

THE SOUND OF MUSICONS: INVESTIGATING THE DESIGN OF MUSICALLY DERIVED AUDIO CUES

Ross McLachlan¹, Marilyn McGee-Lennon² and Stephen Brewster²

Glasgow Interactive Systems Group,
School of Computing Science,
University of Glasgow,
Glasgow, G12 8QQ

r.mclachlan.1@research.gla.ac.uk¹, {first.last}@glasgow.ac.uk²

ABSTRACT

Musicons (brief samples of well-known music used in auditory interface design) have been shown to be memorable and easy to learn. However, little is known about what actually makes a good Musicon and how they can be created. This paper reports on an empirical user study (N=15) exploring the recognition rate and preference ratings for a set of Musicons that were created by allowing users to self-select 5 second sections from (a) a selection of their own music and (b) a set of control tracks. It was observed that sampling a 0.5 second Musicon from a 5-second musical section resulted in easily identifiable and well liked Musicons. Qualitative analysis highlighted some of the underlying properties of the musical sections that resulted in 'good' Musicons. A preliminary set of guidelines is presented that provides a greater understanding of how to create effective and identifiable Musicons for future auditory interfaces.

1. INTRODUCTION

Musicons [1] are musically-derived auditory stimuli. They are short snippets of music which can be linked meaningfully to an interface element or message (in a similar way to an Earcon for example). Musicons have so far been found to be recognisable, memorable over time and easy to learn [1]. Little is yet known, however, about what makes a 'good' Musicon. The choice of music from which to create Musicons is a key research question. Previous studies [1], [2] have shown that familiar pieces of music (such as current chart hits and musical 'memes') can be recognised from very brief samples. There has been no work investigating how recognition or preferences are affected when users themselves can select the music that the Musicons are based on.

Given that we only need a brief snippet from an entire music track to create a useful Musicon [1], [3], we need to investigate potential guidelines to aid designers in choosing the right section of the music to sample to create a useful Musicon. If the wrong section of music is selected the user may not recognise the track at all. The existing literature on Audio Thumbnailing could provide useful insights into the process of automatically extracting a representative portion of a song (such as in [4-6]). However, since different parts of a song will have different meanings to different users and since there could be more than one Musicon created from a single track, there is currently no clear way to find a Musicon algorithmically. If we could identify guidelines for Musicon creation that were able to take user's

subjective preferences into account we could begin to automate the creation process based on a user's music collection.

It might also be desirable to exploit any existing relationships and emotive memories users may have with their own music tracks to enable the creation of more personalized Musicons. A Musicon personalized to a user might be more confidential to that user, easier to learn and/or remember. Understanding more about how best to create these more personalized Musicons and how well they are recognized and/or rated subjectively will provide some much needed groundwork in order that personalized Musicons can be explored in auditory interface design more thoroughly.

This paper presents a user study investigating the effectiveness of Musicons created from a user's own musical tracks. In Phase 1 of the study users were invited to upload their own music tracks and select 5 second sections based on two criteria (1) the section the user felt was most representative of the piece of music and (2) the section of the music he/she personally preferred. In Phase 2, the resulting Musicons were presented to users and evaluated in terms of both recognition and preference. Phase 3 involved analysing the resulting Musicons in terms of the underlying musical properties of the selected sections to understand better what makes a good or bad Musicon. The paper concludes with some initial guidelines for the design of successful Musicons.

2. BACKGROUND

Auditory notifications are used to alert users in a variety of applications such as calendars, social networking tools, instant messengers or SMS and telephony services. Auditory cues can take many different forms ranging from speech, to metaphorical mappings using everyday sounds (such as Auditory Icons [7]) to abstract representations with musical tones (such as Earcons [8], [9]) and super speeded-up speech (Spearcons [2]). The nature of these auditory stimuli makes each more or less appropriate depending on the user, the task and the context.

In selecting auditory cues there is an intrinsic trade-off between ease of comprehension and confidentiality; as stimuli become easier to learn they often become less private and *vice versa*. The following section briefly reviews the auditory design space with respect to ease of comprehension, confidentiality and ease of creation – all crucial factors in the design of effective and usable auditory cues. The final section evaluates Musicons in terms of each of these auditory design factors.

2.1. Comprehension

Speech messages require little or no learning if you understand the language they are spoken in. Meaningful speech messages, however, can be slower to output than other auditory cues [10]. Auditory Icons (described by Gaver [7] as “everyday sounds mapped to computer events by analogy with everyday sound-producing events”) create realistic or metaphorical mappings between signifier and signified using real world sounds to represent virtual objects or actions. Since the sounds share a semantic relationship with the messages they communicate they can be easy to learn and remember. Their success, however, is fundamentally dependent on the success of the metaphor used [11] and since there is not always a mapping between a real world sound and a virtual interface action, it can be difficult to design a set of universally successful Auditory Icons. Earcons are abstract and have to be learned. They are defined by Blattner *et al.* [8] as “non-verbal audio messages used in the user-computer interface” and by Brewster *et al.* [9] as “abstract, synthetic tones that can be used in structured combinations to create audio messages”. Once the association between the signifier and signified is learned, however, Earcons have been demonstrated to be a successful way to deliver auditory messages [9].

