Musipass: Authenticating Me Softly with "My" Song

Marcia Gibson
University of Bedfordshire
Park Square
Luton, United Kingdom
Marcia.Gibson@beds.ac.uk

Karen Renaud
University of Glasgow
17 Lilybank Gardens
Glasgow, United Kingdom
karen@dcs.gla.ac.uk

Marc Conrad
University of Bedfordshire
Park Square
Luton, United Kingdom
Marc.Conrad@beds.ac.uk

Carsten Maple
University of Bedfordshire
Park Square
Luton, United Kingdom
Carsten.Maple@beds.ac.uk

ABSTRACT

The modern world increasingly requires us to prove our identity. When this has to be done remotely, as is the case when people make use of web sites, the most popular technique is the password. Unfortunately the profusion of web sites and the associated passwords reduces their efficacy and puts severe strain on users' limited cognitive resources. There is clearly a need for some creativity in terms of providing viable alternatives to passwords. This paper reports experiences of the use of a musical password, one composed of melodies instead of alphanumerics. Music is universal all over the globe and humans have superior memory for music.

We report here on the evaluation of a prototype of such a musical password system, which demonstrates superior memorability and acceptance by users and is particularly useful to those with impaired memory or cognitive function.

Categories and Subject Descriptors

D.4.6, H.1.2 [**Authentication, User Factors**]

General Terms

Design, Experimentation, Human Factors, Security

Keywords

Authentication, musical password

1. INTRODUCTION

People have to authenticate themselves on the Web many times a day. This is most often achieved by means of a shared secret, termed a password. Unfortunately, the sheer numbers of secret passwords people are expected to remember is placing them under undue pressure, and they are responding by behaving insecurely: writing down or sharing "secrets", using personal details or reusing passwords for various systems [23]. The main problem is that passwords rely on uncued recall, which is increasingly difficult for humans as they age, or for those with memory problems.

It would therefore be beneficial to use some alternative technique to strengthen the knowledge-based mechanism and to make it easier for people to remember their secrets. Ideally, the alternatives should rely on recognition rather than recall, to reduce cognitive load.

A number of alternative systems have been trialed. Some systems rely on recognition of images from a challenge set [17, 10, 44, 22, 57]. Since humans have superior picture memory these mechanisms have the potential to perform better than traditional passwords and are an area of promising research.

Other systems rely on memory of an association. There is evidence that passwords based on associative memory are more memorable and harder for other people to guess [42]. Associative passwords have been trialed for sound clips [33] and for other words [51, 42]. Word association works well, but is very time consuming, both at enrollment and authentication. Liddell *et al.* [33] tested the association between sound and image. Unfortunately participants merely memorized the image and did not listen to the sound or use it as a cue.

Both Conrad *et al.* [14] and Chiasson *et al.* [13] have argued for the use of music in authentication. Music is arguably universal: there is no doubt that humans everywhere, irrespective of culture or creed, enjoy listening to music [45]. Indeed, Peretz et al. [40] suggest that our auditory pathways have been specifically wired to process music-related stimuli. Another suggestion comes from Scherer and Zentner [46]. They argue that music expresses emotion, or that we express emotion by means of music. Reimer argues that wherever humans are found, there you can find music too [43], stating that the value of music is in "enhancing the depth, quality, scope, and intensity of inner human experience in ways particular to how music operates; ways that distinguish music from other human endeavors". Luckily people do not need to have musical training in order to have some musical capability: it appears to be inherent. Bigand and and Poulin-Charronnat [7] investigated capacity for processing music amongst both trained musicians and untrained people. They found that the human brain is predisposed to process music and that this is triggered by everyday exposure to music, which occurs over the person's lifetime, and is not due to any training the person may have received. This means that a musical password would not require any specific musical training, which enhances the accessibility thereof — all that is required is the ability to hear.

NSPW'09, September 8–11, 2009, Oxford, United Kingdom

This paper presents a system called *Musipass*, which uses music to authenticate users. The general idea is that people choose a series of sound clips from a set of options at enrollment, which together form their "password". They then choose "their" clips from a larger group at authentication. The excess clips are referred to as *distractors*. A number of issues need to be resolved if this mechanism is to be viable. These include the mechanism for choosing the user's clips, the length of the clips, the number of clips to be used, and how to present these to the person being enrolled and/or authenticated.

Section 2 provides a justification for the use of music in authentication and explains what kind of sound clips have to be used in order to maximize their efficacy in this context. The section concludes by provising a justification for the use of alternative authentication mechanisms such as Musipass.

Section 3 details the design and implementation of the authentication software system. Section 4 provides a security analysis, Section 5 discusses a user study of our implementation, Section 6 presents the results and Section 7 discusses the results. Section 8 concludes.

2. EFFICACY

In contemplating any knowledge-based authentication mechanism, we have to consider two primary aspects: memorability and security. Memorability is important because the knowledge is intended to be secret and therefore should not be physically recorded. The more memorable it is, the less need there is for a tangible record of the secret to be kept. The security needs require the mechanism to ensure that only authorized users gain access to the system. The first two subsections discuss memorability and security aspects of music in authentication, and the third subsection presents implementation decisions made for the Musipass system, based on the presented principles.

2.1 Memorability

The core issue here, in terms of using music to authenticate people, is: "How well do people remember melodies?".

Jäncke [29] argues that music evokes emotions and, due to this, hearing music will tend to lead to formation of memories. These memories could be related to the music itself, or to particular episodes related to the music. Eschrich *et al.* [20] tested both arousal rates and emotion when participants listened to 20-30 second music clips. They showed that emotion, but not arousal, was related to the memory of the sound clips. Furthermore, researchers have found that music plays a role in the laying down of autobiographical memories [28]. These memories can then be evoked when the music is heard again.

It is a well established fact that advertisements using music are better remembered than adverts without music [52, 2]. Scherer and Zentner [46] point out that music is involved in memory formation for two reasons. The first is that music often accompanies emotionally charged events such as religious ceremonies, weddings, funerals and other celebrations. The second is that music, they argue, could be processed at a lower level of the brain than other semantic memories [32] and that musical memories could thus be more resistant to interference than other memories.

In conclusion, like images, music may also have memorial advantages that we can tap, and it is worth experimenting with its use in authentication.

2.2 Security

The *security* of authentication mechanisms can be judged in terms of *guessability, observability, recordability* [16].

2.2.1 Guessability

If we want to use music in an authentication system, we have to ensure that their secret sound clips are not *guessable*. People are usually open about their music preferences. Some musical preferences are fairly predictable [26, 27]. Therefore, it is infeasible to allow users to upload their own favorite sound clips: we will have to use a fixed set of clips and ask users to choose from these. This might have an impact on the memorability of the clips but since the *raison d'être* of the musical password is to implement security, we cannot be too flexible in this respect.

2.2.2 Observability

Observability of musical passwords is potentially less of a problem than it is for image-based systems. If the user is not alone, he or she merely inserts headphones into the computer, and anyone observing the user interacting with the system will get no clues as to the secret clips. With the increasing popularity of MP3 players and mobile phones that play music, many people carry earphones with them. Figure 1 shows that ATM machine producers are beginning to take this into account in their designs.

If the playback volume of music clips is loud enough, a person in the close vicinity could still listen in as they are selected. Many headphone manufacturers design earphones to reduce audio leakage, since many people wish to listen in public areas without intruding on others. Certainly when we are exposed to music suddenly (for example when we first turn our car on in the morning), the volume of the radio can be uncomfortable to bear, although at the time we were listening to it, our perception had adjusted for the volume. By using multiple short clips separated by a short period of silence, instead of one continuous clip, we can recreate this effect. Playback of discrete clips at a high volume setting would at the very least be unnerving, if not uncomfortable to the majority of users.

Still, attacks utilizing specialized hardware, such as directional microphones cannot be categorically ruled out. Further research is required into the possibility and success rates of these. The requirement for additional hardware would make them more difficult to carry out than would a simple glancing over of keys during alphanumeric password entry.

Additionally, a scheme wherein a user needs to tell an authentication system that he or she recognizes a sound, would inherently require some mapping of the sound to a control that the user can interact with so as to signal recognition. Thus musical passwords might be observable by ignoring the musical tunes altogether, and simply recreating the sequence of user interactions. In a system where the user selects from a number of distractor clips, this problem can be remedied by "shuffling" the placement of the clips in the interface with distractor clips between log in sessions. Shuffling the placement rules out the possibility for users to make selections using neuromuscular, or "muscle" memory. However, if the clips themselves are as memorable as we hope, this should not be necessary to elicit successful authentication.

2.2.3 Recordability

With respect to *recordability*, the musical password is neither stronger nor weaker than traditional knowledge-based mechanisms. Someone who wishes to write down the titles of his or her secret sound clips can do so. However, if the enhanced memorability of the sound clips becomes evident, this might make it less necessary to write them down.

Figure 1: ATM taking ear phones

2.3 Implementation Choices

We need to deliberate about the characteristics of the sound clips to be used in our Musipass system. The following aspects are important:

- *Familiarity*: The first decision we made was that the sound clips would come from melodies that were familiar to most people — so-called *old familiar* tunes. This was based on the findings of researchers which show that recognition of highly familiar melodies will tend to be both pleasurable and effortless.

  In 1932 Smith [50] wrote a paper titled "The Pleasures of Recognition". He points out that we gain great pleasure from the familiarity of things, and that this lies behind our inertia for change. In discussing memory for tunes, he argues that tunes *"seem, by dint of frequent repetition, to hollow out a sort of channel in the memory, a channel that never loses its shape or contours"* (p82). Smith argues that popular music is characterized by the fact that it has fragments of easily recognizable tunes in it and concludes that humans crave something to recognize and that composers would do well to heed this craving. Brain scans show that when people listen to music a number of different regions of the brain become active [19]. When people like a particular piece of music, Blood and Zatorre [8] found that the brain areas associated with pleasure are activated. This finding provides physical confirmation of Smith's arguments. Lazarsfeld and Field [31] carried out an experiment to determine preferences for one of three types of music: popular, classical and *old familiar*. He surveyed 2200 people and found that 76% liked to listen to music and, of those, 16% favored familiar music. Lower percentages preferred the other music types. Once again there is clear evidence of preference for the familiar.

