised T-scores using the software referred to in [31]. T-scores are standardised to have a mean of 50 and a standard deviation of 10. Most scores were well within a standard deviation of the mean, with some higher scores within 1.3 standard deviations. This indicates that no participants were aware of particular problems with their memory, even though one participant mentioned "poor memory" on our questionnaire. Only one subject, a younger female, reported a significant difference between retrospective and prospective memory. Older and younger participants did not differ significantly in average scores.

2.3. Stimulus Design

In the pager condition, a short, simple chime (wav file) was played to indicate a reminder had occurred. In this case, no information regarding the object or action were contained within the audio itself. Instead, the users could "check" what the reminder was via by pressing the HINT button on the device control screen [see Fig 3]. This button was always available to allow users to check the action required and the appliance to be controlled. On selecting HINT, the instruction would appear textually in a pop up box in the centre of the screen. The format of the instruction was "Turn APPLIANCE DIRECTION", where APPLIANCE was one of "heating", "television", or "fan" and DIRECTION was one of "up" or "down". Users then had to click on "ok" to return to the main device control screen.

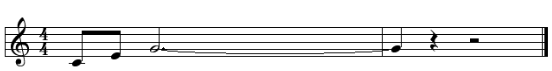

Figure 1: *Sequence of notes corresponding to turning an appliance up*

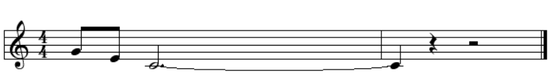

Figure 2: *Sequence of notes corresponding to turning an appliance down*

The earcons were designed to be simple increasing (up, Fig. 1) or decreasing (down, Fig 2) sequences of three MIDI produced notes. Appliances were signalled by the the instrument parameter: the TV was associated with the marimba, the heating with the clarinet, and the fan with the harpsichord.

The earcons do not contain much acoustic variation, but the notes have a pronounced serial order, forming a chord. This potentially increases interference with the recall of the digit span sequence, since non-speech audio stimuli with a strong serial order interfere more with serial recall than unordered stimuli (cf. the literature review in [21]). They were chosen to be simple abstract mappings to each appliance and were considered easily distinguishable from each other given that they were from three distinct families of musical instrument. Any semantic mapping between the earcon and the appliance would be incidental and not a design choice. However, the choice of an ascending sequence for "up" and a descending sequence for "down" is clearly mnemonic and will aid recall. Since the main aim of this experiment was comparing different types of audio reminder, we did not optimise earcon design any further; this will be left to future work.

The speech reminders were produced using the speech synthesis package Cerevoice [32], a high-quality state-of-the-art unit selection system. In unit selection speech synthesis, output is assembled from a large database of speech recorded by a single speaker. We chose to use synthetised speech instead of prerecorded prompts because synthetic speech is easier to adapt to different domains as well as to user preferences such as speaking style, gender, and persona of voice. The text of the reminders followed the pattern "Please turn APPLIANCE DIRECTION", where APPLIANCE was one of "television, heating, fan" and DIRECTION was either "up" or "down". This pattern was chosen because it was brief, yet polite. The voice used was the Cerevoice Scottish female voice "heather".

2.4. Procedure

The experiment was implemented on a handheld computer (PDA, Dell Axim X51). This was a considered design choice since we believe that home care systems might typically be controlled via mobile phones or handheld computers in many circumstances.

The user engaged in a simple number memory task (the Digit Span task) on the PDA. The digits were presented visually on the screen for 1 second with intervals of 1 second between digits. The users were then prompted to repeat back to the experimenter verbally the digit sequence they had just seen. Users were given 8 seconds to respond. After this time had elapsed, the message "Task complete" was displayed on the screen.

Immediately prior to the experiment, each subject was given the digit span test in the same format they would receive it during the experiment. The digit span sequence was increased by one each time they repeated the correct sequence until they got a sequence length incorrect twice. The participants' maximum digit span was then recorded and used in the main experiment. This ensured that the digit span task was difficult enough to be cognitively demanding and not so difficult that the user became frustrated or unable to perform the task.

During the experiment, the highest sequence of digits that would appear for a subject was N (the maximum digit span for that subject) and the minimum length of sequence was N-3. During the experiment the length of the digit span was randomized between these values to reduce the expectation for a fixed length of sequence. This resulted in participants having to concentrate both on the digit sequence as well as the reminder instructions if and when they received one.