2.2. Confidentiality

Privacy or confidentiality can be an important factor in designing notifications since the messages they deliver may contain personal or sensitive information (such as with medical or personal hygiene reminders), or be delivered in a public context where only the recipient wants to intercept the message (such as when using a mobile device in a public place). Notifications that are easier to learn (such as speech or Auditory Icons) do not always offer the same level of confidentiality as more abstract auditory notifications (such as Earcons). Earcons bear no semantic relationship with the content they communicate and so those who do not know the relationship will not automatically understand the messages. Earcons (once learned) can therefore be more confidential than Auditory Icons or speech.

Spearcons are “super speeded up speech” [2] which aim to solve some of the problems associated with speech output. Text to be communicated is sped up to the point where it is not necessarily recognizable as speech yet the message can still be comprehended [2], [12]. This type of cue may provide a level of privacy not afforded with conventional speech output; if a person is not the intended recipient then the message is more difficult to intercept unintentionally. More abstract notifications can potentially offer a greater level of confidentiality since there is no semantic relationship between signifier and signified.

2.3. Creation

The key to using auditory stimuli to convey information successfully is the ability to parameterise the elements of the sound in order to encode information. With speech or Spearcons this is achieved by the concatenation of individual words in order to make sentences or structures that convey the meaning. When using speech or Spearcons, menus can be rearranged or augmented dynamically without disturbing the mapping between sounds and menu items, thus allowing interfaces to evolve without having to extend the audio design.

When using Earcons or Auditory Icons, the mapping from

sound to meaning has to be created either abstractly or through a metaphor. The key difference between Earcons and Auditory Icons is the ease of parameterisation. Elements that make up an Earcon such as timbre, melody and pitch, can be extracted, analysed and manipulated using some musical skill and standard musical tools to create classes of sounds. Brewster *et al.* [13], for example, define a set of guidelines for the creation of Earcons that include recommendations of which parameters to use and how to manipulate them to maximise distinguishability. Earcons allow creation of families of sounds such that notifications and alerts that are related sound similar. Furthermore, if Earcons are designed around a grammar, a user need only learn a set of rules to understand a larger number of notifications [14].

Despite the fact that an Auditory Icon is composed of a collection of sonic elements, it is generally recorded as an atomic unit. This makes auditory icons more difficult to parameterise. There is work on the use of physical models, for example, to allow the simulation and manipulation of real-world sounds but there still remain only a small number of good models and manipulations [15]. This can make the creation of dynamic sets of Auditory Icons difficult.

In summary, there is a clear trade-off between ease of comprehension and confidentiality when using audio stimuli, one which is inherent in the difference between the abstract and metaphorical mapping of signifier to signified. Privacy issues arise with metaphorical mappings since others can potentially overhear the explicit reminders. On the other hand, the recipient may find abstract mappings more difficult to learn. The ease of creation also impacts on the usefulness of audio stimuli, since those that are easier to create make extending the audio design simpler, thus allowing the user interface to be more flexible.

2.4. Musicons

Musicons are defined as “extremely brief samples of well-known music used in auditory interface design” and have been proposed as another solution to address this gap in the audio design space [1]. By sampling a short snippet of a music track, a distinct auditory cue can be created. Musicons can enable designers to exploit existing associations and emotive memories a user may have with a piece of music to create reminders that are abstract in their relationship with the signified as well as being more memorable and potentially easier to learn.

Garzonis *et al.* [11] used pieces of music in some of their auditory icons. The BBC News and the 20th Century Fox themes were used for news and entertainment notifications, respectively. Users were able to use these effectively so this supports the notion that music may be a useful medium through which to convey information. Shellenberg *et al.* [3] asked users to identify pop tracks from short snippets of music and suggested that people could identify pieces of music well from very short snippets. McGee-Lennon *et al.* [1] created Musicons from well-known pieces of music and mapped them to everyday reminders showing that users achieved a high level of recognition (89%) sustained over a 1 week testing period.

In some respects, Musicons are comparable to Spearcons in terms of confidentiality. They can be much shorter than other types of audio stimuli and, if people do not know the association of message to musical track, the notification can provide confidentiality for the target user. Butz and Jung [16] demonstrated the use of a system that communicated notifications to a user in

musical motifs that appeared in ambient background music. Privacy was increased because the motifs used were specific to a user and would simply sound like part of the music to others. Furthermore, the notifications would not disrupt those for whom they were not intended. However, the authors concluded that the method was impractical because of the high overhead involved in composing a piece of music into which the notifications could be inserted seamlessly. The full potential of Musicons for delivery of more personalised and/or confidential messages has yet to be fully explored, though Musicons do not have as high a compositional overhead as the technique described above.

