

Mobile Service Audio Notifications: intuitive semantics and noises

Stavros Garzonis
Dept. of Computer Science
University of Bath
Bath, BA2 7AY
S.Garzonis@bath.ac.uk

Chris Bevan
Dept. of Computer Science
University of Bath
Bath, BA2 7AY
bevan.chris@gmail.com

Eamonn O'Neill
Dept. of Computer Science
University of Bath
Bath, BA2 7AY
eamonn@cs.bath.ac.uk

ABSTRACT

It is hoped that context-aware systems will present users with an increasing number of relevant services in an increasingly wide range of contexts. With this expansion, numerous service notifications could overwhelm users. Therefore, careful design of the notification mechanism is needed. In this paper, we investigate how semantic richness of different types of audio stimuli can be utilised to shape the intuitiveness of mobile service notifications. In order to do so, we first develop a categorisation of mobile services so that clustered services can share the same notifications. Not surprisingly, it was found that overall speech performed better than non-speech sounds, and auditory icons performed overall better than earcons. However, exceptions were observed when richer semantics were utilised in the seemingly poorer medium. We argue that success and subjective preference of auditory mobile service notifications heavily depends on the success and level of directness of the metaphors used.

Categories and Subject Descriptors

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6)

General Terms

Human Factors

Keywords

Mobile services categorisation; mobile audio notifications; auditory icons; earcons; intuitiveness of audio notifications; context awareness.

1. INTRODUCTION

Network carriers often provide their own proprietary platforms for accessing web data, with services such as “find” (location based service providing information about restaurants, bars etc.) and “travel & journey” (rail timetables, driving directions etc.). These services might be available only in certain contexts, for example location-specific, user-specific or time-specific. As a consequence, users may find it difficult to know which services are available to them in any given context. The vision for context-aware systems is that they will be able to infer user intention in any given situation, and suggest to them the most appropriate

services, such as directions to points of interest, timetable updates or timely reminders to buy their groceries. However, user annoyance is certain when systems fail and notifications of irrelevant services come through. There are many reported cases where warning systems have been rejected and deactivated by users because the audio notifications are perceived as annoying [e.g. 4].

Therefore, careful design of mobile service notifications is needed. In situations where audio notifications are appropriate, sound may be utilised to intuitively convey the nature of the service being delivered, facilitating users’ choice to ignore the device if the service is not desired in that situation. Analogous to the cocktail party effect, mobile service notifications should seamlessly be interweaved in our everyday activities, such as driving or conversing, and demand attention to only when truly needed.

However, current mobile audio notifications are often intrinsically meaningless and it is only through extended usage that semantics might be learned. For example, if a user assigns a distinct ringtone for a specific friend in an arbitrary manner, she will learn this association (the more frequent the incoming calls, the quicker the association is established), and in time it will even be perceived as intuitive. How much quicker would she reach this level of intuitiveness if the assigned ringtone were a song characterising the caller? How does this compare to the situation where the ringtone is a pre-recorded message in the voice of the person who is calling?

Intuition typically is defined as the act or faculty of knowing or sensing without the use of rational processes, or in other words *immediate cognition*. However, intuition is typically the by-product of intensive long-term and mostly implicit learning. Apart from survival instinct reflexes that are somewhat hardwired in our brain, everyday life teaches us how to act and react to certain stimuli. The more persistent the association between stimulus and reaction, the more hardwired it becomes, eventually leading to seemingly intuitive behaviours.

In this paper, we investigate how intuitive meaning of different types of audio stimuli can be utilised to shape the intuitiveness of mobile service notifications. As mobile services rise in numbers, an obvious challenge is the ability of the user to learn and retain the associations between notifications and services. If untrained users can automatically and consistently identify these associations, it means that their underlying implicit cognitive links are present, and hence easily learned [26].

We therefore conducted a controlled experiment to measure recognition rates of 3 types of audio as notifications for mobile services. The services were first conceptually categorised to

OZCHI 2008, December 8-12, 2008, Cairns, QLD, Australia.

Copyright the author(s) and CHISIG.

Additional copies are available at the ACM Digital Library (<http://portal.acm.org/dl.cfm>) or can be ordered from CHISIG(secretary@chisig.org)

OZCHI 2008 Proceedings ISBN: 0-9803063-4-5

mitigate the potential problem of learnability of huge number of notifications, as clustered services can share the same notifications. Furthermore, we have gathered a variety of subjective data in relation to user preference on the different types of audio utilised in this study.

Next, we present some background information of the three sound types used in our experiment. In section 3 we present the conceptual service categorization and in section 4 we report the experiment that measured the intuitiveness of the sound-service associations. Finally, discussion of findings and future work are presented in sections 5 and 6 respectively.