This background task was included because people attending to reminders in their homes will typically be engaged in a variety of other primary tasks, which are disrupted by presentation of the reminder. Digit span was chosen because it is easy to adapt to individual ability levels and because it provides an easy measure of the distraction caused by the reminders.

To attend to a reminder users had to:

1. select the button "switch to device control" from the interface screen
2. (for some trials) select the "Hint" button, read the help message, and click on the help message to return to the device control screen
3. select the household appliance to control (heating, TV, fan)
4. select the operation to perform (turn up +, turn down -)
5. select the button "Return" to complete the action

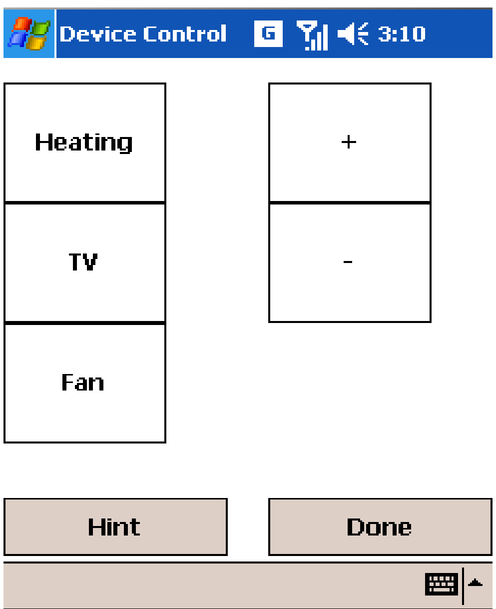

Figure 3: *The Device Control Interface*

Users could change their selections until they were happy to confirm the instruction. Users could also choose to receive a hint if they wished to check what the reminder was (cf. Section 2.3). Users were asked to verbally recall the sequence of digits for each trial. In conditions where an audio reminder was played, this required that the users attend to the reminder (as described above) before repeating the number sequence. The verbally recalled digit span sequences were transcribed from the audio recordings by two of the authors.

Participants were shown the PDA interface prior to the actual trial and played each of the possible reminders they would hear during the experiment. In addition, participants had the chance to listen to each of the reminders. For the earcons, participants were asked to guess the message until they got each type of reminder correct at least once.

2.5. Questionnaire

A post experimental questionnaire was designed and administered to collect users opinions on the audio reminder system and in particular the different reminder types. Questions addressed perceived difficulty of each of the tasks required of them (digit span and reminder, see Section 2.4) as well as ease of use of the experimental interface.

We also asked users to rate the perceived (a) helpfulness, (b) annoyance, and (c) pleasantness of each of the reminder types. To investigate the suitability for reminders in different contexts, users were asked which type of reminder they would prefer in two different contexts, being alone versus with others present. This was included because previous interviews had revealed that the preferred modality of the reminder might depend both on the content of the message and the context in which it is received.

We believe this qualitative analysis to be an important addition in auditory interface research as many results include error rates and reaction times without consideration for the true usability of the reminders in practice. In the home care setting in particular, it is essential that multimodal and auditory interfaces are designed not solely for accurate and speedy responses but also for interfaces that might be usable and acceptable to users over a prolonged period within their own homes.

3. RESULTS

3.1. Performance

In this section, we examine the effect of reminder type on participants' performance both on the primary task (attending to reminders) and the background task (serial recall of visually presented digit span). The maximum score on the trials without reminders was 18 (i.e. all 18 digit spans were repeated correctly); the maximum score on the trials with reminders was 6 for speech, 6 for earcons, and 6 for the pager condition. (For a definition of our target variables, see Section 2.1.) In our results tables, we present median scores averaged across all speakers. In *Hypothesis H1*, we assumed that speech would disrupt performance on the digit span performance to a larger extent than earcons or simple pager alerts. This is not borne out by our results: There is no significant difference in digit span scores between the three reminder conditions (Kruskal-Wallis test, df=2, χ^2=0.9636, p<0.9). Although raw scores suggest that participants tend to perform worst when presented with pager-style reminders and earcons (cf. Table 1), there is considerable variance in the data, as the boxplot of results in Fig. 4 shows. In fact, the distribution of digit span scores appears to be bimodal for speech and skewed towards lower scores for the pager and earcon conditions (cf. Table 2).