One potential advantage of Musicons over Earcons or Auditory Icons is that they could be simpler to create. A designer only needs to pick a piece of music and take a short, identifiable snippet to create a Musicon. No musical or sound design expertise is needed and there is a large amount of source material to choose from. Users could also easily create their own Musicons and they could be created automatically once Musicons are more fully understood. Schellenberg *et al.* [3] selected snippets to be “maximally representative” of the track based on the experimenter’s judgment. However, except that snippets were selected to start on the downbeat at the beginning of a bar, no other guidelines were given for suitable sections from a musical track that we could use to create Musicons.

Previous work on audio thumbnailing could provide a useful insight into the creation of Musicons. An audio thumbnail is a short, representative sample of a piece of music used as a preview in order to aid search and retrieval of music tracks from a large collection [17]. However, such methods only aim to create one representative thumbnail per track [4], [6] which would be used by all users. Since we are interested in exploiting existing personal relationships and emotive memories users may have with their own music tracks, we need to investigate more subjective assessment of representativeness, a question which we address in this paper.

3. MUSICON EXPERIMENT - OVERVIEW

Previous studies have shown that pieces of music can be recognized from snippets as short as 0.2 seconds in length [1], [3]. Very little is known, however, about what makes a snippet good or bad for use as a Musicon. It is not clear *how* to pick the particular section of the music track from which to create the Musicon in terms of either performance (recognition and memorability) or preference (how pleasant it sounds).

The selection of the right section of the music to use for creating Musicons is potentially highly subjective. There is no universal metric to define ‘representativeness’ in terms of a section of a piece of music. We cannot assume that a universal set of Musicons is possible or ideal, and so it is necessary to test performance and preference for Musicons generated from music selected by users themselves from their own music collections.

In Phase 1 of a three part study we asked users to bring 5 music tracks from their own private collection for use in generating personalised Musicons. In Phase 2, recognition performance and preference for the Musicons were investigated. In Phase 3 we explored the underlying properties of good and bad Musicons. The following section will present each phase of the study in turn and then discuss how our findings might be used to offer initial guidelines for the design of good Musicons.

4. PHASE 1 – MUSICON CREATION

To investigate the most salient and useful features of musical tracks from which to create Musicons, an example set of Musicons was required. Results from [1] and [3] suggested that people can identify well-known tracks from very short snippets chosen by experts but there have been no studies investigating how well users can recognise snippets from tracks they have chosen themselves. To investigate this, participants were asked to supply tracks from their own music library from which a number of Musicons could be generated.

The same fifteen participants took part in both Phase 1 and 2. There were 6 females and 9 males, aged 19 - 53, none of whom reported any hearing problems. Nine of the participants reported having had formal musical training (two had a degree in music and 7 had some private tuition or training during secondary school). The remainder had no musical training.

Participants were asked to supply 5 tracks from their own music library - *Participant Tracks*. In addition, 5 *Control Tracks* were used to create Musicons that were the same across all participants. The Control Tracks, which included those used in [12], were:

- The Rembrants: I’ll be there for you (Friends TV show theme)
- Ray Parker Jr: Ghostbusters
- Johan Pachelbel: Canon
- John Williams: Theme from Jurassic Park
- Theme from James Bond

These tracks were chosen because they had strong thematic associations with popular culture for the sample group of westernized adults living in the UK and the first four had proved to be effective in a previous study of Musicons [1].

By including both control and participant supplied music, the effect of **Track Type** (*Participant* vs. *Control*) on Musicon recognition and preference could be studied in Phase 2. Each participant was also asked to choose two ‘selections’ from each musical track (both their own tracks and the Control Tracks). The first task was to select the section that was their personal favourite part of the track (*Favourite*). The second was to select the section they felt was most representative of the track in general (*Essence*). Participants were asked to choose both *Favourite* and *Essence* to help us understand the different motivations behind the selection of the portion of music users might want to use for creating a Musicon from a known piece of music. Participants choose these sections on their own, using custom software. For each track, the software presented two slider bars (the knob on which corresponded to a five second slice of the song), one for ‘Favourite’ and one for an Essence section). Participants could adjust the sliders and play the selected clips until they were happy with their choices. Once they confirmed their selections, the software moved onto the next track. The order in which tracks were presented to participants was randomised. It was entirely possible that these two selections would overlap, or indeed be exactly the same. This, if it turned out to be the case, would in itself provide useful information.

Each of the sections selected were 5 seconds long. The decision to choose this length was made to balance the trade-off between how easy the task would be for participants and how much music there would be from which to generate Musicons. Choosing shorter selections could have been too difficult for

participants and having anything longer would have resulted in too much material from which to generate good Musicons.

Once participants had chosen all 20 of the five second sections (5 Control/Favourite, 5 Control/Essence, 5 Participant/Favourite and 5 Participant/Essence), six Musicons were generated from each – a *short* (0.2 second) Musicon from the start, middle and end of each section, and a *medium* (0.5 second) Musicon from the start, middle and end of each section. Two durations were used to analyse the effects of Musicon length on performance and preference.