  There is another advantage in using familiar well-known tunes. It is more likely that people will have been exposed to them if they have been around long enough to become familiar, since they will be played on the radio and television at regular intervals, and possibly in restaurants [36] and stores [45] as well. Research has shown that background music has an impact on behavior without people being aware of it. People may be unaware, at least consciously, of all their different exposures to music, but such exposures will nevertheless serve to strengthen the memory trace related to that particular melody. Furthermore, Willems *et al.* [59] found that familiarity influences preference, especially in a recognition-based task.

- *Length*: The next decision to be made relates to the length of the sound clips. Bella *et al.* [4] carried out experiments investigating the process of melody recognition. They found that a feeling of familiarity for a familiar melody could be experienced after only 3 to 6 notes and recognition could occur after a further 2 notes were heard. In general they also found that high-familiarity melodies were recognized faster than low familiarity melodies. Unfamiliar melodies could only be judged after far more notes had been heard.

- *Rhythm*: When one considers sound clips, there are many ways of analyzing them. For example, one can consider the rhythm, tempo, contour, timbre, loudness, reverberation and melody of the piece.

  Wells *et al.* [58] argue that the kind of rhythmic music that encourages people to tap their feet or snap their fingers has enhanced memorability. Mélen and Deliége [34] carried out an experiment to determine what it was that caused melodies to be recognized. They compared melodic contour, harmonic structure and local surface cues for efficacy in assisting recognition. They transformed a melody using reduction and two different kinds of rhythmic transformations. They found that recognition was better for the rhythmic transformations but this only happened when the local surface cues were preserved in the transformation. They concluded that melody recognition was assisted by particular surface cues in the melody which seem to be encoded when the music is heard, and these are then used to elicit recognition when it is heard again. We should therefore exploit this proven memorability by offering our participants a range of different clips with an easily-recognized rhythm.

- *Vocal*: The next decision relates to whether the sound clips will feature voice or be merely instrumental. Vocal music is more memorable than instrumental music for the average listener, as shown in research into the use of music in advertising [5, 56, 48].

  Researchers argue that the lyric and the tune of a melody in a vocal are processed independently in the brain [6]. However, there are strong connections between the parts of the brain processing them [41, 24]. Crowder and Repp [15] showed that the lyrics and melody of a song could cue each other. These effects only exhibit when the music is familiar to the listener.

- *Hooks*: Burns [12] quotes Monaco and Riordan's [38] definition of a hook as: "*a musical or lyrical phrase that stands out and is easily remembered*" (p178). A hook could consist of a series of notes, could include some repetition and is the part of the song that stays in the person's memory. Examples are: "With a little help from my friends", or "Michelle, my belle". The use of hook sound clips would obviously be a good way of increasing the possibility that the sound clip will be recognized.

Musipass used short sound clips which had as many of the above characteristics as possible. This should, based on the literature, maximize the efficacy of the sound clips in terms of memorability.

A number of security-related implementation choices also need to be made:

- How many sound clips will be displayed at a time? Here we have a balance between convenience and security. Working through all the different options is far more time-consuming than for images, which are taken in at a glance. The user has to listen sequentially to each clip in order to identify his or her clip. On the other hand, we have to offer a decent number of alternatives in order to provide an acceptable level of security.

- How many sets of sound clips will the user progress through to be authenticated? This has a direct bearing on the security and convenience of the mechanism. An aligned question is whether we include decoy screens, which display a full set of distractors if a user makes an incorrect choice at any stage during authentication.

The Musipass system will display nine icons at a time, each representing one sound clip, and work through four screens to authenticate. Decoy screens will be deployed to provide error feedback to the user without providing additional clues to any would-be intruder. Asking users to select a relatively short password sequence from a small set of distractors will allow us to gain a reliable picture of how they react to the idea of authenticating with music, which is of primary importance at this early stage in our investigations. In addition, should users react positively to the system, we will have an established baseline with which we can compare future implementations that can be more tailored toward providing increased password space.

The final decisions relate to user convenience:

- How will the user activate the sound clip in order to hear it? Either users have to click on an icon, or move the mouse over it. The latter is advisable since it requires less effort.

- Will we include a training session after enrollment? Use of a training section should enhance memorability, but will make enrollment far more time-consuming. On the other hand, the ease of recognition for familiar tunes might make training unnecessary.

In Musipass, users will hear the clips when they move their mouse over the icon. Since we we plan to compare with traditional password systems, on which users will typically "train" by re-entering the password they select, our users will also work through a training session at enrolment. If people remember melodies by their physical nature in addition to the associations triggered while listening, we should focus training on both repetition and association. Our results can then be compared in future studies where training has not been implemented.

2.4 Why Musical Passwords?

There are large numbers of users who have great difficulties with alphanumeric passwords. Some have disabilities such as dyslexia [9] or dyspraxia [18]. The former leads to unpredictable spelling. Dyspraxic users have difficulties in sequencing of numbers and letters.

Younger users may have developmental or language difficulties which makes entering an obfuscated password challenging and frustrating [47]. On the other hand, people are living longer than before and the elderly sector of society is growing yearly. It is well known that memory is less reliable as we age, especially when it comes to retaining newly learnt information [49]. Elderly users consequently find it very difficult to remember passwords. The use of password alternatives is also particularly useful where the intended users of the system are illiterate [30], have poor reading skills [47], or use a different alphabet [35].

The problems these users experience are due to the fact that they have to recall their password without a cue and enter it correctly without feedback. The alphanumeric password is clearly "user-hostile" in many situations [3].

Whilst biometrics and token-based systems, when implemented thoughtfully, can provide sufficient levels of security and convenience, they clearly generate new dependencies and possible exclusions of minorities. Contributing factors include, cost of purchase, lack of installation and operating knowledge or, specifically in the case of biometrics, incompatibility between the characteristics to be measured and those of the person requesting access. In addition, instantaneous access is often not achievable, until the hardware is first forwarded (possibly collected by hand or delivered by mail) to the person requesting access. If we wish to promote inclusion in society for members who are economically, physically, cognitively, and technologically impaired, then it seems we should certainly be considering software based alternatives, particularly with the advent of initiatives such as e-voting and e-banking.

Alternative, software based approaches aim to address the problems of the traditional password. Firstly they provide the user with cues and typically require recognition or cued recall rather than free recall. Secondly, these systems do not require correct entry of a precise alphanumeric string, requiring only the simple click of a mouse.

A variety of image-based password mechanisms have been proposed [17, 10, 44, 22, 57]. Schemes wherein the user typically is required to identify a selection of previously chosen images from a larger repository of possible options are conceptually closest to our envisaged sound-based design. Depending upon implementation, image-based schemes are often shown to provide greater resistance to statistical attacks than an equivalent length text password, and can show a high rate of successful registrations and logins. However, they do encompass inherent drawbacks of their own, in that they cannot be used by those who are blind and are likely to be difficult for those who are partially sighted [21]. Image-based passwords are restricted in that they cannot be used in situations where it is not possible to use a screen, such as when authentication is required over the telephone [14].

Musical passwords can potentially retain the memorial and ease of use advantages of image-based schemes [14], whilst providing a solution where a graphical user interface is not available, or when the user is visually or otherwise impaired. Natural areas for system deployment hence include,

- **Mobile banking:** Particularly in developing nations, where internet access is not always so easy to come by [37], and handsets may not be so advanced [1], or where many members of the community share access to a limited number of phones [54].

- **Internet web sites and ATM interfaces:** Especially for those users who otherwise experience difficulties gaining access with passwords and PINS and who are visually or otherwise impaired.

3. MUSIPASS

3.1 Enrolling

During the enrollment phase, users work through a four-stage process. Each stage being carried out on a different "setup screen".

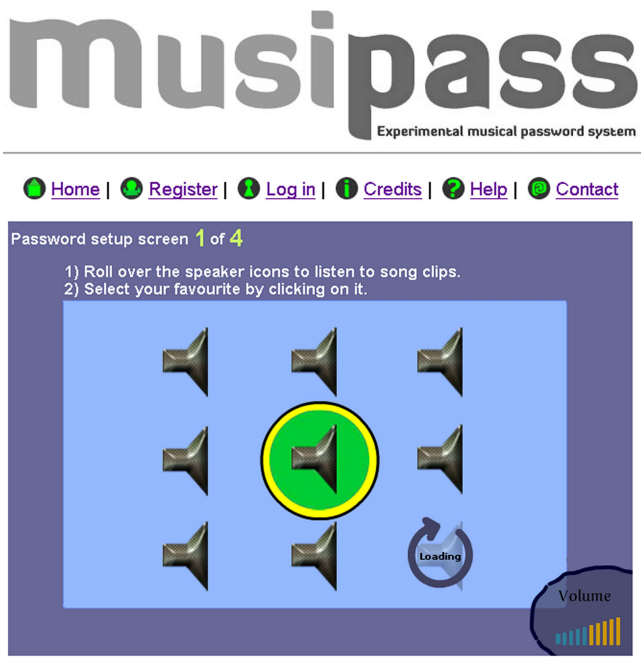

Figure 2: Enrolling with Musipass

Figure 3: Training with Musipass

An example of one of the setup screens is shown in Figure 2. Each screen displays icons representing 9 different sound clips. The user hears the sound by floating their mouse cursor over the icon. They can listen to clips as many times as they like and then choose *their* clip by clicking on it. The server itself holds 201 sound clips, although only 36 are used for each user's password alphabet. Due to the randomization in selecting the alphabet, these clips would, in general differ somewhat between users, thus increasing the global space.