2. TYPES OF AUDIO NOTIFICATION: EARCONS, AUDITORY ICONS AND SPEECH

In auditory interfaces, several types of sounds are commonly employed. Such sound types range from a simple electronic ‘beep’ to the digital reproduction of complex sounds such as human speech. Although these sounds vary greatly in their attributes and complexity, they can be classified according to their relationship to the interface event they refer to. This event could for example be the emptying of the recycling bin on a desktop computer [12], a warning signal for ground proximity in a cockpit [22,26], the presence of users in a network [9], or the success (or potential success) of hitting the target button on a mobile interface [6]. The degree to which these concepts are related to the sounds referring to them can be expressed as a continuum, ranging from a completely arbitrary relationship (e.g. a ‘beep’ sound) to a highly semantic relationship (e.g. speech) and the directness of these mappings is believed to affect their ease of learning and retention [18].

The terminology and definitions of the different types of sounds utilised in the auditory interfaces literature vary. Although there may be some inconsistencies amongst them, we have identified two broad types into which all of these non-speech sounds may be considered to fall: *earcons* and *auditory icons*¹. In the following section, we present definitions and a short background for these sound types, along with speech.

2.1 Earcons

‘Earcons’ are defined by Blattner et al. [3] as “nonverbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation or interaction”, and are characterised as “the aural counterparts of icons”[3]. Brewster et al. [8] later described earcons as “abstract, musical tones that can be used in structured combinations to create auditory messages” and are “composed of short, rhythmic sequences of pitches with variable intensity, timbre and register”.

One of the advantages of using earcons is that they are flexible, and can be tailored to each application or context in which they are used. Earcons may be composed into families of sounds by sharing some of their properties (such as rhythm or pitch), while varying other properties (such as timbre or register). This way, compounds of individual earcons can be used to deliver more complex messages by creating a sort of grammar, where similar objects are represented by one family of sounds and similar

actions by another [e.g. 8]. In addition, families of earcons can be created hierarchically, when one of their parameters is changed according to the level within the hierarchy. For example, in an experiment by Brewster [7], over 80% of the participants were able to identify the correct location of a given earcon in a hierarchy of 27 nodes over four levels. Earcons have also been utilised to create auditory maps [2], to summarise equations for visually impaired users [25], and to increase performance in navigational tasks on mobile interfaces [19, 6].

A possible weakness of earcons is that they usually have no pre-existing relationship or intuitive meaning with their referent – a potential problem since they must therefore be learned “without benefit of past experience” [11]. However, it is suggested that through deliberately attempting to create metaphorical mappings between earcons and their referents, this could improve their memorability, as users tend to construct meanings for the sounds that they hear [7]. This is supported further by Cohen [9], who found that users often constructed ‘stories’ in order to give meaning to a series of unrelated sounds (in this case, sounds provided to monitor background activities on a computer network). Additionally, there is also evidence to suggest that some attributes of sound such as loudness, pitch, tempo and onset can also relate to levels of otherwise unrelated concepts such as temperature, pressure, size and rate [23, 27].

2.2 Auditory Icons

‘Auditory icons’ are sounds designed using the concept of what Gaver [11] describes as ‘everyday listening’. Everyday listening is “the experience of hearing sounds in terms of their sources”, so that “instead of mapping information to *sounds*, we can map information to *events*” [11]. Unlike earcons, auditory icons are conceptually similar to graphical icons in that they utilise a metaphor that relates them to their virtual counterparts. Gaver [13] suggests that “if a good mapping between a source of sound and a source of data can be found, the meaning of an auditory icon should be easily learned and remembered.” Some examples include ‘shattering dishes’ for dropping an object into the recycle bin [12], ‘door slamming’ for remote users logging out the network [9], and ‘tyre-skidding’ for collision warning while operating a driving simulator [15]. However, metaphoric mappings are not always easy to find, and audio feedback can be confused with actual environmental sounds [9].

A variant of auditory icons, ‘*parameterised* auditory icons’ are constructed sounds that imitate everyday sounds. However, by manipulating them along dimensions relevant to sound producing events, they can produce families of related sounds. For example, Gaver’s *SonicFinder* application [12] incorporates sounds of different materials (e.g. wood or metal) for different virtual objects (e.g. files or applications), and different actions performed on them (e.g. knocking or scraping) to denote actions on the system interface (e.g. selecting or dragging an object). Further mapping can be applied, such as variation of frequency to denote variation on size of objects. Parameterised auditory icons are more flexible in auditory interface design, but at the same time less intuitive in the metaphors they are carrying. It is worth noting that what we refer as parameterised auditory icons have also been referred as ‘representational earcons’ [3], and were initially introduced by Gaver as being defined as “caricatures of naturally occurring sounds” that “don’t really sound like the objects they represent, but that capture their essential features” [13].

¹ The literature also suggests a third type: *arbitrary sounds* (or ‘tones’), which refer to random, simple sounds that are often used as a control condition when comparing other types of audio notification.

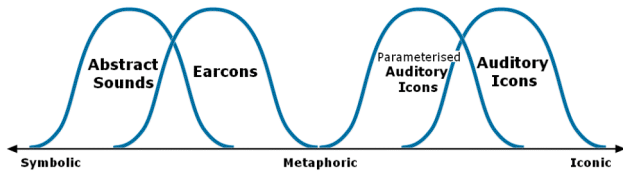


Figure 1. Non-speech sounds on the spectrum of semantics

The types of sounds described so far can be depicted along a semantics spectrum such as the one presented in Figure 1.