The start, middle and end of the sections were used to generate a range of Musicons as we did not know where the most representative part within the section was located. Most of the songs selected by users could be described as, or as a sub-genre of, modern westernised pop or rock. Only one song was selected by more than one participant. Of all of the participant supplied tracks, there were only three fully instrumental tracks while the rest contained at least one singer. This resulted in a set of 120 Musicons for each participant, which was then evaluated with the same set of users in Phase 2. Each participant only evaluated his or her own set of 120 Musicons.

5. PHASE 2 – MUSICON RECALL TEST

The second phase of the study took the set of Musicons generated in Phase 1 and tested them with users to investigate recognition of, and preference for the set of Musicons. Phase 2 used a within-subjects design and took place during the same session as Phase 1. As introduced in Phase 1, the three Independent Variables were:

Track Type: whether a participant picked a piece of music from his/her own collection or whether it was from the control set (*Participant Track/Control Track*);

Selection: whether participants picked the section of the track as either favourite or essence (*Favourite/Essence*);

Length: the length of the Musicon (*0.2 s / 0.5 s*).

For each participant, Phase 1 produced 120 unique Musicons: 10 Tracks (5 Control and 5 Participant tracks) x 2 Selections (Favourite and Essence) x 2 Lengths (0.5s and 0.2s) x 3 Positions (Start, Middle and End). In Phase 2, participants were asked to listen to each of the Musicons and to identify the track from which it was created.

Musicons were presented in a randomised order. On hearing a Musicon, participants were asked to press a button on the experiment interface corresponding to the correct track. In total there were 10 buttons, one for each track in the experiment (5 control tracks, 5 participant tracks). This provided a measure of recognition performance for each Musicon. In addition, participants were asked to rate each of the Musicons in terms of preference. The three Dependent Variables measured were:

Identifiability: Whether or not the participant was able to correctly identify the track from which the Musicon was generated;

Number of Replays: Participants were allowed to replay each Musicon up to three times before submitting their answer. From this, it would be possible to investigate not only if a track could be identified but also how difficult it was to identify;

Preference: Participants were asked to rate each Musicon in terms of preference on a 5 point Likert scale (Strong Dislike, Dislike, Neutral, Like, Strong Like) based on whether they found the Musicon pleasant sounding.

5.1. Hypotheses

H1: Recognition rate for Musicons generated from *Participant Tracks* will be greater than those from *Control Tracks*. Measured by higher number of *correctly identified tracks* and a lower *number of replays*;

H2: Participants will have a higher *preference rating* for the Musicons from *Participant Tracks* than *Control Tracks*;

H3: Recognition rate for Musicons created from *Essence Selections* will be higher than that for Musicons created from *Favourite Selections*. Measured by the number of *correctly identified tracks* and the *number of replays*;

H4: Recognition rate for the *0.5s* Musicons will be higher than for the *0.2s* ones. Measured by the number of *correctly identified tracks* and the *number of replays*.

5.2. Results – Recognition Rate

The recognition rate of each Musicon is shown in Table 1. Totals shown are out of 225 (15 participants x 5 Songs x 3 Positions (Start, Middle and End). Musicons that performed better than others were more correctly identified with a fewer number of replays.

	Length	
	0.2s	0.5s
Control Tracks		
Favourite	73%(165)	94% (212)
Essence	78% (176)	89% (200)
Participant Tracks		
Favourite	69% (157)	84% (190)
Essence	72% (162)	85% (192)

Table 1: Number of correctly identified Musicons.

5.2.1. Identifiability

A three-factor, repeated-measures ANOVA on Track Type, Selection and Length for the number of correctly identified Musicons showed a significant main effect for Track Type ($F(1,74)=5.513, p<0.05$) and a significant main effect for Length ($F(1,74)=81.799, p<0.01$). The main effect for Selection was not significant ($F(1,74)=0.148, p=0.70$). There were no significant interactions, Track Type x Selection ($F(1,74)=0.278, p=0.6$), Track Type x Length ($F(1,74)=0.369, p=0.545$), Selection x Length ($F(1,74)=3.286, p=0.07$) and Track Type x Selection x Length ($F(1,74)=2.426, p=0.124$).

The Musicons generated from the Control Tracks were correctly identified significantly more often than those generated from the Participant Tracks and the 0.5s Musicons were correctly identified significantly more often than the 0.2s Musicons. This partially rejects Hypothesis 1 and partially confirms H4. There was no evidence for H3.

5.2.2. Number of Replays

A Musicon could be replayed up to three times. Figure 1 shows the total number of replays over all participants for the whole experiment. The average number of replays per Musicon was small ($M=0.51$, $SD=0.84$), however, as can be seen in Figure 2, the total number of replays for 0.2s Musicons was higher than the total number of replays for 0.5s Musicons.