Once the user has chosen one sound clip from each of the four setup screens, these become their secret password and they progress on to a *training session*. During this session they listen to their password clips again, and are asked to enter some text to describe each one, as shown in Figure 3. This is done to ensure that the memory trace for the sound clip is well established and will be less likely to fade.

Training is followed by an authentication stage which requires the user to step through 4 screens. Each will have an icon representing one of their chosen clips and also the 8 randomly chosen *distractor* clips.

3.2 Authentication

At authentication, the user provides an identification email address which is used to instantiate the system into loading the correct password alphabet. They then proceed to step through 4 screens, almost identical to the ones shown in the enrollment phase, but with the addition of a *recovery* button, as shown in Figure 4. The placement of the sound icons is randomized to guard against shoulder surfing attacks. If at any stage during the authentication task an error is made the system will proceed to display only decoy sound clips in subsequent screens (when this happens the screens are referred to as "decoy screens"), the unfamiliarity providing useful feedback for the authentic account owner. The sounds contained in a given user's decoy screens are selected arbitrarily from the remaining 165 sounds not already in use as part of the password alphabet. The set of decoy sounds selected for each decoy screen does not change between log in attempts. It should be noted that use of decoy clips can lead to the possibility of an intersection attack if not implemented securely, further discussion of such an attack and details for a secure implementation can be found in Section 4.3.

In order to guard against timing attacks that might take advantage of the caching of sound clips in a web browser, which would result in decoy music clips taking longer to load than authentic sets, an override mechanism was deployed so that no sounds were cached locally during the experiment. Although this meant that users would wait slightly longer to hear sounds, it put less strain on the (shared) web server than would taking the opposite approach, which would involve downloading a potentially very large corpus of information onto the client machine.

The recovery button was provided to support legitimate users should they suspect that a selection error had been made during interactions with one of the previous log in screens.

If the user clicks the recovery button, he or she can start the authentication process again. Dhamija and Perrig's [17] Déjà vu system implemented this kind of recovery option, and found that users were able to recover by using this button after they had made a mistake. If, as we hope, melody memory is as strong, or stronger than, image memory, then users should easily detect the absence of one of *their* melodies and recover. Attackers, on the other hand, should continue undeterred, thereafter to be denied access to the system.

3.3 Replacement

If a user has forgotten which tunes were chosen during enrollment, or suspects that the secret key has been discovered, he or she can request a re-registration. Once this has been approved, the user steps through the enrollment process once again, choosing a new set of sound clips. It should be noted that this functionality was disabled for our evaluation system so that we could investigate recognition and general reactions, without at this stage concerning ourselves with the issue of memory interference.

Figure 4: Authenticating with Musipass

One easily implementable mechanism to minimize interference effects is to ensure that the re-registering user is not offered the same sound clips as those previously chosen. A particular strength of a sound-based system is that the dictionary is potentially very large and thus excluding a particular set of sounds from those offered to the user is not problematical.

4. SECURITY ANALYSIS

We consider the following typical scenarios where an attacker Mallory attempts to access Alice's account

4.1 Brute force attack (online)

Mallory knows Alice's username and enters it at the interface. The system responds by returning the first subset of Alice's individual log in alphabet. From here Mallory selects sounds at random until the correct password sequence is obtained.

The number of possible permutations in Alice's alphabet is q^r, where q is the number of sounds contained in the alphabet and r is the length of the password sequence.

In our implemented system, users choose password clips over four selection screens, each offering a choice of nine clips, therefore $q = 9$ and $r = 4$ this gives a total of 6561 permutations and means the average brute force attack would elicit the correct sequence in 6561/2 attempts. This is lower even than the space offered by a four digit PIN over an alphabet of ten digits, although for practical purposes we can mitigate this risk by blocking authentication after a given number of attempts is exceeded.

This approach does not rule out a so called, "low and slow" attack, where Mallory circumvents the lock out policy by distributing his guesses over a number of user accounts (i.e. he isn't concerned with whose account he accesses, he is only concerned that he will gain access). As countermeasure, we can increase the length of the password sequence and/or the number of distractors provided, until the average number of guesses required is substantially higher than the available number of accounts we want to support in the system, multiplied by the number of attempts we allow during authentication, although further research is required to ascertain whether and how password memorability and level of interference with distractors would be affected.

4.2 Brute force attack (offline)

Internally to the Musipass system, Alice's musical password is represented as a character string. Let us assume that the string can be encoded or hashed in some way so as to obfuscate its original form to anyone viewing its encoded representation. We will refer to this as the password "hash" (since this is conceptually similar to what happens with traditional passwords), although specifics and methods for the encoding a musical alphabet is still a current area of research. Let us then assume that attacker Mallory gains access to Alice's hashed musical password string. He also has access to the entire list of sounds held on the server and their cleartext equivalents. Mallory then systematically creates password sequences using the list of sounds, applies the same encoding function implemented in the system to their cleartext equivalents and compares the result with the hashed version of Alice's password. Since we prevent the user from selecting the same clips more than once, the maximum possible number of permutations Mallory can create based on the number of sounds in the list is:

$$permutations = \frac{q!}{(q-r)!}$$

In our implemented system the database contains $q = 201$ sounds, of which $r = 4$ are selected. The total number of permutations is hence, $q!/(q-r)! = 1,583,960,400$. This is approximately equivalent to a 5 character traditional password over an alphabet of 62 letters (mixed case letters and digits a-z, A-Z and 0-9).

However, if Mallory has not been exposed to an enrolment sequence for the system, or because the length of the password sequence may vary between users, Mallory may not be able to extrapolate the length of the password based upon its encoded string representation. This is especially so, as most hashing algorithms return a string of a set size regardless of the length of the input string. In this case, Mallory would need to set a value for the maximum password length he would like to compute, and then search exhaustively through them (e.g. passwords of length 1, followed by passwords of length 2, followed by passwords of length 3...) until he finds the correct sequence.

The following equation describes how many permutations are possible in this scenario, where r_{max} is the maximum length of password sequence Mallory decides to test.

For Musipass passwords (with repetition disallowed)

$$\text{permutations} = \sum_{r=1}^{r_{max}} \frac{q!}{(q-r)!}$$

For traditional passwords (with repetition allowed)

$$\text{permutations} = \sum_{r=1}^{r_{max}} q^r$$

Here we see that a potentially much larger space must be searched depending on the value of r. For demonstration purposes, we can continue the example, and imagine Mallory sets length r_{max} = 4, with the actual password length, $r = 4$ in both the traditional and musical password schemes. The number of possible permutations in Musipass is then, 1,640,402,004 compared to that of the traditional password, 15,018,570. Regardless of whether Mallory knows the length or not, we can conclude that a password created in Musipass is more secure against offline brute force attacks than a tra-

ditional password of equal length, due to the inherent strength afforded by its large alphabet.

4.3 Intersection attack

Mallory enters Alice's username at the interface and is presented with the first subset of her password alphabet. Mallory then selects each song clip in that subset once, making a note of the clips that appear on the subsequent screen, each time clicking the recovery button to restart the authentication attempt. Due to the use of decoy screens, it is possible for Mallory to identify the correct password element if, on selection, it leads to a screen that contains a different set of song clips than the ones returned by the other non-password elements. This attack possible due to the existence of a M:1 mapping between distractor clips and decoy screens. It has a recognizable signature, and hence measures could be put into place to monitor for and safeguard against it. This signature would take the following form:

In each of the first three stages (screens) of authentication: A minimum of 2 distractor sounds selected, in addition to the password element, with each of the selections separated by a restart of the authentication attempt.

In our testing of the Musipass prototype, discussed in Section 5, the maximum number of restarts taken by any given participant who was presented with decoy screens during an authentication attempt was 2 (3 users) over both phases of testing. Clearly participants did not attempt to attack the system in this way.

Although monitoring might be used to thwart this attack, it can also lead to false positives, where an authentic user takes the same path through the system as described, during authentication. It is therefore advisable to remove the vulnerability altogether, by removing the M:1 mapping between distractor clips and decoy screens, replacing them with 1:1 relationships (i.e. so that each clip, on selection, leads to a unique screen). Implementation could be as follows:

We store the catalog of music clips $U = \{u_1, u_2, ...u_n\}$ on a trusted server, where u_k is an individual music clip, $k = 1, 2, ...n$. From this we derive individual user alphabet $A \subset U$ which is separated into two subsets, one containing log in clips, that will be presented during a normal authentication session, $L \subset A$ and the other consisting of "decoy" clips $D \subseteq (A \setminus L)$ that can be presented after a selection error is made.

L is then split for sequential presentation, in Musipass this takes the form $M = \{N_1, N_2, N_3, N_4\} \subseteq L$ since authentication is achieved over four screens. All members of M contain nine music clip elements, of which one is selected to become part of the password sequence, the rest are classified as distractors. $N_r \subset L$, where $r = 1, 2, 3, 4$ and $\sum_r N_r \subseteq L$. We refer to the set of user selected password elements as $P = \{p_1, p_2, p_3, p_4\}$. $P \subset M$, moreover, $p_r \in N_r$.

We create a "decision tree", where if the correct password element is picked in each N, the authentication attempt is successful. An example N is $N_1 = \{p_1, n_1, n_2, ..., n_8\}$ for each correct password clip, p_1 that is selected, a password screen N_2 is presented.