Score	Reminder Type			
	None	**Pager**	**Earcon**	**Speech**
Digit Span Correct	16	3	3	5
Reminder Correct	NA	6	5	6

Table 1: *Median scores for digit span and reminder task*

Low scores on the pager condition may be due to the need to use the HINT facility when presented with the pager beep. This not only involves another step in operating the tool, it also requires users to read and process a visually presented verbal message. This in turn potentially affects both the phonological loop and the visuospatial store. In contrast, speech only requires auditory processing facilities and the phonological loop, while earcons place a load on auditory processing and memory. This additional load on memory may be the reason earcons did not outperform speech. Table 3

Digit Span Correct	Reminder Type		
	Pager	Earcon	Speech
0-3	6	6	4
4-5	2	4	3
6	3	1	4

Table 2: *Distribution of digit-span scores*

provides data on users' reactions to reminders. As predicted in *Hypothesis H2*, reminder responses to the speech prompts are always correct. This result is statistically significant ($p<0.05$, χ^2=6.6112, df=2). Although medians suggest perfect performance in the pager condition, the detailed results tell a different story. Only 6 out of 11 participants attain the maximum score in the pager condition, as opposed to 11 of 11 in the speech condition (Table 4). This result is partly due to participants failing to use the HINT facility in the pager condition. The two participants who performed worst under the beep condition almost never checked the action to be taken. Even though it is tempting to dismiss this reluctance to check as a fluke, we suggest that it might be even more pervasive in a field context, where the experimenter is not present, and the temptation to just guess is even stronger, because users think they know what to do next.

Furthermore, even two of the eight participants who always checked the box made one mistake each. This suggests that the additional step of having to look up the action is sufficient to introduce potential errors. This finding clearly needs to be investigated further, because in the home care domain, it is often critical that users attend to reminders quickly and process them correctly. Our results certainly suggest that the simple pager alerts used in systems like those discussed in [1, 2] need to be rethought.

Reaction	Reminder Type		
	Pager	Earcon	Speech
Reminder Score	6	5	6
HINT Used	6	0	0

Table 3: *Reactions to reminder*

Score	Reminder Type		
	Pager	Earcon	Speech
0-3	2	3	0
4-5	3	4	0
6	6	4	11

Table 4: *Distribution of reminder score*

3.2. Performance and Memory

Since we did not select our participants according to memory capacity, these analyses can only offer some post-hoc insights into the role memory may play in intra-individual differences. The digit span task as it was used in this experiment mainly measures retention of information in short-term memory. Participants' performance on the pure digit span task only correlates with their performance for the pager-style reminder (ρ=0.86,$p<0.001$), somewhat less with performance for speech reminders, (ρ=0.685,$p<0.02$), and not at all for earcons (ρ=0.483). There are no significant correlations between self-reported memory problems as reflected in the PRMQ and participants' performance. More specific probes of cognitive ability are clearly needed.

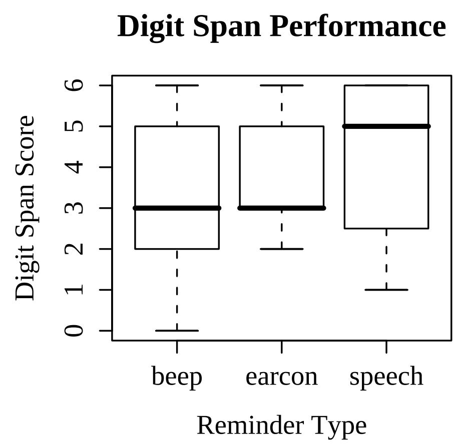

Figure 4: *Performance on the Digit Span Background Task (averaged across participants)*

3.3. Participants' Responses to the Experiment Interface

In our questionnaire, we first assessed how comfortable participants were with operating the reminder interface. While most people (N=9) responded that they were confident (N=7) or very confident (N=2), two participants responded that they were not confident. This indicates that in future experiments, participants need to be given more time to familiarise themselves both with the novel interface and with the different audio stimuli and their meanings.