A three-factor, repeated-measures ANOVA on Track Type, Selection and Length for the number of replays showed no effect for Track Type ($F(1,224)=2.113$, $p=0.147$), providing no evidence for H1. The main effect for Length was significant ($F(1,224)=125.55$, $p<0.001$), as was the main effect for Selection ($F(1, 224)=4.40$, $p<0.05$). There were no significant interactions (Track Type x Length $F(1,224)=0.159$, $p=0.69$, Track Type x Selection $F(1,224)=0.051$, $p=0.822$, Length x Selection $F(1,224)=0.722$, $p=0.397$, Track Type x Selection x Length $F(1,224)=2.154$, $p=0.144$). Musicons of 0.2s ($M=0.68$, $SD=0.93$) were replayed significantly more often than those of 0.5s ($M=0.29$, $SD=0.66$), partially confirming H4. Musicons generated from *favourite* Selections ($M=0.52$, $SD=0.86$) were replayed significantly more than *essence* Selections ($M=0.45$, $SD=0.79$), partially confirming H3.

5.3. Results – Musicon Preference

Friedman’s analysis of variance by ranks was used on the preference ratings. Differences across all factors were significant, $\chi^2(3)=403.067$, $p<0.001$. *Post hoc* pairwise Wilcoxon tests with Bonferroni correction were carried out. A significant difference was observed between Musicon Lengths, $p<0.001$ and between Song Type, $p<0.001$. In general, participants preferred 0.5s (Median Rating = Like) Musicons over 0.2s Musicons (Median Rating = Neutral) and participants preferred Musicons created from the Participant supplied songs (Median Rating = Like) over those created from the Control songs (Median Rating = Neutral). There was no evidence to suggest that Section, either favourite or essence, had any effect on the preference ratings.

5.4. Discussion

The hypothesis that recognition rate for Musicons generated from Participant Tracks will be greater than Control Tracks (H1) was not supported. The Control Tracks used in this study were chosen because they had strong thematic associations with popular culture for the participant group and the results confirm that this assumption was true. The accuracy for the Participant Tracks was 78% overall, which is good, but not as high as the rates observed for the Control Tracks in this experiment (83%) and in [1] (89%). This suggests that there may be something inherently more ‘identifiable’ about the Control Tracks carefully chosen by experts, or that participants were more able to pick easily identifiable sections from the Control Tracks.

The hypothesis that participants would perform better with 0.5s Musicons than with the 0.2s Musicons (H4) was supported. This also confirmed the results observed by McGee-Lennon *et al.* [1] who found the same result. That the 0.2s Musicons were replayed more than 0.5s ones suggests that participants found them more difficult to recognise and adds weight to the claim that 0.5s Musicons is the most appropriate length for a Musicon.

The hypothesis that participants would perform better with Musicons created from essence sections over favourite sections

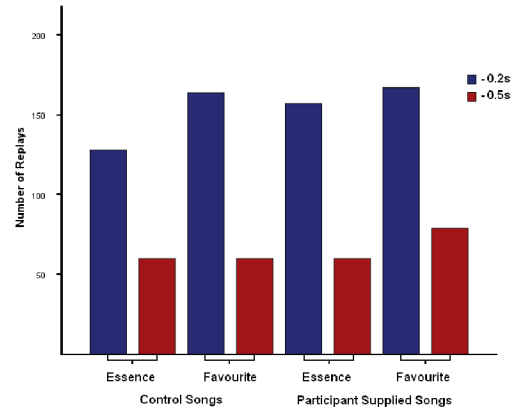


Figure 1: Overview of the number of Musicon replays.

(H3) was supported: there was no evidence to suggest that Selection had any effect on recognition rate but essence sections were replayed significantly less often than favourite ones.

The hypothesis that participants would prefer Musicons created from their own tracks over those created from Control Tracks (H2) was confirmed. The participants’ responses to the Control Tracks suggested that they did not find them unpleasant but that they simply did not feel strongly either way.

6. PHASE 3 – MUSICAL SECTION ANALYSIS

The results presented above do not reveal anything about the underlying nature of the 5 second sections from which the Musicons were created. In this phase, we address two questions: (1) what are the key properties of the sections that were chosen in Phase 1? and (2) are there any similarities between the sections? We performed a qualitative analysis in which we looked at where the 5 second sections chosen in Phase 1 occurred within whole track and what musical content they contained to understand if knowledge of the properties of the music within the section may contribute to the design of good Musicons.

The analysis was designed to identify the similarities between the musical sections chosen by participants. If we could spot features that were common across well liked and easily identifiable Musicons it might help in choosing the right parts of *any* given piece of music on which to base a Musicon. The qualitative analysis involved the experimenter listening to the sections several times and looking at the underlying musical properties of the sounds to identify common compositional features between the different favourite and essence sections.

The study of the composition of a piece of music is well established in the area of Musical Analysis [18]. This broad discipline is interested in identifying the fundamental parameters or elements of a piece of music. Such analysis can highlight the underlying similarities or differences between two pieces, styles or historical periods of music by considering aspects such as form, structure, timbre and harmony. We used this approach in our analysis. Four main categories of labels were used to drive the analysis. These were derived by one of the researchers before the analysis began, based on standard definitions of musical terms which can be found on Oxford Music Online [19] and are now discussed in turn.