The comparative effectiveness of auditory icons and earcons as a method of system to user event notification remains a contested issue. Many studies have investigated and compared the effectiveness of different audio stimuli in notifications – often focussing on warning signals in cognitively demanding environments such as power plants [16], collision avoidance systems in aircraft cockpits [22, 26] or in motor vehicles [15] where the degree to which sounds are accurately interpreted in terms of the event that generated them is potentially life-critical.

Auditory icons have been found to produce quicker reactions than earcons in collision avoidance warnings [15], and there are indications to suggest that associations of auditory icons with their referents are both easier to learn [10, 5, 26] and retain [20] than earcons – although there is also some evidence that memory performance depends more on a sound-by-sound basis rather than a sound type-by-sound type basis [10]. Leung et al. [20] also found that speech warnings and auditory icons were easier to learn and remember than abstract sounds, although different training methods also affect learning and retaining of sounds within memory, regardless of the type of sound involved [7, 10].

However, the findings of Issacs et al are not typical of the effectiveness of auditory icons *per se*. On the iconic side of the semantic scale, auditory icons score much higher than earcons in terms of learnability and memorability – probably due to the iconic relation between signifier and signified. However the superiority of auditory icons drops sharply when the relations become less direct or less intuitive. This finding is also supported by Keller & Stevens [18], indicating that learnability of mappings is better in direct relations, followed by ecological and metaphorical relations, and worst with random relations.

Finally, there are contradictory reports with regard to users' preferences for sound types. For example, in Isaacs et al. [17] earcons were described as annoying, and as a result were not potentially as useful as they might be. In Cohen [9] on the other hand, it was auditory icons that were annoying to users. Preference, it would appear, depends heavily upon the specific sounds chosen and their particular context of use.

2.3 Speech

Speech – in a language known to the user – removes the difficulty of finding direct or metaphorical associations between the sound (signifier) and the system action (signified). Although the mappings between words and meaning are mainly arbitrary in most languages, they have been learnt and practised to the extent that their interpretation is automatic. This particularly benefits infrequent, high-priority speech warnings given by computers but might become annoying for messages that are less important and heard frequently. Speech notifications however can be

successfully deployed to give high-level information to several people at the same time – as for example when used in the tannoy systems of airports and train stations. Furthermore, it has been argued that “the human auditory apparatus is focused primarily on distinguishing speech from all other sounds” [21]. This means that speech auditory warnings have an increased probability to be picked up in an acoustically rich and complex environment.

The main disadvantage of speech notifications is that they are generally not nearly as brief and concise as their non-speech audio counterparts. As icons (through the use of metaphors) can present more information in less space than text can, non-speech sounds can take less time to convey more information. Similarly, speech (as text) is not universally understood as it refers only to the population familiar with the particular language, while non-speech audio metaphors can pervade across many cultures. Finally, privacy concerns might make speech the least preferred option for public notifications. One such example can be found in [1], where the perceived privacy concerns surrounding a proposed speech-enabled ATM led to it being rejected by its users.

In the following section, we present our categorisation of mobile services, devised in order to reduce the number of notifications needed. The results of this categorisation process will produce a sub-set of mobile services against which the intuitiveness of the 3 different sound types will be experimentally tested in section 4.

3. MOBILE SERVICE CATEGORISATION

Currently, the categorisations of mobile services presented by mobile network providers are very different. These various categorisations exist not just to inform users and potential customers of their services but also to promote and encourage their usage. One of the potential problems with these categorisations is that they are not strictly hierarchical, with the same service often found in several subcategories.

Suggesting a comprehensive universal categorisation of mobile services is not a trivial task – the choice of categories can be radically diverse. For example, a ‘download’ category can include ‘games’ and ‘wallpapers’ but the same services could be under categories such as ‘entertainment’ and ‘device personalisation’ respectively. Therefore, one needs to create a categorisation of services depending on the purpose they will serve. Since the existing categorisations do not fit the purpose of notifications, we need to create a new way of classifying services based on how similar (or identical) notifications of similar but distinct services need to be.

An aggregation of 52 services provided by 3 major UK network operators was collected, mapped with a mind-mapping tool and compared against the categorisations each operator used. Although more sophisticated methods such as cluster analysis could have been employed, such methods were discarded for three reasons: the classifications were too few to be fed into a cluster analysis; the classifications provided were not strict hierarchies as services often appeared in more than one category and finally, the classifications seem conceptually very dissimilar, defying comparison by cluster analysis. For example, there were some categories based on the delivery method (alerts) and some on the content (entertainment).

A significant problem remains that metrics such as ‘delivery method’ or ‘content’ do not adequately characterise all mobile services (e.g. ‘calls’ have no content and ‘sports information’ can be delivered in many different ways). However, one metric that

did fit our purpose is the ‘source’ or ‘origin’ of the incoming services. When a phone rings for example, the first question for a user is likely to be ‘who is it?’ For incoming calls, different ringtones can be assigned to individuals or groups (e.g. ‘family’). Of course, applying the same rationale to all services (such as calls, sports information or calendar reminders), we need different audio notifications to denote their origin.