This is further borne out by participants' assessments of the usability of the experimental interface. The most difficult aspect seems to have been selecting the device. Only 7 participants found this to be easy or very easy, whereas all participants found the digit span task and reading the screen easy or very easy. 9 out of 11 participants found the reminder facility easy or very easy to use. This may explain why some participants did not check reminders as often as they needed to. Despite this, all but one person rated the HINT facility as either helpful (N=1) or very helpful (N=10).

3.4. Acceptability

Any audio reminder system can only be deployed successfully if the audio reminders are acceptable to the user. Table 6 shows considerable variation in user preferences. Our *hypothesis (H3)* that shorter reminders would be preferred was not borne out: Roughly half of our participants liked earcons, whereas the other half preferred speech.

The questionnaire also yielded rich data on the reasons for participants' preferences: Many people commented on the fact that speech was the easiest to get the information from but the caused the most interference with the number task. This is not reflected in the results as summarised in Table 1. More fine-grained analyses of the digit span production data are needed to determine whether

Attitude	Reminder Type		
	Pager	**Earcon**	**Speech**
Helpfulness	3	3	5
Annoyance	1	2	2
Pleasantness	3	4	4

Table 5: *Median scores for attitudes to each reminder type*

there may be a difference in the amount of effort needed to repeat the digit sequence. It was also pointed out that the reminders that are the easiest to understand may become annoying over time. This needs to be tested in long-term case studies.

We asked users about their preferences in two contexts, once when alone, and once with others present. Remarkably, the responses were similar for both contexts, except for two participants. One preferred earcons when alone, and speech when with others, the other preferred speech when alone, and earcons when with others. One user liked non-speech audio stimuli in general (pager or earcon), while another liked meaningful auditory stimuli (earcons or speech). Since two participants preferred more than one reminder type, the totals in Table 6 add up to 13.

When asked for reasons for their preferences, the same feature of a reminder would be seen as both positive and negative. For example, some users felt that the explicitness of speech was an advantage when others were present, because the reminders would not need explaining to guests. Others, however, considered this explicitness inappropriate for some types of alerts, such as medication reminders. These alerts were judged as too private and should be delivered using earcons. Feedback like this underscores the need to couple experimental testing with qualitative data.

In our questionnaire, we further elicited opinions on three aspects of acceptablity: whether the reminders were helpful, whether they were annoying, and whether they were pleasant. All three reminder types were rated on a five-point Likert scale. Table 5 presents the median scores for each property. The only significant difference was in ratings of helpfulness: Speech was clearly perceived to be the most helpful (Kruskal-Wallis χ^2=9.3553, df=2, p$<$0.001). The differences in annoyance (χ^2=0.693, df=2, p$>$0.7) and pleasantness (χ^2=1.3621, df=2, p$>$0.5) were not significant. Again, inter-subject variation was considerable. As Figure 5 shows, speech was unanimously perceived as very helpful, whereas opinions about the usefulness of earcons and pager reminders varied enormously. While the variation for speech and earcon judgements is similar, pager ratings range over the whole spectrum (cf Fig. 6). Reasons for ratings reflected personal experience. For example, one user disliked the pager alert because it sounded like the Windows chime.

3.5. Preference versus performance: Is there a link?

User preferences cannot be predicted reliably from a subject's digit span score: Of the five participants who prefer speech or speech and earcons, only three had their highest score in the speech condition. For the five participants who preferred earcons, only two had the highest digit span score in the earcon condition. The discrepancy is largest for participants who preferred pager-style chimes: Only one of three performend best with the chime.

The correlation between performance on the reminder task itself and preference ratings is stronger: All of the five participants who preferred speech performed best in the speech condi-

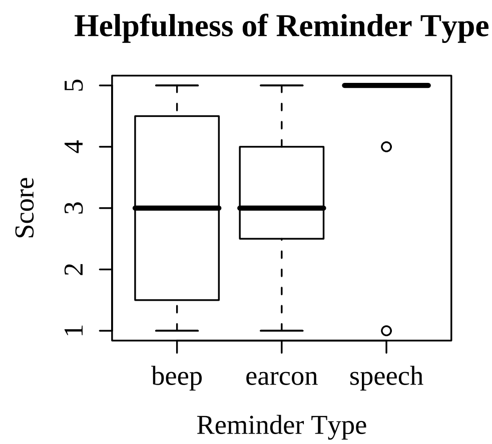

Figure 5: *Helpfulness ratings for each reminder type averaged across participants)*

tion, whereas for earcons and pager-style reminders, there was no link between preference and performance.