For each n, distractor clip, there is a mapping to a specific, unique, decoy screen i.e. from the first log in screen there is a possibility of 8 unique decoy screens being returned, each containing 9 music clips that are not in the set M. The total number of clips contained in unique decoy screens is hence, in the 1^{st} order decoy screens, $3x(8x9)$, in the 2^{nd} order decoy screens, $(2x(8x9))x9$, and in the 3^{rd} order decoy screens, $(((1x(8x9))x9))x9$. This gives the total number of decoy screen clips $= 7344$.

We did not implement this 1:1 mapped design in our prototype, since it would require a large set of song clips, for which we also wanted to assess relative popularity as part of our analysis into vulnerability to dictionary attacks, (next section). Using such a large number of clips would have required us to attract an infeasibly large number of participants in order to gain accurate results.

4.4 Dictionary attack

Dictionary attacks utilize non-standard frequencies in the distribution of selected passwords over the available space. In a text-based system, instead of trying every possible combination of characters, attacker Mallory uses a list of historically common password sequences and submits these to the system in an attempt to elicit access to Alice's account.

The principal requisite for a successful attack of this type is that non-standard frequencies must exist in the passwords selected (i.e. there needs to exist common passwords). The strength of a musical password against an attack of this type then, is a result of the relative popularity of the song clips used to form the alphabet. If a small group of sound clips *are* popular for users, then it is likely that they would be chosen as password "letters" more often, in turn increasing the likelihood that common passwords will be created in the system and therefore making it more vulnerable to entry through this attack.

If we look to the example of text-based schemes, we are reminded that the reason users choose passwords that are vulnerable in this way, is because strings that are meaningful, that contain patterns, or that can be pronounced, intrinsically allow for the easy creation of cues that can assist recall; while wholly arbitrary character sequences do not. We hope that as Musipass relies on clip recognition and not recall, the presented sounds themselves will act as external cues for the user, and hence creation of additional cues will be unnecessary. We can test this hypothesis using data about the password selections our participants made during testing (full discussion of user tests and results can be found in Sections 5 and 6).

We can rate the popularity of sound clips by assigning a statistical significance level to each, based upon its number of appearances during enrollment and the number of times it was selected to become part of a password sequence. We then assess each sound by defining a null hypothesis, the hypothesis to be tested for rejection, and its alternative.

$$\begin{aligned} \text{Alternative hypothesis, } H_A &\equiv \text{Sound is popular} \\ \text{True null hypothesis, } H_T &\equiv \text{Sound is unpopular} \\ \text{Effective null hypothesis, } H_0 &\equiv \text{Sound is not popular or unpopular} \end{aligned}$$

We then apply the following tests:

$$H_0 : \mu_{\text{popularity}} = \mu_{\text{binomial}}$$

$$H_T : \mu_{\text{popularity}} \leq \mu_{\text{binomial}}$$

If both the true null hypothesis and the effective null hypothesis are rejected, then the alternative hypothesis is accepted (i.e. the sound is said to be "popular" for our users). $\mu_{\text{binomial}} = a\rho$ is defined by a, the number of times the specific sound appears and $\rho = \frac{\bar{\mu}_{\text{popularity}}}{\bar{a}}$ ($\bar{\mu}_{\text{popularity}}$, the averaged amount of times any sound is chosen and $\bar{a}$ is the averaged amount of times any sound appears). $\mu_{\text{popularity}}$ is the number of times a specific sound is chosen and μ_{binomial} is the binomial average for the specific sound. In deciding to accept or reject the effective null hypothesis, we will use a critical value, which we will take to be 5%, (specifically 0.05) and use this as a tolerance threshold for acceptance.

After applying this to our data, we found that 62% of the alphabet was neither popular or unpopular and 36% were biased in that they were either too popular or unpopular (15% and 21% respectively).

In practice user selected traditional passwords tend to utilize a much smaller subset of the space they can theoretically provide, although this alphabet is available to all of the users, all of the time. When we scale our figures over the 36 alphabet sounds a user in our system would be exposed to during enrollment, we would expect, on average to see around 5 popular sounds to be included. However in order to be able to form the sounds into a password, at least one of them would need to appear on each selection screen, the chance of this happening is 27% (104,976 of 272,016 possible permutations). This is seemingly why, after comparing the passwords created in the system (133 accounts) we can confirm that no identical passwords were selected (i.e. in practice there were no common passwords selected by users).

Even so, if it is possible to predict which songs are more popular than others in this context, it may be possible to reduce the bias in future systems, this, along with any ramifications, is an interesting area for further research.

Finally, regardless of whether or not common passwords are selected by users, an important aspect of our design is that each Musipass system can be populated with an individual sound clip alphabet. We could altogether do away with the practice of attackers employing standard password cracking dictionaries to gain access, since any dictionary created would only be of use against the system it was originally created from. This is because the letters used to create a commonly selected password sequence therein, would not be present in another system.

4.5 Prediction attack

Attacker Mallory comes to know Alice's musical preferences along with her username. He proceeds to enter the username at the interface and the system responds with the first 9 clips from Alice's password alphabet. Mallory then attempts to predict Alice's clips given her taste, in an attempt to gain access.

There is evidence of people having particular music preferences [46, 25, 11] and by way of a countermeasure, it therefore reasonable to ask people to choose one out of a set of proffered sounds. It might still be possible for guessable passwords to be selected in the scheme, and it is therefore important to ensure that the overall number and diversity of clips stored globally in the system is large enough to support the number of sounds implemented in individualized alphabets without too much repetition of similar artists and genres appearing on sequential password selection screens. A possible solution (not implemented due to time restrictions) might be to ask the user to provide details about their preference during enrollment. If the user specifies for example that he or she mostly enjoys ballads, it might be possible to populate her entire alphabet with clips belonging to that category. This may however, also result in a reduction of memorability. On the other hand, if the user is familiar with the song clips presented, by perhaps listening to them regularly, memorability might be enhanced, therefore this is an area we must leave open to further research and debate.

5. TESTING MUSIPASS

We developed a prototype of the Musipass system, embedding a Flash application within a Web page. Using Flash enabled rapid prototype development and platform independence.

The experiment was advertised via email and the Facebook social networking site, both directly to personal contacts and via groups including one for students at the University of Bedfordshire, one entitled, "Information Security" and one, "Promote Web Accessibility and Web Standards". These groups were selected in particular, because both the Information Security and Web Accessibility groups are open to an international audience and the University of Bedfordshire has a culturally diverse student body with nearly one-third of students coming from outside of the UK. In addition it was felt members of these groups might be interested in our work.

Visitors to our site were provided with a description of the experiment and proceeded where they opted to participate. During the experiment they were asked to carry out the following steps:

1. Provide an email address (as unique identifier and means for communication).
2. Provide a text-based password that had not been used previously and that they felt would be secure.
3. Re-enter the password (a practice traditionally followed to ensure the password has not been mistyped and to strengthen the memory trace).
4. Authenticate using the new password

Once the user had authenticated (or gave up their attempt), a page containing the Musipass interface was loaded. We then asked the participant to:

1. Select a password song clip from a choice of nine over each of four screens. Participants using a mouse listened to clips by floating the pointer over icons and made selections by clicking their mouse. Those using a keyboard listened to clips by tabbing into the icons and selected by pressing the "Enter" key.
2. Enter a short description for each of the chosen clips. (to strengthen the memory trace – akin to asking users to re-type their password).
3. Authenticate with the new musical password.
4. (Optional) Fill out a questionnaire to express their opinion.

A week later, participants were sent an email inviting them to return and attempt a second authentication with their text-based and Musipass passwords. Any user returning before the end of the seven day period was prohibited from accessing the test. On completion, returning participants were given the opportunity to complete a final post-evaluation questionnaire.

5.1 Playback difficulties

Some people who had enrolled to participate reported that they were unable to hear the sound clips. This was the result of two separate issues:

1. On speaking with some participants, we found that level of technical expertise could be an obstacle. Some did not know that their computer could play sound and did not have the speakers switched on, or had audio muted and did not know how to switch it on. It would therefore be beneficial in future versions, to include an instructional sound configuration page.

2. Other users could hear audio on their computer, but not from Musipass. Data gathered during the course of the experiment included information about the software architecture of access devices. We analyzed this, and found that some participants with identical operating system, web browser and Flash plugin configuration could hear the sounds, whilst others could not; indicating the problem is not merely one of compatibility as we had suspected, but is more likely to be a low level error resulting from the way playback objects are interpreted and organized in the Flashplayer

environment at runtime. This, more abstruse, problem requires further investigation. If it is decided to implement future versions of Musipass in Flash, then porting the code over to Actionscript 3, with it's new Sound API might relieve the issue.

The results outlined in Section 6 exclude data gathered from those participants experiencing playback problems, since it would not have been possible to separate the efficacy of musical passwords from the technical issues.

6. RESULTS

The experiment ran for 52 days, during which time 133 people carried out the initial enrollment and authentication process (we will refer to this as *phase one*), with 94 returning for a second authentication attempt seven or more days later (referred to as, *phase two*).

6.1 Participant demographics

6.1.1 Age

A central factor to be investigated in a study of this nature is memorability. However results could well be affected by the age of participants. For this reason we asked them to provide details about their age group. During both phases of the experiment, the biggest majority were aged between 26 and 35 years (39.85% overall during phase one and 39.36% in phase two). The participation of older users was small, with only 2.26% (three users) in the 56-65 age group, and 1.5% (two users) in the over 65 category during phase one. During phase two, two users returned to participate from each of the 56-65 and over 65 groups (Figures 5 and 6).