Looking more closely at the services currently provided by network operators, we were able to distinguish three different distinct sources of origin: an impersonal third party (such as a company or a server), another person (known or unknown to the user) or the user herself (e.g. diary reminders). This provided us with a hierarchy that conceptually separated the services into three major categories: world-to-user, person-to-user, and user-to-self. World-to-User (W2U) services deliver information or content, such as news headlines, traffic information and songs. W2U can be further broken down into ‘information retrieval’ services (fun: cinema, sports, music etc.; work: stock market, traffic etc.; directory: finding people, businesses, places etc.), ‘downloads’ (entertainment: music, videos, games etc.; personalisation: ringtones, wallpapers etc.) and ‘streaming media’ (TV: movies, series, documentaries etc.; Radio: music radio, talk shows etc.). User-to-User (U2U) services are predominantly communication services amongst users and can be further broken down to ‘calls’ (voice, video, conference calls etc.), ‘messaging’ (sms, mms) and ‘mobile community’ services (instant messaging, chat, friends’ location etc.). Finally, User-to-Self (U2S) services include services created by (and for) the user, and can be broken down to ‘calendar reminders’ (meetings, lectures, birthdays etc.), ‘to-do reminders’ (further sub-categorised by users), and ‘synchronise devices’ (mobile with laptop, online account etc.). Using this categorisation system, we focused on nine common mobile services (Figure 2).

The choice to concentrate on nine services was not an arbitrary one. There is evidence that people can quickly learn and retain 4 to 6 different warning sounds, while learning slows down considerably when exposed to 10 warning sounds [22]. Elsewhere [e.g. 24] it is suggested that a maximum of 4 to 6 sounds is optimal, due to learning time and the requirements of memory in terms of long-term retention. Finally, there is evidence that narrow hierarchies yield better navigation performance on mobile devices [14]. Our choice of nine services allowed us to maintain a narrow hierarchy, and provided a range of choices that was manageable in terms of memory.

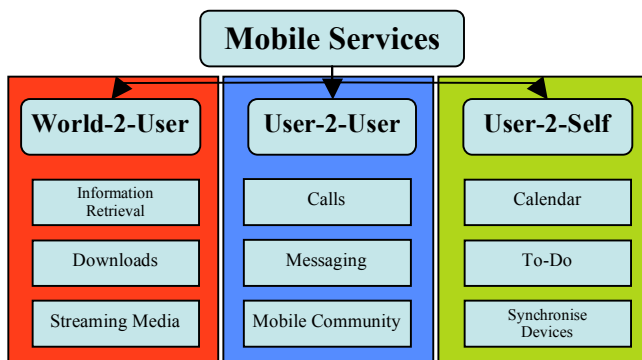


Figure 2. The top levels of our mobile services categorisation

4. EVALUATION

To compare the effectiveness of the three different sound types (earcons, auditory icons, speech) in terms of how accurately users could relate them to the service they represent, we conducted an experiment in which we asked participants to associate each sound stimulus to one of nine potential services. For clarity, and since ‘streaming media’ and ‘mobile community’ services were most likely to be unfamiliar to participants, we replaced them with the most popular instantiations of them: ‘TV’ and ‘instant messaging’ respectively.

4.1 Experimental design

The experiment had a within participants design. The independent variable (‘sound type’) was the type of audio notification used: *earcons*, *auditory icons* or *speech*. The dependent variables were *accuracy* (i.e. the participant chose the correct service from the nine possibilities) and response time. It was predicted that speech notifications would lead to the quickest and most accurate responses, followed by auditory icons and then earcons.

4.2 Participants

Following pilot trials with 2 participants, the experimental participants were 29 university undergraduate students: 13 female, 16 male, with an age range of 18-21 ($M=19.7$, $SD 1.37$). No reward was given for their participation.

4.3 Apparatus

The experimental setup was an IBM desktop personal computer (PC) connected to a 42” plasma monitor with touch screen functionality. User responses to our auditory notifications were collected using the touch screen.

Stimuli

Our experimental stimuli were a collection of 27 sounds, nine per sound type (earcon, auditory icon, speech). To compare the effectiveness of the three types of sounds, we had to ensure that each sound was related to its corresponding service as directly as possible. The inherent difference between the sound types, and the necessity to find the most appropriate metaphors to relate to the services meant that the length (in time) of the notifications could not be controlled. Therefore, the time-length of our sounds varied from 0.5 sec to 4.21 sec ($M=2.27$, $SD 0.99$). Not surprisingly, the length variance for auditory icons was the highest as the appropriate sounds had to vary considerably in order to achieve the realism needed. Ideas for how best to represent each service were produced during a brainstorming session amongst the experimenters and are described below.