These findings suggests that our hypothesis H4 may have to be rejected or at least further qualified, since performance on the primary task appeared to have a greater influence on acceptability than performance on the background task. The data are too sparse to statistically test correlations—this aspect of the experiment would benefit from testing a larger number of users. Despite this, our findings confirm that user performance needs to be carefully balanced against acceptability. If users really prefer meaningful non-speech audio, then the system designer has to create a set of sounds that are both easy to interpret and not too disruptive.

Preferred . . .	Pager	Earcon	Speech
. . . when alone	3	5	5
. . . when with company	3	5	5

Table 6: *Preferences for reminder types - multiple choices possible*

3.6. Older versus Younger Participants

Although our data does not allow us to draw statistically significant conclusions about differences between younger and older participants, a brief analysis suggests that there is as much variation between older participants as there is between younger ones. Two of the three participants for whom earcons were least intrusive were older. The older participants also tend to be in the lower mode of the bimodal performance distributions. However, with the right modality, two of the three older participants perform significantly better. This underscores the need to adapt reminders to users' abilities and to test various options with the user.

4. CONCLUSION

This study highlights the need for continued research into both the effectiveness and acceptability of various auditory reminders

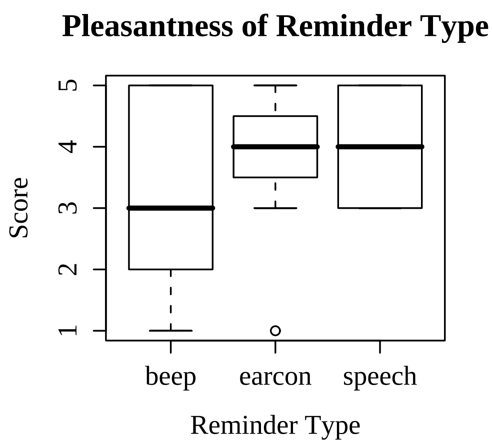

Figure 6: *Pleasantness ratings for each reminder type averaged across participants*

within the home. Despite the fact that speech was easiest to process and resulted in only moderate disruption of background tasks, a significant number of participants preferred earcons, which were perceived to be less obtrusive. This underscores again that in interface design, quantitative, experimental data needs to be supplemented by qualitative methods. Our results also suggest that systems relying on simple pager-style chimes may need to be rethought because of the potential disruption to the user.

It is crucial that auditory reminder research continues to compare the different auditory modalities (pager, earcon, speech) rather than attempt prove one modality as being "best" in all cases. The optimal choice of auditory reminder will depend on the task or operation to be attended to, the urgency of the reminder, the importance of correctly attending to the reminder, the degree of disruption caused by the reminder, the context in which the reminder is being received, previous exposure [33], and the users' perceptual and cognitive abilities. Some of these can be measured quantitatively (such as perceptual abilities), others need to be assessed qualitatively (such as previous exposure).

In some applications, such as home care reminder systems, the variation in all of these factors will be such that the user(s) will need to be able to personalize the system to their needs and modify these settings as their requirements change over time and space.

4.1. Multimodality

Any full implementation of a reminder system needs to be multimodal [34], but each component of the multimodal system needs to be adapted to user preferences [33]. Further work is required on both user performances and user preferences for different types of reminders to ensure that users are presented with enough design options to create a customised interface that is both usable and acceptable. Users must be able to choose in which modalities they wish to receive the reminders and to switch between different modalities depending on the content of the message and the context in which a message will be received.

4.2. Older vs. Younger Users

A significant percentage of users of reminder systems will be older. One factor is the increasing percentage of older people in the population. Moreover, reminder systems are a key part of home care and telecare systems. Since people tend to prefer to stay in their own homes and manage their known care needs into old age, older people are a key user group for these systems. Therefore, auditory interface research must investigate the effect age-related changes in perception and cognition have on usability. We plan to extend this to include comparisons of different age groups both in terms of performance and preference for each of these types of auditory reminder within the home.