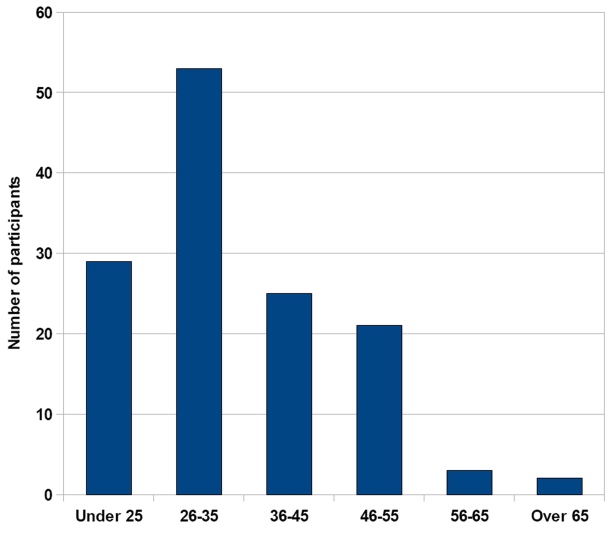

Figure 5: Phase one participant age groups

6.1.2 Geographical distribution

Although we did not ask participants about their cultural backgrounds in the questionnaire, we were able to look up the locations of machines used to source page requests based upon IP addresses. We found that most were from the United States (60%) and Great Britain (19%). This came as no surprise given our advertisement strategy. A further 14% originated from Europe, including Germany, France, Ireland, Austria, Sweden, Norway, Italy and Spain.

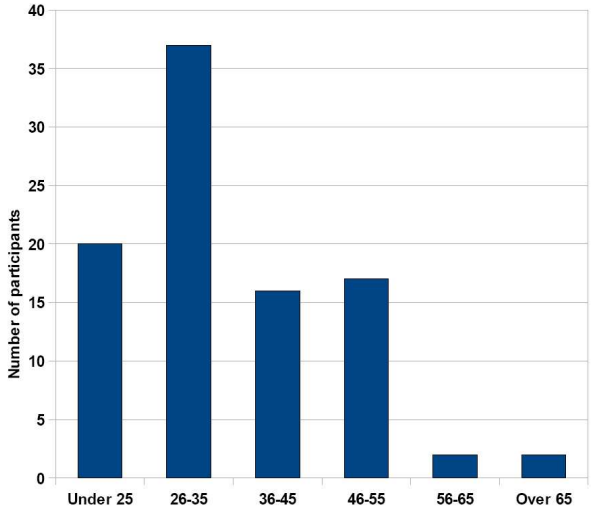

Figure 6: Phase two participant age groups

Just over 1% of requests were from Australia. Countries making up less than 1% each of page requests were Canada, Jordan, the Russian Federation, Mexico and Japan. Location of a further 1% of requests could not be resolved.

It should be stressed that geographical distribution of page requests is only *indicative* of cultural demographic, since not every page request would necessarily lead to experiment participation. If we assume that a roughly equal proportion of page requests lead to registration for all countries, then the figures suggest that our results are relevant mainly to users from the West. Particularly as those from outside of this region may not have been exposed to the clips we deemed as belonging to the "old familiar" category. One strength of using music to form passwords, is that users can be supplied an individual alphabet based upon their own cultural needs and experiences.

6.1.3 Impairments that could affect results

We asked participants about their hearing ability. Almost all phase one respondents said that they have full hearing, with the only exceptions being one person with mild hearing loss, and one whose hearing is corrected to normal with an aid. One of the two returned for the second phase of the experiment.

Quality of vision could also affect the way in which people interact with the system. Four of the phase one participants (3.01%) said that they had suffered a 20% loss of vision, all others had normal or corrected to normal vision. All four returned to participate in phase two. Four participants signified that they had a disability, one had a color vision deficiency, one was dyslexic, and one had attention deficit hyperactivity disorder. One participant opted not to disclose the nature of the disability. Three of the four participants returned to take part in phase two.

6.1.4 Musical experience

We considered that the musical experience of our participants might play a role in their performance when remembering musical passwords. We asked participants to categorize their experience as follows:

Musical Experience	No. Participants	Overall %
None	18	13.53
Listen frequently	49	36.84
Play instrument	44	33.08
Professional Musician	22	16.54

Table 1: Participants grouped by musical experience (phase one)

Musical Experience	No. Participants	Overall %
None	11	11.7
Listen frequently	33	35.11
Play instrument	33	35.11
Professional Musician	17	18.09

Table 2: Participants grouped by musical experience (phase two)

- None
- Listen frequently
- Play instrument
- Professional musician

The number of participants in each category for both phases of the experiment are shown in Tables 1 and 2. These figures include nineteen participants who specified that their experience was "Other", but whose level of expertise could be mapped to one of the four categories. For example, one participant who described their experience as *"married to a pianist and composer"* was re-classified as, "Listen frequently". Another participant said: *"degree in e-music; musical performer since age 5 - some pro; improvisor"* and was re-categorized as "Professional musician".

6.1.5 *Download speed*

On average our participants had a fast download speed, with connections typically ranging between 300 and 10,000 Kbps. The lowest recorded download speed was 198 Kbps, this user was able to authenticate successfully in both phases of the experiment. Due to the average speeds, our results might not be relevant for those with very slow connections.

6.2 Memorability of passwords at phase one

The results for each Musipass authentication attempt were coded as follows:

Successful where the participant successfully logged in.

Failed where the participant at the end of the experiment had failed to recognize and hence recover from instances of decoy sound set presentation after incorrectly guessing a song clip, or who had reached the final log in screen and had selected a non-password element, and

Quit where the participant realized that decoy songs were being offered, but when given the choice of re-attempting the authentication or proceeding to the questionnaire, opted to quit.

Table 3 shows that text passwords were more memorable for our participants than musical passwords right after they had been set up. The values contained in the *Failed* and *Quit* columns were combined to give an overall unsuccessful number of authentication attempts, these along with number of successful attempts for both

	Successful	Failed	Quit	% Success
Traditional	133	0	N/A	100
Musipass	131	0	2	98.4

Fisher's $p = 0.4982$.

Table 3: Phase one authentication results

	Successful	Failed	Quit	% Success
Traditional	58	36	N/A	62
Musipass	86	5	3	91

Fisher's $p = 0.000002$.

Table 4: Phase two authentication results

traditional and Musipass passwords underwent Fisher's exact test, the result of which however, showed the difference to be statistically insignificant (Fisher's $p > 0.05$)

6.3 Memorability of passwords at phase two

During phase two, we found that participants found it easier to remember their Musical passwords than the traditional passwords with a 91% authentication success rate in Musipass compared to a 62% success rate for traditional passwords (Table 4). Many participants returned after a period of disuse that was much longer than seven days (the mode was seven, but the mean was nine), with one user successfully authenticating with Musipass after 36 days away from the system. Full details of success rates grouped by number of days passing between the two phases are given in Figure 7. A high success rate (88-100%) was achieved, at least up until the eleventh day, after which time the data set becomes sparse and we begin to observe less of a marked difference in success rates between the two authentication modes.

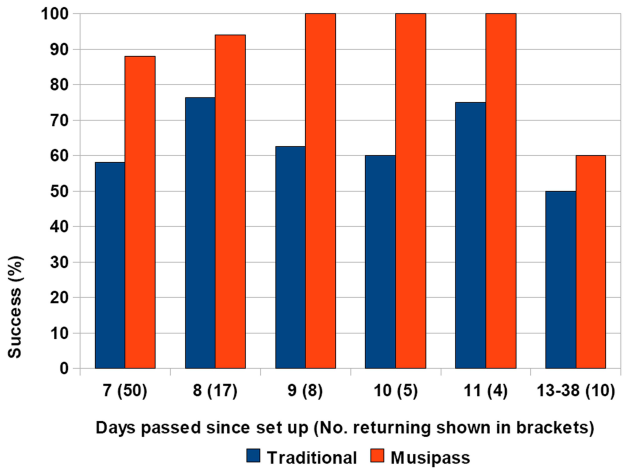

Figure 7: Phase two authentication results grouped by number of days passed since initial set up

6.3.1 *Testing the relationship between age and memorability*

The phase two results were broken down by age group. There were too few participants in the older age groups to analyze performance rates, so we focused our attention on participants up to the age of 55.

We found that those under 25 years of age had the highest success rate with Musipass (100%), followed by the 25-35 age group (91.89%). These in turn were followed by the 36 to 45 group with an 81.25% success rate. For every successive increase in age group, in these first three, there seemed to be an approximate drop of 10 percentage points in Musipass log in success rate. However, those in the 46-55 group outperformed the 36-45 group, contradicting this hypothesis.

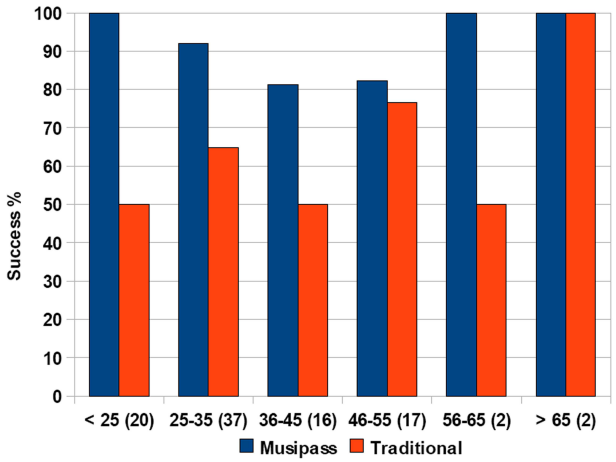

Figure 8: Phase two authentication results by age group

This leveling of success rate been the two groups could indicate one of three possibilities: First, that the memory undergoes a gradual deterioration up to the age of 36-45, after which time it plateaus (at least up until around 55 years). Secondly, that the participants in the 36-45 group were closer to the age 45 than the participants in the 46-55 group were to 55, biasing the result. Or, thirdly that the varied length of time in returning for phase two or another factor, such as musical ability, affected the results.