Earcons

One of the strongest advantages of earcons is that they can be composed to form families of sounds, with each family representing different actions on the same object or similar actions on different objects [12]. In this case, each one of the super-categories of services (W2U, U2U, U2S) is conceptually distinct to each other, and each of them contains sub-categories of services that are semantically related. Therefore, it made sense to create three families of earcons that were different enough from one another to represent the three different categories. This distinction was created by the choice of instrument for each category: an oboe for W2U, a piano for U2U, and bells for U2S (highly distinct instrument timbres). In addition, the musician composing our earcons took into consideration Brewster’s

guidelines [8] and designed them to be as distinct as possible within each category by varying their pitch and rhythm.

Since mobile services are conceptually complex, finding an intuitive way to relate to earcons was challenging. However, in an attempt to maximise intuitiveness, we strove to make our earcons as metaphorical as possible. In some cases, the metaphor was with respect to the syllables of the words. For example, the ‘to-do’ earcon was represented by the sound of two notes dropping in pitch from one note to the next, while ‘calendar’ had 3 sounds. In other cases a conceptually higher link was possible. For example, the ‘synchronizing devices’ earcons was a collection of notes followed by a short rest, and then the same collection of notes repeated. Finally, stronger – albeit culturally specific – metaphors were used: the ‘TV’ earcon, for example, mimicked the intro sound of the popular British television show ‘Eastenders’.

Auditory Icons

Finding iconic mappings between auditory icons and services in some cases was straightforward. For example, an obvious choice for an incoming call is a classic telephone ring, and ‘TV’ was represented by the sound of a TV switching on to white noise. However, in some cases it is not always obvious how to represent a service with an everyday sound. In these cases we tried to apply a more metaphorical mapping. For example, the sound of a consistent dripping sound was used for the ‘Downloads’ service. This was meant to represent the constant stream of data being received while downloading a file.

Speech

The speech notifications were prepared using a pre-recorded and non-synthetic female voice describing each of the services. For example, the ‘calls’ service was notified as “You have an incoming call” and ‘downloads’ was notified as “You have a new download”.

4.4 Procedure

Each participant was informed about the nature of the experiment and was given a brief definition of the three different sound types. A training session preceded the experimental phase, and participants were invited to complete a short questionnaire immediately subsequent to the experiment. The whole procedure was conducted with the experimenter present. Details of each phase of the procedure are described in the following sections.

4.4.1 Training

In most research with audio notifications, participants are informed of the associations between sounds and referents during training in the first part of the experiment. However, since we wanted to measure the intuitiveness of these associations, we deemed it necessary to exclude this kind of training phase. So, although participants were familiarised with the three different types of sounds, they were not exposed to the experimental stimuli during training.

During the training phase, we used animal sounds as they are easily understood and distinct from the sounds used during the experimental phase. Participants were trained in two steps. First, they were given time to become familiar with the visual interface and the three different types of sound. The interface was a large screen with nine buttons (arranged in a 3x2 grid), each one allowing the user to play a different sound when pressed (Table 1). Participants were asked to interact with the application until they felt comfortable with the interface.

Table 1. Interface for familiarising with the sound types

Elephant (Speech)	Mouse (Speech)
Elephant (Auditory Icon)	Mouse (Auditory Icon)
Elephant (Earcon)	Mouse (Earcon)

Table 2. Training interface

Donkey	Mouse	Sheep
Dolphin	Ape	Pigeon
Elephant	Mule	Snail

I

in the second step of the training, an example evaluation trial was played so that participants were familiarised with the evaluation procedure. We presented a series of 9 different animals on the screen (Table 2) and played 3 representations for two of them (total 3x2=6) in randomised order. For each sound, the user was asked to select the animal that they thought the sound represented. Each sound was heard up to 3 times (or until the participant made a choice), followed by a two second gap before the next sound was played.

4.4.2 Experimental Phase

At the beginning of the experimental phase, the participant was presented with the nine services (Messaging, Information Retrieval, Downloads, Television, Calls, Instant Messaging, Calendar, To Do and Synchronise Devices) and was given a short definition for each of them. During this time, the participant was given time to learn the fixed position of each service on the screen, and was informed that each service had 3 sound representations, and that the same sounds would be heard more than once during the course of the evaluation.

The experimental phase comprised of three blocks of 27 trials (total: 81), in which participants were presented with all three types of sound for each of the nine services. In order to address possible learning effects, the order of presentation of the notifications was pseudo-randomised, ensuring that the same type of sound was not played in two consecutive trials.

Once the test phase was complete, the participant was asked to complete a short questionnaire, capturing their preferences and subjective comparisons of the three sound types. Each participant took approximately 25 minutes to complete the experiment.

4.5 Results

The dependent variable was the number of errors (selection of an incorrect service) recorded by a logging software. In addition, the

TABLE 3. Mean Scores and Response Times

Sound Type	Earcon	Auditory Icons	Speech
Accuracy (target service)	4.86 (18%)	10.10 (37.4%)	25.90 (95.9%)
Response time (MSec)	5116.32	4818.57	3263.61

system logged the response times of the participants across the trials. Subjective data were collected in a post-test questionnaire.

4.5.1 Quantitative data

The results presented in this section were generated from two measures: *accuracy* and *response time (RT)*. Mean and percentage scores for all measures were calculated and are presented in Table 3. Accuracy mean scores for the three sound types are also presented in Figure 3 for each service separately. Prior to any statistical analysis, the distribution of accuracy scores were normalised using a logarithmic transformation.