4.3. Future Work

In future experiments, we will explore inter-subject variation in preferences and performance more systematically. Because of the large inter-subject variation found in this study, we will increase the number of participants tested in future experiments. In particular, we will focus on comparing older and younger participants. We plan to look at other tasks which are less dependent on serial recall, such as proofreading [35] and investigate the influence of participants' perceptual and cognitive skills and abilities on preferences and performance.

5. ACKNOWLEDGEMENTS

This research was funded by Scottish Funding Council grant no. HR04016 (MATCH). We would like to thank Steve Brewster and Phil Gray for helpful discussion, and our participants for taking part in our research.

6. REFERENCES

[1] E. Inglis, A. Szymkowiak, P. Gregor, A. F. Newell, N. Hine, B. A. Wilson, J. Evans, and P. Shah, "Usable technology? challenges in designing a memory aid with current electronic devices," *Journal of Neuropsychological Rehabilitation*, vol. 14, pp. 77–87, 2004.

[2] P. Wright, N Rogers, C. Hall, B. Wilson, J. Evans, H. Emslie, and C. Bartram, "Comparison of pocket-computer memory aids for people with brain injury," *Brain Injury*, vol. 9, pp. 787–800, 2001.

[3] Fausto Sainz de Salces, David England, and Paul Vickers, "Household appliances control device for the elderly," in *Proceedings of the 2003 International Conference on Auditory Display*, 2003.

[4] Fausto J Sainz de Salces, David England, and David Llewellyn-Jones, "Designing for all in the house," in *CLIHC '05: Proceedings of the 2005 Latin American conference on Human-computer interaction*, New York, NY, USA, 2005, pp. 283–288, ACM Press.

[5] M. Pollack, L. Brown, D. Colbry, C. E. McCarthy, C. Orosz, B. Peintner, S. Ramakrishman, and I. Tsamardinos, "Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment," *Robotics and Autonomous Systems*, vol. 44, pp. 273–282, 2003.

[6] M. Pollack, "Intelligent Technology for an Aging Population: The Use of AI to Assist Elders with Cognitive Impairment," *AI Magazine*, vol. 26, pp. 9–24, 2005.

[7] M. Blattner, D. Sumikawa, and R. Greenberg, "Earcons and icons: Their structure and common design principles.," *Human Computer Interaction*, vol. 4, pp. 11–44, 1989.

[8] S.A. Brewster, "Non-speech auditory output.," in *The Human-Computer Interaction Handbook*, J.A. Jacko and A. Sears, Eds., pp. 220–239. Lawrence Erlbaum, Mahwah, NJ, 2002.

[9] W. Gaver, "The sonicfinder: An interface that uses auditory icons," *Human-Computer Interaction*, vol. 4, no. 1, pp. 67–94, 1989.

[10] L.A. Roberts and C.A. Sikora, "Optimising feedback signals for multimedia devices: Earcons vs. auditory icons vs. speech," in *Proceedings of IEA, Tampere, Finland*, 1997.

[11] C.A. Sikora, L.A. Roberts, and Murray L., "Musical vs real world feedback signals," in *Proceedings of CHI, Denver, Colorado*, 1995.

[12] L. Lines and K. S. Hone, "Older Adults' Comprehension and Evaluation of Speech as Alarm System Output Within the Domestic Environment," in *2nd International Conference on Universal Access in Human Computer Interaction, Crete, Greece.*, 2003.

[13] L. Lines and K. S. Hone, "Eliciting user requirements with older adults: lessons from the design of an interactive domestic alarm system," *Universal Access in the Information Society*, vol. 3, pp. 141–148, 2004.

[14] M. Perry, A. Dowdall, L. Lines, and K. Hone, "Multimodal and ubiquitous computing systems: Supporting independent-living older users," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, pp. 258–270, 2004.

[15] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: challenges and results," *Robotics and Autonomous Systems*, vol. 42, pp. 271–281, 2003.

[16] P.M. Bronstad, K. Lewis, and J. Slatin, "Conveying contextual information using non-speech audio cues reduces workload," in *Technology and Persons with Disabilities Conference*, 2003.