Structural Features: These are features relating to how the piece of the music is structured and, more specifically, where a

particular 5-second section falls within the structure. High-level structural features, such as introductions, verses and refrains, are examples. Such features are useful since they, if found to be relevant, would provide a pointer to a specific passage within a piece of music that shares a similar structure.

Timbral Features: The timbre of a piece of music refers to the overall sound and is normally defined as properties of the sound independent of rhythm or pitch. For the purposes of the analysis this is defined in terms of what instruments are present or absent with respect to the entire track, which will allow us to assess how ‘full’ or ‘empty’ the sound of this particular section is with respect to the rest of the track.

Melodic Features: These would describe whether the 5-second section contains any prominent melodic riffs, motifs or repeated melodic lines in the piece. These could be either instrumental or vocal.

Tonal and Rhythmic Features: These are features describing the salient tonal or rhythmic features of the sections. These could include, for instance, modulations (where the pitch of the track is changed substantially for effect), changes in tempo or prominent rhythmic patterns.

It was useful to augment each label with an indication of where the section lay within the whole track. For example, if a section was labelled ‘Chorus/Refrain’, it was useful to specify whether it was positioned nearer the start or end of the Chorus/Refrain. That the section was positioned to contain the *very* start of the chorus also proved salient (where *very* indicates that the section included the *absolute* starting point of the label, e.g. Chorus/Refrain, or contained the transition from the previous structural label, e.g. the transition from the Verse to the Chorus/Refrain). Subsequently, the augmentations ‘Start’, ‘Very Start’, ‘End’ and ‘Very End’ were included for each label.

The categories outlined above were used to guide the analysis, though the principles of Grounded Theory [20] were used to allow additional categories or themes to emerge from the data. The researcher listened to the 5 second sections several times and labelled each with as many of the features that were applicable. On each iteration, if it became clear that there were a number of sections with a common feature that was not currently being considered in the analysis, those sections would be labelled with that feature, and the feature would be considered for all sections on the next iteration. When no new features emerged, the analysis stopped.

6.1. Results

Each of the 5 second sections was labelled descriptively by the experimenter according to the underlying qualitative musical properties of each section. An overview of the labels and their frequencies can be found in Table 2 (labels with less than 5 occurrences have been omitted for brevity).

6.1.1. Control Tracks

The Control tracks were the same across participants (and did not come from the participant’s own music collection). We were primarily interested in how to create Musicons from a user’s own music collection. Therefore, the control tracks were not considered alongside the participant supplied songs in the detailed analysis. However, the Musicons generated from Control tracks were correctly identified more often than those created

from the participant supplied ones, which either suggests that there may be something inherently more ‘identifiable’ about them, or that participants were better at picking easily identifiable sections from the Control tracks.

The 5 second sections that were chosen from the Control tracks were remarkably similar over all the participants. For example, of the 5 second sections chosen from The Rembrants ‘I’ll be there for you’, 40% were of the main introduction guitar riff and 37% were of the section of the chorus during which the lyric ‘I’ll be there for you’ is sung, while only 23% of the sections were chosen to be from other parts of the song. Similarly, of the 5-second sections chosen from Ray Parker Jr ‘Ghostbusters’, 53% of the sections were chosen from the verse (either where the vocalist begins to sing, or where the word ‘Ghostbusters’ is sung) and 37% of the sections were of the main instrumental riff, while only 10% of the sections were chosen from other parts of the song. The trend is similar for the James Bond Theme, though does not hold for either John Williams ‘Theme from Jurassic Park’ or Johan Pachelbel ‘Canon’. The exact reasons for why the pattern is not repeated for these tracks is unknown, but both of these tracks do not contain vocals, are more classical in nature and do not have the same general structure as the western pop songs. It could be the case that the participants were more familiar with the Friends and Ghostbusters tracks, or with western pop/rock in general, and were subsequently able to make better selections. Although no strong conclusions can be drawn, it is still interesting to note the similarity between the sections. It suggests that if there are many people who are familiar with a particular song, they may have similar views on what is ‘representative’ of that song.

6.1.2. Participant Tracks

The majority of the labels emerging were structural in nature. Structural labels were useful in this context as they were able to transcend musical differences in genre, melody, rhythm, timbre and other intrinsically musical properties with which a composer makes a track unique. Structural similarities can group very disparate pieces of music and thus are useful for Musicon analysis. Since almost all of the user contributed songs were examples of modern western pop or rock, they were all structured in a similar way. Each song normally featured an introduction section, followed by one or more verses which were then followed by a chorus/refrain. Therefore, identifying which structural segment (e.g. introduction/verse/chorus) the 5 second section fell into was a useful way of identifying similarities between all of the 5 second sections. In total there were 150 sections (15 participants x (5 Favourite sections + 5 Essence sections)).