Comparative accuracy rates across the conditions are presented in Figure 4. Comparisons of the accuracy scores and response times across the three *sound type* conditions were performed using *two-tailed* paired-samples *t*-tests. However, due to the accuracy rates for the speech condition (almost 100%), comparisons including the *speech* condition were not considered for further analysis. A statistically significant increase in accuracy scores was observed for *auditory icons* ($M=0.98$, $SD =.23$) over *earcons* [$M=0.63$, $SD=.23$, $t(28)=-7.704$, $p<.001$].

Response times across the conditions are presented in Figure 5. No significant difference was found for the response times between the *earcon* and *auditory icon* conditions [$M = 297.74$, $SD = 902.20$, $t(28)=1.777$, $p=0.086$ (n.s)], indicating that the time taken to make an identification was the same regardless of whether the sound was an *earcon* or an *auditory icon*. Unsurprisingly, responses to speech stimuli were significantly faster than either earcons [$M = 1852.70$, $SD = 972.65$, $t(28) = 10.26$, $p<.001$] or auditory icons [$M = 1554.96$, $SD = 657.32$, $t(28) = 12.74$, $p<.001$]. Although participants made decisions that were faster and more accurate for every service with speech notifications, the exception to the rule was with the ‘call’ service, where the auditory icon scored higher. Similarly, although auditory icons outperformed earcons on average, earcons for SMS and instant messaging were more accurate.

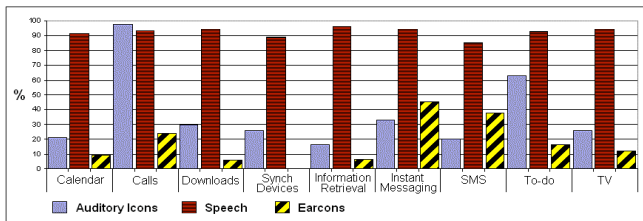


Figure 3. Mean scores of the three sound types for each

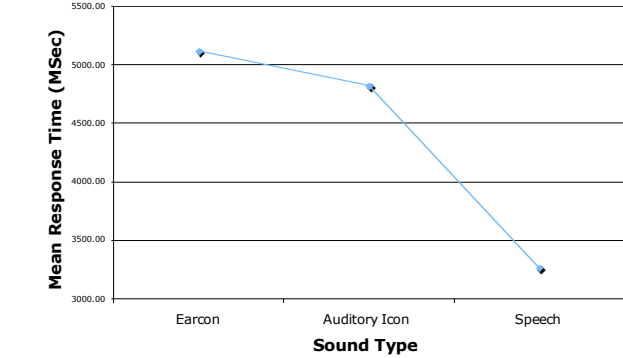
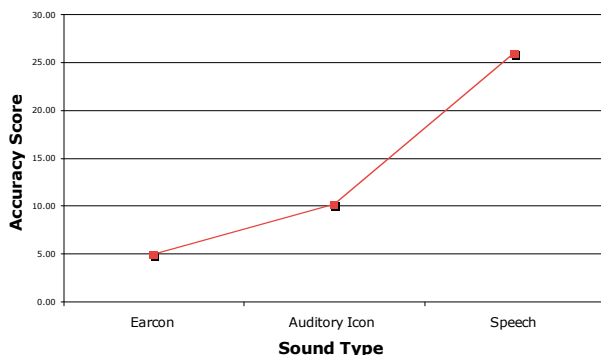


Figure 5. Mean response times across all sound type conditions

Figure 4. Mean Accuracy Scores for Group only and Target Sound across all Sound Type Conditions

4.5.2 Qualitative data

Overall, participants’ subjective results are in agreement with their performance. In the post-test questionnaire we asked participants to evaluate sound comprehension, their own personal liking for the sounds, and which notification type they would like to have on their phones. Also, they were asked to estimate how long it would take them to learn and memorise them if they were using them on their mobile device. Their responses were collected on Likert scales, which were then converted to a score out of 100. Results for each sound type are reported separately.

Speech

Speech notifications were (unsurprisingly) the easiest to understand. On average participants scored speech at 91% in terms of comprehension and almost 90% of them believed it would take them less than a week to learn and remember the speech associations.

Although participants rated how much they liked speech notifications at over 78%, there were mixed views on whether they would like to have them on their own mobile phones. Just over 1/3 of people would want to have them, while just over 1/3 were indifferent to this prospect (scoring 54% on average). Finally, over 40% of the participants reported finding speech notifications boring or annoying or expressed concern with regard to privacy.

Auditory Icons

Overall, participants rated auditory icon notifications lower than speech but higher than earcons. On ease of comprehension and preference, auditory icons scored exactly 50%, with equal number of participants liking or disliking them. As with speech, there were mixed views on whether they would like to have them on their phones, with an average score of 50%. On learning and

Table 4. Mean scores - subjective assessments of sound types

Scores (%)	Speech	Auditory Icons	Earcons
Comprehension	91	50	31
Preference	78	50	38
Would like to have	54	50	43
Perceived ease to learn	97	78	66

remembering, 83% believed that they would need no more than 2 weeks.