[17] Nitin Sawhney and Chris Schmandt, "Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments," *ACM Trans. Comput.-Hum. Interact.*, vol. 7, no. 3, pp. 353–383, 2000.

[18] A. Ashburn, E. Stack, R.M. Pickering, and C.D. Ward, "A community-dwelling sample of people with parkinson's disease: characteristics of fallers and non-fallers," *Age and Ageing*, vol. 30, no. 1, pp. 47–52, 2001.

[19] T. Hart, K. Hawkey, and J. Whyte, "Use of a portable voice organizer to remember therapy goals in traumatic brain injury rehabilitation: a within-subjects trial," *J.Head Trauma Rehabil.*, vol. 17, pp. 556–570, 2002.

[20] G. Hein, T. Schubert, and D.Y. von Cramon, "Closed head injury and perceptual processing in dual-task situations," *Experimental Brain Research*, vol. 160, no. 2, pp. 223–234, 2005.

[21] S.P. Banbury, W.J. Macken, S. Tremblay, and D.M. Jones, "Audiotry distraction and short-term memory: Phenomena and practical implications," *Human Factors*, vol. 43, no. 1, pp. 12–29, 2001.

[22] H.A. Colle and A. Welsh, "Acoustic masking in promary memory," *Journal of Verbal Learning and Verbal Behaviour*, vol. 15, pp. 17–31, 1976.

[23] P. Salame and A.D. Baddeley, "Disruption of short-term memory by unattended speech: Implications for the structure of working memory," *Journal of Verbal Learning and Verbal Behaviour*, vol. 21, pp. 150–164, 1982.

[24] D.M. Jones, C.A. Madden, and C. Miles, "Privileged access by irrelevant speech to short-term memory," *Quarterly Journal of Experimental Psychology*, vol. 44A, pp. 645–659, 1992.

[25] D.M. Jones and W.J. Macken, "Irrelevant tones prodice an irrelevant speech effect: Implications for phonological coding in working memory," *JEP: Learning, Memory, and Cognition*, vol. 19, pp. 369–381, 1993.

[26] S. Tremblay, A.P. Nicholls, D. Alford, and D.M. Jones, "The irrelevant sound effect: Does speech play a special role?," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 26, pp. 1750–1754, 2000.

[27] R. Vilimek and T. Hempel, "Effects of speech and non-speech sounds on short-term memory and possible implications for in-vehicle use," in *Proceedings of ICAD*, 2005.

[28] L. Baeckman, B. J. Small, A. Wahlin, J. E. Birren, and K. W. Schaie, *Aging and Memory: Cognitive and Biological Perspectives*, pp. 349–377, Academic Press, San Diego, CA etc., 2001.

[29] T. A. Salthouse, R. L. Babcock, and R. J. Shaw, "Effects of adult age on structural and operational capacities in working memory," *Psychol.Aging*, vol. 6, pp. 118–127, 1991.

[30] P. Rabbitt, M. M. Anderson, E. Bialystok, and F. I. Craik, *The lacunae of loss? Aging and the differentiation of human abilities*, Oxford University Press, New York, NY, 2005.

[31] J. R. Crawford, G. Smith, E. A. Maylor, S. della Sala, and R. H. Logie, "The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample," *Memory*, vol. 11, pp. 261–275, ,Psychological physiopathology 2003.

[32] Matthew A. Aylett, Christopher J. Pidcock, and Mark E. Fraser, "The cerevoice blizzard entry 2006: A prototype database unit selection engine," in *Proceedings of Blizzard Challenge Workshop, Pittsburgh, PA*, 2006.

[33] E. Arroyo, T. Selker, and A. Stouffs, "Interruptions as multimodal outputs: Which are less disruptive," in *Proceedings of ICMI*, 2002.

[34] Elizabeth Mynatt, Douglas Blattner, Meera M. Blattner, Blair MacIntyre, and Jennifer Mankoff, "Augmenting home and office environments," in *Assets '98: Proceedings of the third international ACM conference on Assistive technologies*, New York, NY, USA, 1998, pp. 169–172, ACM Press.

[35] D.M Jones, C. Miles, and J. Page, "Disruption of proof-reading by irrelevant speech: Effect of attention, arousal, or memory," *Applied Cognitive Psychology*, vol. 4, pp. 89–108, 1990.