In addition to the structural labels, a number of melodic and timbral labels emerged as salient. The melodic labels generally indicated the presence of a strong or prominent melodic feature, such as a main riff (e.g. the main riff in Stevie Wonder ‘Superstition’ or in blink-182 ‘Apple Shampoo’) or instrumental solo (e.g. the guitar solo in Santana ‘Smooth’, or the brass solo in Louis Prima ‘Angelina, Zooma, Zooma’). The timbral features that emerged as salient generally distinguished between the presence or absence of vocals in the section. Of all of the participant supplied tracks, there were only three fully instrumental tracks while the rest contained at least one singer. Of the tracks with vocals, whether the participant’s chose sections that featured the singer proved highly salient.

The most frequently observed property was the presence of a vocalist, observed in 73% of sections. In modern pop or rock music, the vocalist is often carrying the main melody. Thus, picking a section of the track containing the vocalist is important

Label	Category	Frequency
Vocals	Timbral	109
Chorus/Refrain	Structural	48
Main Riff	Melodic	44
Instrumental	Timbral	41
Verse	Structural	36
Contains Track Title	Timbral	31
Chorus/Refrain – Very Start	Structural	31
Main Riff – Very Start	Melodic	30
First Verse	Structural	29
Introduction	Structural	25
Verse – Very Start	Structural	19
First Verse – Very Start	Structural	19
Full Instrumentation	Timbral	13
Introduction – Very Start	Structural	13
Middle 8	Structural	9
Instrumental Solo	Melodic	9
Climactic End-Section	Structural	8
Main Melodic Theme	Melodic	6
Chorus/Refrain – Very End	Structural	5

Table 2: Occurrences of labels in the analysis [19].

to picking a section that contains the main melody - a feature from which the track may be easily identified.

The next two highest ranking labels were ‘Chorus/Refrain’ (32% of sections) and ‘Main Riff’ (29%), which account for the highest ranking structural and melodic labels. Both the Chorus/Refrain and the Main Riff are also typically representative of western modern pop or rock music. There were a number of labels which appeared nearly as frequently as both ‘Chorus/Refrain’ and ‘Main Riff’. The label ‘Verse’ appeared frequently (24%), as did ‘Introduction’ (17%).

Labels augmented with ‘Very Start’ also occurred frequently. If a 5 second section contained the very start of the Chorus/Refrain it was labelled with both ‘Chorus/Refrain’ and ‘Chorus/Refrain – Very Start’. From this it was possible to analyse the proportion of 5 second sections *within* a particular label (e.g. all the 5 second section that were labelled with Chorus/Refrain) that were also labelled with an indication of position (e.g. Start, Very Start, End, Very End). As can be seen in Table 2, two of the highest ranking labels (Main Riff and Chorus/Refrain), have a high proportion of labels with an indication of position. 64% of all sections labelled with ‘Chorus/Refrain’ were also labelled with ‘Chorus/Refrain – Very Start’ and 68% of all sections labelled with ‘Main Riff’ were also labelled with ‘Main Riff – Very Start’. This pattern continued with the labels ‘Verse’ (53% also have ‘Verse – Very Start’) and ‘Introduction’ (52% also have ‘Introduction – Very Start’). All labels with this pattern are either Structural or Melodic in nature. The data suggest that if a melodic or structural feature is identified as highly representa-

tive of the track, it is likely that the *very start* of that melodic or structural feature is considered highly representative of the track.

The data suggest that there was a preference for sections that appeared nearer the beginning of tracks (sections between the introduction and first chorus). For example, labels such as ‘Middle 8’ (6% of sections), ‘Climactic End-Section’ (characterised as a unique section, appearing at the end of a song that is normally intense/exciting - it acts as a climax to the song) (5% of sections) and ‘Outro’ (1% of sections) occurred infrequently.

The sample of user supplied music in the study was almost entirely limited to western popular music. It is true that many underlying similarities were discovered in the 5 second sections chosen from these tracks; however, the presence of these similarities cannot be extended beyond this musical genre. This can be demonstrated with one track featured in the experiment: Duke Ellington and John Coltrane’s ‘The Feeling of Jazz’. This piece does not share many of the features with the other tracks in the study: it is not structured in the same way, nor does it contain any of the same salient features. The participant who chose this piece picked the very start of the introduction as his or her *Essence* selection and a section labelled ‘Instrumental Solo’, which occurred roughly half way through the track, as his or her *Favourite* selection. However, since the track is not structured in the common ‘Introduction-Verse-Chorus’ form of Western Pop/Rock music, it was difficult to draw strong comparisons between this track and all of the others in the experiment.

Overall, there was a great deal of similarity between the selections made by participants across the songs used in the experiment, suggesting that there may be common musical features that can be used to aid the selection of music from which Musicons that are representative of the piece can be created. The most frequently appearing label in the Musicon analysis was *Vocals*, suggesting that when selecting sections from which to make Musicons, the presence of a vocalist is a property that people consider representative.