Earcons

Earcon notifications achieved the lowest scores with comprehension at 31% on average and preference at 38%. The perceived ease of learning also scored comparatively poorly (66% on average) with 10% of the participants estimating that they would never learn and retain them.

At the end of the questionnaire there were several comparative questions with regard to self-assessment of recognition rate, user preference in sound type, and potential usage. The results of these questions were generally in agreement with the actual performance. However, more participants chose earcons over auditory icons as a sound type they would like to use in the future (15 against 11). This contradicts their responses for each sound type separately, where auditory icons scored higher (see Table 4).

5. Discussion

Our evaluation found that auditory icons performed significantly better than earcons in terms of recognition accuracy exactly because auditory icons are usually richer in their semantics. Everyday sounds are more familiar to people, and they have already been associated with the specific context in which they normally arise. If auditory notifications are to take advantage of this pre-existing relationship, they need to represent a service that is relevant to that context. Users may then recognise these associations with minimum or no training at all, as they did in our experiment. We believe that everyday and widespread technology such as mobile devices should be designed to be as intuitive as possible with minimum user training requirements.

A further interesting result with regard to semantics was that 61% of all earcon notifications were assigned by participants to only three services: messaging, instant messaging and calls. Since the earcons used were designed to be distinguishable, we believe the reason these 3 services were associated with more than 1 earcon each is the pre-existing association between earcon sounds and these services. It is true that traditionally all notifications on mobile phones sound very similar to earcons (rather than speech or auditory icons) and mostly notify users of incoming messages or calls on mobile phones (i.e. ringtones) or instant messaging applications (such as MSN) on computers. Therefore, it is most likely that participants followed their intuition, according to which earcons are related to messaging services and calls. This preconception may have interfered with our participants' consideration of new associations between earcons and other services.

Supporting this suggestion, 'messaging' and 'instant messaging' were the only two services where earcons outperformed auditory icons. Although these earcons were not identical to the widespread notifications on current popular mobile phones, one could argue that their resemblance made these earcons more effective. Interestingly, such common everyday notifications (e.g. the classic Nokia sound for incoming sms) blur the boundaries between auditory icons and earcons. As the association is established through repetitive widespread usage, they are appropriated and cross over to the auditory icons arena. The same could be argued for the traditional telephone ring itself. When it was first introduced it was nothing more than a noise or arbitrary sound. Through the years it became synonymous with the telephone. Therefore, it is not surprising that we found certain

auditory icons to be very accurate (e.g. old style phone ringing for 'calls', and TV turning on with white noise for 'TV'). This further suggests that there are everyday sounds which, if utilised as auditory notifications, can create strong intuitive connection with certain mobile services.

Furthermore, the subjective data collected through the post-test questionnaire follow the same trend, with speech being the easiest to comprehend and most liked, followed by auditory icons and then earcons. However, these gaps in preference get smaller when users are asked to envision which notification type they would like to have on their device. It is interesting to note for example that the superiority of speech notifications (both in terms of our quantitative and qualitative data) was not reflected in users' expressed preferences. This might indicate that speech notifications are not be the best choice for audio notifications on mobile phones, despite their apparent superiority in conveying semantics. However, it is difficult to draw conclusions on user preference based on a single lab study. It is difficult for people to envision daily usage of these sound types, especially in the rich and complex audioscape of everyday life.

6. CONCLUSIONS/FUTURE WORK

In this paper we investigated the importance of intuitiveness in audio notifications for mobile services. First, we created a hierarchy of mobile services in order to reduce the potentially huge number of notifications to a much smaller number corresponding to clusters of similar services. Then, we experimentally tested the intuitiveness of the different audio notifications by asking users to guess the audio-service mappings. Our experimental hypothesis was supported, with auditory icons performing significantly better than earcons. Speech not surprisingly outperformed both non-speech sound types. The user preference and perceived ease of learning of the sounds followed the same trend, with speech preferred over auditory icons, and auditory icons preferred over earcons. However, answers on choice of sound type for actual potential use by users were inconclusive.

We argue that the success of auditory icons over earcons heavily depends on the success of the metaphors used. Regardless of sound type, associations were more successfully guessed when the mapping between service and notification was iconic (e.g. auditory icon for 'TV' service and earcon for 'Messaging' service). This suggests that pre-existing semantics should be utilised for intuitive mobile service notifications.

In future work, we intend to investigate whether the intuitiveness of mappings affects learnability and memorability over longer periods. Also, if both intuitive and arbitrary associations between services and audio notifications are learned, will one set be forgotten more quickly than the other? Furthermore, we are investigating the user preference outcomes of a longitudinal study, where participants spend days or weeks interacting with the same auditory notifications.

7. ACKNOWLEDGMENTS

We thank Vodafone Group R&D, Francis Binns, Mike Crocker, Daniel Goldstein, Iain Kingston, Andrew Shakespear, Erxiong Xu, Dalia Khader.