In order to investigate the first two scenarios we would need the exact age from all participants (something that we did not have). However we could test the third possibility by removing the variable of time passed, breaking the results down further by age and musical experience and then examining the data further for correlations. If there is a relationship between age and memorability, we would expect to observe a relative increase or decrease in success rates between people in the different age groups, but with the same level of musical expertise.

We therefore isolated the results from participants who had returned to complete phase two on the seventh day only (removing the days passed variable), and then reordered the data by age and experience (Table 5, and Figures 9 and 10). We did not find a correlation between age group and ability to authenticate with either password type up to the 46-55 age group when the data was analyzed in this way, leading us to conclude that in our particular sample the age variable had no affect on authentication success.

6.3.2 Testing the relationship between musical experience and memorability

We broke down the authentication results by the four categories of musical experience (Figure 11). Initially there did not seem to be a correlation between memorability and musical experience. Even though the professional musician's overall authentication success rate with Musipass was slightly higher than that of the non-musical participants, this was also the case with traditional passwords. This suggests that perhaps professional musicians had better memory capabilities overall (not just for melodies) or that another non-music related memory factor might have affected results, such as number of days passing between phases one and two.

Musical experience	Age	No. Participants	Musipass success %	Traditional success %
None	Under 25	1	100	0
	25-35	0	0	0
	36-45	1	0	100
	46-55	2	100	100
Listen	Under 25	3	100	66.67
frequently	25-35	8	87.5	37.5
	36-45	3	100	0
	46-55	3	100	66.67
Play	Under 25	4	100	50
instrument	25-35	6	100	66.67
	36-45	4	75	25
	46-55	1	100	100
Professional	Under 25	4	100	75
Musician	25-35	3	100	100
	36-45	3	100	50
	46-55	2	100	100

Table 5: Phase one authentication results

In order to illuminate matters, we decided again to isolate results only from those participants returning on the seventh day, this time grouping the data by musical experience (Figure 12). Here we found that professional musicians and non-musicians tested similarly for traditional password authentication success rates, but that there was a higher success rate in Musipass from professional musicians than there was from the non-musical group. However, only four non-musical participants returned on the seventh day and without more data we still could not be sure about the existence of a relationship between level of musical experience and the ability to log in with Musipass.

Data from participants from the Listen frequently, Play instrument and Professional musician categories was more abundant. We observed a positive correlation for traditional password memorability – the more experienced people were, the better they remembered their password strings. Since ability to recall text strings is not usually associated with musical ability, we can conclude that level of musical experience most likely affects memorial ability in general. We believe this is possibly due to the way musicians are trained to recall complicated patterns whilst performing, strengthening memory as a whole.

6.4 Overall perceptions

In addition to the demographic questions highlighted in Section 6.1.1 we used the questionnaire to gather information about our participants' attitudes towards authentication in general and their perceptions of, and reactions to, Musipass.

6.4.1 Phase one questionnaire

We asked participants whether they usually had difficulties in remembering their passwords and PINS and whether they usually write them down. 39.1% of participants said that they had difficulties remembering, and 33.08% said that they write them down.

We asked users to tell us how long it took for them to recognize

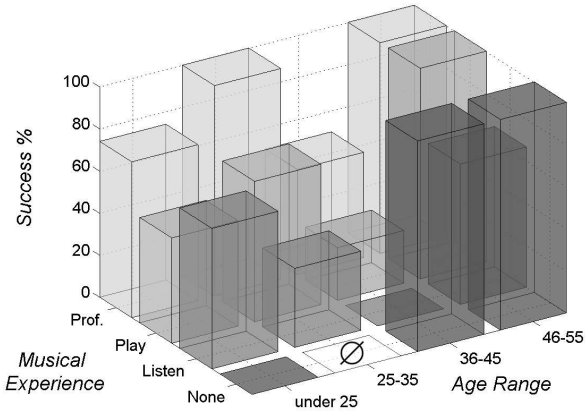

Figure 9: Traditional password success rate grouped by musical experience and age

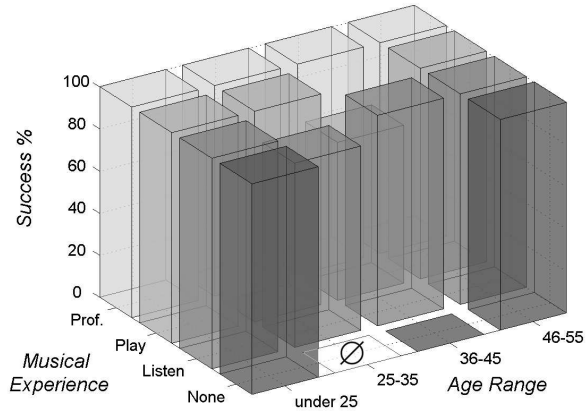

Figure 10: Musipass success rate grouped by musical experience and age

their sound clips, by selecting from one of four options: "Almost immediately", "After 2-3 seconds", "Only after a full clip" or, "I needed to listen more than once". Most participants (74.44%), recognized their clips almost immediately, whilst a further 19.55% recognized their clips after 2-3 seconds. The remaining minority (6.01%) said that they had to listen to the full clip, or that they needed to listen more than once.

When asked to rate how much they liked Musipass on a five point Likert scale, with 1 being *disliked very much* and 5 *liked very much*, the mode average response given was 4, showing that most users liked the system, but not to any extreme.

When asked to rate how easy it was to remember their sound clips on a scale from 1 (*very difficult*) to 5 (*very easy*). The mode average response given was 5: *very easy*.

We asked users how satisfied they were after Musipass set up and training was complete, with the amount of time it took to carry out the final log in on a scale from 1 (*Very dissatisfied*) to 5 (*Very satisfied*).

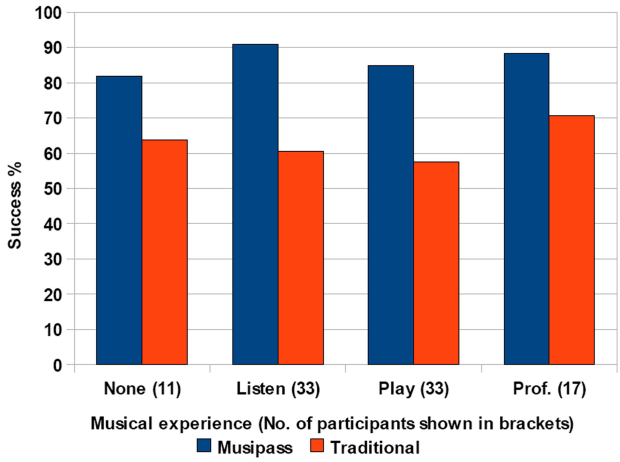

Figure 11: Phase two authentication results grouped by musical experience

Overall our participants were not satisfied with the amount of time it took to authenticate, with the mode response given being 2: (*dissatisfied*).

When asked how easy it was to go through the process of choosing their password sounds on a scale from 1 (*Very difficult*) to 5 (*Very easy*) the most common response was 4, suggesting that most users found this easy enough to do.

Participants rated how time consuming it was to choose their password sounds on a scale from 1 (*Not time consuming*) to 5 (*Very time consuming*). The mode average response given was 4, showing us that users felt it took too long to enrol.

We asked participants how much mental effort was involved in choosing their sounds on a scale from 1 (*Very little effort*) to 5 (*A great deal of effort*). The average response was 2.

When asked if they thought someone who knew them well would be able to guess the songs they chose, on a scale ranging from 1(*Yes*) to 5 (*No*). The mode average response given was 2, indicating that most people thought their musical passwords might be guessable by friends and family.

6.4.2 Phase two questionnaire

We were pleasantly surprised at how positive the overall response from phase two participants was. When asked how much mental effort was involved in logging in with Musipass on a scale of 1 (*very little mental effort*) to 5, (*too much mental effort*). The mode average response was 1: *very little mental effort*.

Likewise, most participants found recognizing their song clips very easy, rating this aspect on average 1, on a scale of, 1 (*very easy*) to 5 (*very difficult*).

We now include some example comments received about memorability below that are typical (all emphasized text that follows are quotations from the responses, for authenticity spelling and grammatical errors have been retained):

> *"I thought it worked very well and found it very easy, without cheating! Though I did recognize some of the other songs in each group I knew instinctively that they were not the right ones. When I came across my choice I immediately knew it and moved on without listening to the others. Million times more easy to recall using*

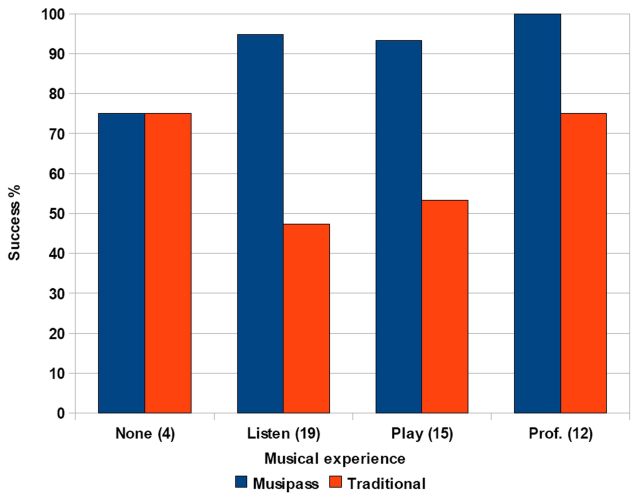

Figure 12: Phase two authentication results for people returning on day 7

the songs, I couldn't remember my text password having a random word. Very impressed with the system Cheers and good luck"

"...i'm also amazed on how easy it was to identify the selected clips (believe it was by emotional connections but also by negative relationships with the other clips; btw i still can't remember my textual password!)"

"After trying unsuccessfully to log in with my alphanumeric password, the usefulness of Musipass became clear. I wouldn't want to have to use it for things I log in to often, such as online banking or email, but for lesser used passwords, this would be a better option than guessing, trying to retrieve it from a note or document, or waiting to have it emailed to me..."