8. REFERENCES

- [1] Baber, C. and Noyes, J. M., Eds. (1993). *Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers*. Taylor & Francis, Inc.
- [2] Blattner, M. M., Papp, A. L. III, & Glinert, E. P. (1994). Sonic enhancement of two-dimensional graphic displays. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 447 - 470.
- [3] Blattner, M. M., Sumikawa, D. A., Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *SIGCHI Bull.* 21(1), pp. 123-124.
- [4] Block Jr F.E., N. L., Ballast B., (1999). Optimization of Alarms: A Study on Alarm Limits, Alarm Sounds, and False Alarms, Intended to Reduce Annoyance. *Journal of Clinical Monitoring and Computing*, 15 75-83.
- [5] Bonebright, T.L., & Nees, M.A. (2007). Memory for auditory icons and earcons with localization cues. *Proc. ICAD 2007 – Thirteenth Meeting of the International Conference on Auditory Display*, pp. 419-422.
- [6] Brewster, S. (2002). Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing*, 6(3), pp. 188-205.
- [7] Brewster, S.A. (1998). Using non-speech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction*, 5(2), pp 224-259.
- [8] Brewster, S. A., Wright, P. C., Edwards, A. D. N. (1993). An Evaluation of Earcons for use in Auditory Human-Computer Interfaces. *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp. 222-227.
- [9] Cohen, J. (1994). Monitoring Background Activities. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification and Auditory interfaces*, pp. 499-522.
- [10] Edworthy, J., Hards, R. (1999). Learning Auditory Warnings: The Effects of Sound Type, Verbal Labelling and Imagery on the Identification of Alarm Sounds. *International Journal of Industrial Ergonomics*, 24(6), pp. 603-618.
- [11] Gaver, W.W. (1997). Auditory Interfaces. *Handbook of Human-Computer Interaction*, Elsevier Science.
- [12] Gaver, W.W. (1989). The SonicFinder: An Interface That Uses Auditory Icons. *Human-Computer Interaction*, 4(1), pp. 67-94.
- [13] Gaver, W.W. (1986) Auditory Icons: Using Sound in Computer Interfaces. In *Human-Computer Interaction*, 2(2), pp. 167-177.
- [14] Geven, A., Sefelin, R., Tscheligi, M. (2006). Depth and Breadth away from the desktop - Optimal Information Hierarchies for Mobile Use, *Proc. MobileHCI*, pp. 157-164.
- [15] Graham, R. (1999). Use of Auditory Icons as Emergency Warnings: Evaluation within a Vehicle Collision Avoidance Application. *Ergonomics*, 42 (9), pp. 1233-1248.
- [16] Hickling, E. M. (1994). Ergonomics and engineering aspects of designing an alarm system for a modern nuclear power plant. In *Human Factors in Alarm Design*, N. Stanton, Ed. Taylor & Francis, Bristol, PA, 165-178.
- [17] Isaacs, E., Walendowski, A., Ranganathan, D. (2002) Hubbub: A Sound-Enhanced Mobile Instant Messenger that Supports Awareness and Opportunistic Interactions. *Proc. Computer-Human Interaction (CHI)*, pp. 179-186.
- [18] Keller, P., Stevens, C. (2004). Meaning from Environmental Sounds: Types of Signal-Referent Relations and their Effect on Recognizing Auditory Icons. *J Exp Psychol Appl*, 10(1), pp. 3-12.
- [19] Leplatre, G. and Brewster, S.A. (2000). Designing non-speech sounds to support navigation in mobile phone menus. In, Cook, P.R., Eds. *6th International Conference on Auditory Display (ICAD)*, pp. 190-199.
- [20] Leung, Y. K., Smith, S., Parker, S., & Martin, R. (1997) Learning and Retention of Auditory Warnings. *Proc. 3rd International Conference on Auditory Display*.
- [21] Nass, C. and Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM* 43, 9, pp. 36-43.
- [22] Patterson, R.D., Mayfield, T.F. (1990). Auditory Warning Sounds in the Work Environment [and Discussion]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 327(1241), Human Factors in Hazardous Situations, pp. 485-492.
- [23] Rigas, D., Alty, J. (2005). The rising pitch metaphor: an empirical study, *International Journal of Human-Computer Studies*, Volume 62, Issue 1, January 2005, pp 1-20.
- [24] Sorkin, R. D. (1987). Design of auditory and tactile displays. In G. Salvendy (Ed.), *Handbook of human factors* (pp.549-576). New York: Wiley & Sons.
- [25] Stevens, R., Brewster, S., Wright, P., & Edwards, A. (1994). Design and evaluation of an auditory glance at algebra for blind readers. In G. Kramer and S. Smith (eds.), *Proc 2nd International Conference on Auditory Display*, pp. 21 - 30.
- [26] Ulfvengren, P. (2003). Design of Natural Warning Sounds in Human-Machine Systems. PhD thesis, Stockholm, Sweden.
- [27] Walker, B. N., Kramer, G. (2005). Mappings and Metaphors in Auditory Displays: An Experimental Assessment. *ACM Transactions on Applied Perception* 2(4), pp. 407-412.