Although most users were able to authenticate successfully with Musipass, a few expressed some valid concerns about memory interference between password clips and the distractors, or should audio passwords be used to access multiple systems, commenting:

"well, I couldn't remember my text password either, so it's a wash! :) it was particularly confusing when there were two or more songs on the same screen that i'm fond of or familiar with; it was hard to find a reason to choose one over the other during password selection, and hard to remember which I'd picked when I returned."

"If several other sites were using Musipass and, potentially, using similar songs I'm not sure it will work. I recognized "the songs I liked" and I would need to be persuaded that I could successfully disambiguate different selections of songs. I hope you'll pursue it."

We asked users how long it took them in general to recognize their sound clips, asking them to select from the same options as given in phase one, these were: "Almost immediately", "After 2-3 seconds", "Only after I had listened to the full clip" or, "I had to listen to the options more than once". The figures had not changed very much from the phase one responses to this question, with most (50%) of participants stating that they were able to recognize their sounds almost immediately, and a further 40.43% recognizing them after 2-3 seconds. The remaining 9.57% listened to the clips in their entirety or needed to listen multiple times. One participant made an interesting point on this aspect, again it related to the issue of interference:

"I was pleasantly surprised by the immediacy of my recognition of the music clips. Perhaps this is because my mind is less clogged with musical passwords than with text-based passwords?"

When asked how frustrated they felt during the authentication process on a scale of 1 (*Not frustrated at all*) to 5, (*Very frustrated*), the average response given was 1 (*Not frustrated at all*).

When asked how they felt about the time involved in logging in to Musipass on a scale from 1 (*It was very quick*) to 5 (*it took too much time*), the average response we received was 4 – Even though participants said they enjoy using the system, they did not seem to think it was practical in terms of the time required to authenticate. Some examples of typical feedback we received are as follows:

"I remebered the music login easily, though had forgotten the text one. Probably if i had to log in more often than the one week gap I prob would have remembered it. Also if I was trying to log in for something specific that I really needed to know then the time required for loggin in may get annoying. However in the scenario given, ie leasurly login with no urgency, it was pretty much the most fun login I've ever done :D".

"I think it was really fun, but I am not sure how I would like it if I had to play so many clips just to login to my email, or something like that, which I want to do quickly."

Finally, we asked participants how much they liked Musipass overall on a scale of 1 (*Disliked very much*) to 5, (*Liked very much*) The average response given was 4, most users liked the system.

A few participants expressed security concerns about the possibility of observation-based attacks and password strength of the system (at no point did we tell them that their password alphabet would differ from someone else's, extending the range of possibilities, or that the placement of songs was shuffled each time they were loaded). However we do feel it is important to include these viewpoints, as perception of security offered is an important factor when considering likelihood of technology acceptance. Some typical comments we received are:

"even though it sometimes takes me several attempts to remember which of 5 passwords I have used for a site, this still felt longer, and i was also conscious that if it wen't wrong I was potentially going to have to do it again which would definately be longer. I also am always using my computer in public space and would not want to long into anything confidential in an audible way that others could overhear"

"it's an interesting concept but I'm not sure how it would work in practice while making the paranoid nerd-types like myself feel secure in our password selections. (nothing like random strings to make you sleep better at night!)"

"Doesn't seem very secure, and takes a while to log in. Reusing an easy-to-remember text password would be faster."

7. DISCUSSION

In Section 2 we argued that the efficacy of an alternative authentication mechanism should be judged based on two criteria: memorability and security. In terms of memorability Musipass appears to pass muster. Users authenticated successfully after a full week away from the system. Furthermore, they appeared to like the system, enjoying the experience of choosing the sound clips and returning to attempt to remember them a week later. They clearly found Musipass easy to use and were not at all frustrated either during enrollment or authentication.

In terms of security, it was clear that our participants had some concerns about the guessability of their choices. Certainly alphanumeric passwords are often too easy to guess if one knows someone well enough, and we had hoped that Musipass would offer superior strength in terms of guessability as well. Guessability may be worsened if we allowed users to upload their own choices into the system. We could also run into problems with digital copyright. The question remains as to exactly how guessable musical passwords are. If our user's perceptions are correct, and passwords in Musipass are guessable, do we accept this as an inherent weakness of the system, or is there a way to make the choices less so? If the former, this does not make Musipass superfluous in the world of authentication. Many systems which ask users to authenticate themselves require this more for their own convenience than to achieve any measure of protection. For these systems, where authentication is required merely to deliver a measure of customization or to attribute contributions, we could feasibly make use of Musipass because of its superior memorability.

Techniques for hardening the system against guessing attacks might include, issuing the user with sound clips rather than allowing freedom of choice. This option is likely to impact negatively on the memorability of the sound clips and therefore may prove untenable. Another option would be to populate the system with many different musical genres, and to vary the genres offered to different users. It is then less likely that users' choices will be predictable although a price might be paid, once again, in terms of memorability.

Finally, the one measure we alluded to, but did not discuss in great detail, was the *convenience* of the mechanism. Even though people complain about passwords, the undeniable fact is that they are very convenient [39]. Although participants indicated that they liked the Musipass system, some of their comments show that some exasperation was experienced due to the time-consuming nature of the authentication process.

We know that computer users are definitely concerned about their convenience, which is not unreasonable of them. If they anticipate that they will have to authenticate using a time-consuming mechanism such as Musipass a number of times a day, one can readily anticipate their dismay. Certainly Liddell *et al.* [33] experienced this reluctance when he hoped that users would listen to music and then choose an associated picture, but they simply chose the picture, considering the listening phase to be too time-consuming.

Liddell's study used students as participants and it must be noted that most students are young and have few memory difficulties. On the contrary, older users are plagued by memory problems and will probably be more willing to accept some inconvenience in return for increased memorability. Indeed, Renaud [44] found that older users were not concerned about the time-consuming nature of the image-based authentication mechanisms tested. Their memory difficulties made the memorial nature of the mechanism far more important than how long it took them to authenticate.

Perhaps the future of Musipass will lie within the context of low-risk systems, which are used infrequently, by users who are more concerned about forgetting than convenience.

This experiment has been a good starting point, and opens the way for much future research, to attempt to address the security concerns and convenience limitations of Musipass, while retaining the superior memorability thereof.

8. CONCLUSION AND FUTURE WORK

In this paper we report on early trials of a scheme we called "Musipass", which used sound clips as the password alphabet. Audio password systems can be designed so that they rely on recognition to authenticate (such as in our design) and therefore are able to offer some of the memorial utility that image-based passwords can offer, whilst at the same time being less prone to shoulder surfing attacks as they can be used with earphones in any public place.

We tested the Musipass system and traditional passwords for overall memorability after a period of disuse. We found that, overall, Musipass offered better performance, with 48% more successful authentication than with traditional passwords.

Participants returned for the second phase of our experiment from seven to thirty-eight days from the date of initial set up. However, we found that when we isolated data from participants returning on the seventh day for phase two, there was no identifiable correlation between the age variable and memorability. We found that there was a correlation between musical experience and ability to authenticate using Musipass. However, this effect was also observed in traditional password system authentication, suggesting that the level of musical experience not only affects the ability of music clip recognition, but the memory as a whole.

Musipass consistently supported a large proportion of users toward their goal of successful authentication throughout all subcategories that were analyzed (including number of days taken to return for the second authentication, age and musical experience). Therefore, regardless of these factors, users were more able to authenticate successfully with musical passwords than when using a text-based scheme. The overall reaction to the system was positive. The majority of participants liked the system and found it easy to use. However (as with all recognition-based authentication systems) one drawback of the design was that participants were less satisfied with was the amount of time required of them.

We hope that the work we have outlined in this paper will encourage further research into the efficacy and feasibility of sound-based passwords to compliment the existing research into image-based schemes. In addition to the questions we have raised, other areas for future work include validation of Smith's [50] and Blood and Zatorre's [8] findings in the context of Musipass – Is it possible that a previously unfamiliar tune might become familiar with use and would a user persevere with it long enough to reach this point? If so, the possibilities for alphabet inclusion are increased. We might, for example, use creative commons licensed music (ours was a research prototype and hence in US and UK copyright law would be treated as "fair use" [53, 55]. In a non-research system, even if that system is non-profit making, this may not be the case. In this situation, the system can no longer be considered "cost less" as the clips will need to be paid for. If we are to remain faithful to our originally intended purpose, ensuring inclusivity for groups who find using traditional approaches difficult, it would seem unfair to ask them to pay, or to view advertisements, only to be given the same opportunity to authenticate that others take for granted. A solution, might be to invite record companies to submit clips, pay-

ment to them would take the form of showcasing work of emerging artists. Since we all have a preference for the familiar, this could in turn, positively affect sales.

Another question relates to scalability over multiple sites. How many musical passwords can people remember? Is there a way to ensure music clips on one web site differ from those on another? If not people might become confused as they select clips that are familiar, but incorrect in the wider context. In this case, implementing Musipass as single sign on might provide a solution.

The majority of participants told us that they didn't have to listen to all of the music clips presented to them, selecting "their" clip as soon as they heard it and moving on. This, along with some of the comments received, serves as anecdotal evidence that people do not find the distractor clips becoming as familiar to them as their selected password clips. Our study involved a lengthy delay between enrolment and authentication. trials involving regular use, perhaps on a daily basis, might be better placed to confirm this as fact.

Reporting task times from our web based study would have been misleading, since it cannot be guaranteed that our participants did not stop during the experiment to do something else (at least one participant told us that they did this). Although many users found the time to authenticate to be too time consuming, the question still remains as to exactly how long it took. Repeating the experiment in a lab-based environment could provide more reliable data on which to carry out an analysis.

Finally, the design outlined in this paper should be considered as an example of a possible implementation for an audible password. The choices we have made for included music clips and the procedure for password selection were based upon research into enhancing memorability and security. It should also be possible to populate the system and to create musical passwords in other ways so as to enhance these properties. We hope that the work outlined in this paper will provide a starting point for development in this field.

Acknowledgments

Thanks are due to David L. Goodwin for invaluable discussion and to Rachna Dhamija and Comac Herley our pre and post workshop shepherds.

9. REFERENCES

[1] All about Symbian. BBC: Developing countries dominate mobile phone sales. Web Document. http://www.allaboutsymbian.com/news/item/4940_BBC_Developing_countries_domin.php

[2] M I Alpert, J I Alpert, and E N. Maltz. Purchase occasion influence on the role of music in advertising. *Journal of Business Research*, 58(3):369–376, March 2005.

[3] B F Barton and M S Barton. User-friendly password methods for computer-mediated information systems. *Computers & Security*, 3(3):186–195, 1984.

[4] S D Bella, I Peretz, and N Aronoff. Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, 65(7):1019–1028, 2003.

[5] A Bellaire. Getting creative money's worth in TV. *Advertising Age*, s-4, Jul 1979.

[6] M Besson, F Faïta, A-M Bonnel, and J Requin. Singing in the brain: Independence of music and tunes. *Psychological Science*, 9(6):494–498, 2002.

[7] E Bigand and B Poulin-Charronnat. Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100:100–130, 2006.

[8] A J Blood and R J Zatorre. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Nat Acad Sci USA*, 98:11818–23, 2001.

[9] British Dyslexia foundation. About Dyslexia. Web Document. http://www.bdadyslexia.org.uk/about-dyslexia.html

[10] S Brostoff and A Sasse. Are passfaces more usable than passwords? a field trial investigation. In S. McDonald, editor, *People and Computers XIV - Usability or Else! Proceedings of HCI 2000*, pages 405–424. Springer, 2000.

[11] B Bryson. "Anything but Heavy Metal": Symbolic Exclusion and Musical Dislikes. *American Sociological Review*, 61(5):884–899, 1996.

[12] G Burns. A typology of 'hooks' in popular records. *Popular Music*, 6(1):1–20, 1987.

[13] S Chiasson, A Forget, and R Biddle. Accessibility and graphical passwords. In *SOAPS*, Pittsburgh, July 2008.

[14] M Conrad, T French, and M Gibson. A pragmatic and musically pleasing production system for sonic events. In *10th IEEE International Conference on Information Visualisation IV06 (5-7 July 2006, London)*, pages 630–635. IEEE Publications, 0-7695-2602-0, 2006.

[15] R G Crowder and M L Serafine. Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition*, 18(5):469–476, 1986.

[16] A De Angeli, L Coventry, G Johnson, and K Renaud. Is a picture really worth a thousand words? reflecting on the usability of graphical authentication systems. *International Journal of Human-Computer Studies: special issue: HCI research on Privacy and Security*, 63(1-2):128–152, July 2005.

[17] R Dhamija and A Perrig. Déjà vu: A user study using images for authentication. In *Proceedings of USENIX Security Symposium*, pages 45–58, Denver, Colorado, August 2000.

[18] Dyspraxia foundation. About Dyspraxia Web Document. http://www.dyspraxiafoundation.org.uk/services/dys_dyspraxia.php

[19] A K Engel and W Singer. Temporal binding and the neural correlates of sensory awareness. *Trend Cognitive Science*, 5:16–25, 2001.

[20] S Eschrich, T F Munte, and E O Altenmüller. Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC Neuroscience*, 9:48, 2008.

[21] K Franklin and J Roberts. A path based model for sonification. *Information Visualization*, 1(1):865–870, July 2004.

[22] S Furnell, I Papadopoulos, and P Dowland. A long-term trial of alternative user authentication technologies. *Information Management & Computer Security*, 12(2):178–190, 2004.

[23] S Gaw and E W Felten. Password management strategies for online accounts. In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security*, pages 44–55, New York, NY, USA, 2006. ACM Press.

[24] S Hébert and I Peretz. Are text and tune of familiar songs separable by brain damage? *Brain and Cognition*, 46(1-2):169–75, 2001.

[25] J D Herrington. Effects of music in service environments: a field study. *Journal of Services Marketing*, 10(2):26–41, 1996.

[26] L E Hirsch. Weaponizing classical music: Crime prevention and symbolic power in the age of repetition. *Journal of Popular Music Studies*, 19(4):342–358, 2007.

[27] M Jackson. Music to deter yobs. BBC news Magazine, 10 Jan 2005

[28] P Janata, S T Tomic, and S K Rakowski. Characterisation of music-evoked autobiographical memories. *Memory*, 15:845–860, 2007.

[29] L Jäncke. Music, memory and emotion. *Journal of Biology*, 7(21), 2008.

[30] D Katre. Using mnemonic techniques as part of pictorial interface for self-identification of illiterate villagers. In *Proc I-HCI 2004*, Bangalore, India, 6-7 Dec 2004. http://www.cdac.in/html/pdf/dkatre.pdf. Accessed: October 2008.

[31] P Lazarsfeld and H Field. *The People Look at Radio*. U.N.C. Press, 1946.

[32] J E LeDoux. Emotion as memory: Anatomical systems underlying indelible neural traces. In S. Christianson, editor, *Handbook of emotion and memory: Theory and research*, pages 269–88. Erlbaum, 1992.

[33] J Liddell, K V Renaud, and A De Angeli. Authenticating users using a combination of sound and images. In *HCI 2003*, Bath, UK, September 2003. Short Paper.

[34] M Mélen and I Deliége. Extraction of cues or underlying harmonic structure: Which guides recognition of familiar melodies? *European Journal of Cognitive Psychology*, 7(1):81–106, March 1995.

[35] T Mendori, M Kubouchi, M Okada, and A Shimizu. Password input interface for primary school children. In *Proceedings of the International Conference on Computers in Education (ICCE02)*, pages 765–766, Auckland, New Zealand, December 3-6 2002.

[36] R E Milliman. The influence of background music on the behavior of restaurant patrons. *Journal of Consumer Research*, 13:286–289, Sept 1986.

[37] Miniwatts Marketing Group. World Internet Usage Statistics. Web Document. http://www.internetworldstats.com/stats.htm

[38] B Monaco and J Riordan. *The Platinum Rainbow*. Sherman Oaks, 1980.

[39] R Morris and K Thomson. Password security: A case history. *Communications of the ACM*, 22(11):594–597, 1979.

[40] I Peretz, A J Blood, V Penhune, and R Zatorre. Cortical deafness to dissonance. *Brain*, 124:928–940, 2001.

[41] I Peretz, M Radeau, and M Arguin. Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition*, 32(1):142–152, 2004.

[42] J Podd, J Bunnell, and R Henderson. Cost-effective computer security: Cognitive and associative passwords. In *6th Australian Conference on Computer-Human Interaction (OZCHI '96)*, Hamilton, NEW ZEALAND, November 24 - 27 1996.

[43] B Reimer. Why do humans value music? Web Document. http://course1.winona.edu/cschmidt/MS298/Why Do Humans Value Music.pdf

[44] K Renaud. A visuo-biometric authenticaton mechanism for older users. In *Proc British HCI 2005. Sept 5-9, Edinburgh*, pages 167–182, 2005.

[45] P J Rentfrow and S D Gosling. The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6):1236–1256, 2003.

[46] K R Scherer and M R Zentner. Emotional effects of music: Production rules. In P N Juslin and J A Sloboda, editors, *Music and emotion: theory and research*, chapter 16, pages 361–392. Oxford University Press, 2001.

[47] A Schmidt, T Kölbl, S Wagner, and W Strassmeier. Enabling access to computers for people with poor reading skills. In C Stary and C Stephanidis, editors, *8th ERCIM Workshop on User Interfaces for All. Lecture Notes in Computer Science (LNCS), Vol. 3196*, pages 96–115, Vienna, Austria, 2004.

[48] M A Sewall and D Sarel. Characteristics of radio commercials and their recall effectiveness. *Journal of Marketing*, pages 52–60, Jan 1986.

[49] A Small, Y Stern, M Tang, and R Mayeux. Selective decline in memory function among healthy elderly. *Neurology*, 52:1392–6, 1999.

[50] A B Smith. The pleasures of recognition. *Music and Letters*, XIII(1):80–84, 1932.

[51] S L Smith. Authenticating users by word association. In G Papp and R Posch, editors, *Proceedings of the Human Factors Society 31st Annual Meeting*, pages 135–138, Wien, 1987.

[52] D Stewart and D Furse. *Effective Television Advertising*. Lexington, 1986.

[53] The UK Copyright Service. Factsheet P-01: UK Copyright Law. Web Document. http://www.copyrightservice.co.uk/copyright/p01_uk_copyright_law

[54] Twist, J. BBC News – Technologies to aid the poor. Web Document. http://news.bbc.co.uk/1/hi/technology/4679015.stm

[55] US Copyright Office. Fair Use. Web Document. http://www.copyright.gov/fls/fl102.html

[56] R R Weeks and W V Marks. Music's power for television advertising. *Southern Journal of Business*, 3(4):35–39, 1968.

[57] D Weinshall and S Kirkpatrick. Passwords you'll never forget, but can't recall. In *Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems*, volume 2 of *Late breaking result papers*, pages 1399–1402, 2004.

[58] W Wells, J Burnett, and S Moriarty. *Advertising: Principles and practice*. Prentice Hall, 1989.

[59] S Willems, M Van der Linden, and C Bastin. The contribution of processing fluenxy to preference: A comparison with familiarity-based recognition. *The European Journal of Cognitive Psychology*, 19(1):119–140, 2007.