7. MUSICON GUIDELINES

From the results of the previous phases the following guidelines for the design of Musicons can be drawn out:

Track Type: Musicons created from tracks that are both familiar to and liked by the user for whom they are intended are more likely to be *preferred* over those created from more generally well known tracks. Therefore, Musicons can be created by sampling snippets of music from tracks chosen by the end user to ensure a higher and more stable level of preference. However, this comes with a trade-off in performance – Musicons from participant supplied tracks were not identified as accurately as those from well known tracks. Future research should aim to investigate whether the trade-off in performance and preference changes over time, once the participant has become more familiar with the stimuli.

Length: Experimental evidence suggests that Musicons which are 0.5s in length are identified correctly and well liked.

Musical Properties: The presence of vocals was the most common feature selected by participants. Choosing a section with vocals is likely to give good Musicon performance if using western pop/rock music.

Start of Chorus/Refrain: It was common for users to select a passage of the track containing the very beginning of the first chorus or refrain. Therefore, Musicons should be sampled from a section of the track that contains vocals and the beginning of the first chorus or refrain, if using western popular music.

Start of any Melodic or Structural Feature: Although the chorus/refrain was the most popular passage in our study, there were others that were selected almost as often. If any melodic or structural feature is identified as highly representative of the track, it is likely that the *very start* of that melodic or structural feature is also considered highly representative of the track. Therefore, when sampling a Musicon from *any* Structural or Melodic passage, sample from the very start of that passage.

8. CONCLUSIONS

This research has demonstrated that by allowing users to self-select subjectively representative sections from their own music tracks, identifiable and well liked Musicons can be created. Furthermore, it was also observed that the self-selected sections were similar enough in their underlying musical features to allow for the possibility of automatic Musicon generation from an arbitrary piece of music.

Future work on Musicons is underway to focus on how well the above design guidelines can be used to create Musicons with performance and preference rates comparable to the ones observed here. Work is also planned to study how well Musicons scale and whether there is an upper limit on the number of Musicons a person can effectively remember.

The effectiveness of Musicons, (both performance and preference) compared to different types of audio stimuli, such as Earcons, Auditory Icons or Spearcons should be further investigated. The guidelines for the design of Musicons presented here provide a starting point for further investigation into the usefulness of Musicons as audio stimuli and deepen our understanding of their structure and basic composition and how this might be used to inform the design of novel auditory interfaces.

9. REFERENCES

- [1] M. McGee-Lennon, M. Wolters, and R. McLachlan, "Name that tune: Music as reminders in the home," in *ACM CHI 2011*, 2011.
- [2] B. Walker and A. Nance, "Spearcons: Speech-based earcons improve navigation performance in auditory menus," *ICAD 2006*, pp. 63-68, 2006.
- [3] E. G. Schellenberg, P. Iverson, and M. C. McKinnon, "Name that tune: identifying popular recordings from brief excerpts," *Psychonomic bulletin & review*, vol. 6, no. 4, pp. 641-6, Dec. 1999.
- [4] M. A. Bartsch and G. H. Wakefield, "Audio thumbnailing of popular music using chroma-based representations," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 96-104, 2005.
- [5] M. Levy, M. Sandier, and M. Casey, "Extraction of high-level musical structure from audio data and its application to thumbnail generation," in *ICASSP*, 2006, vol. 5, no. 1.
- [6] H. Nawata, N. Kamado, H. Saruwatari, and K. Shikano, "Automatic musical thumbnailing based on audio object localization and its evaluation," in *ICASSP 2011*, 2011, no. 4, pp. 41-44.
- [7] W. Gaver, "Auditory interfaces," *Handbook of human-computer interaction*, 1997.
- [8] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human-Computer Interaction*, vol. 4, no. 1, pp. 11-44, 1989.
- [9] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "A detailed investigation into the effectiveness of earcons," in *ICAD*, 1992, vol. 18, pp. 471-498.
- [10] A. Baddeley, "Human memory: Theory and practice," p. 423, 1997.
- [11] S. Garzonis, C. Bevan, and E. O'Neill, "Mobile service audio notifications: Intuitive semantics and noises," in *OZCHI*, 2008, pp. 156-163.
- [12] B. Walker and A. Kogan, "Spearcon performance and preference for auditory menus on a mobile phone," in *UAHCI*, 2009, pp. 445-454.
- [13] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "Experimentally derived guidelines for the creation of earcons," in *Adjunct Proceedings of HCI*, 1995, vol. 95, pp. 155-159.
- [14] S. A. Brewster, A. Capriotti, and C. V. Hall, "Using compound earcons to represent hierarchies," *HCI Letters*, vol. 1, pp. 6-8, 1998.
- [15] D. Rocchesso, *Explorations in Sonic Interaction Design*. Logos Verlag Berlin, 2011.
- [16] A. Butz and R. Jung, "Seamless user notification in ambient soundscapes," in *IUI '05*, 2005, p. 320.
- [17] M. A. Bartsch and G. H. Wakefield, "To catch a chorus: Using chroma-based representations for audio thumbnailing," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, no. October, pp. 15-18.
- [18] I. Bent and A. Pople, "'Analysis,'" *Grove Music Online. Oxford Music Online*. Oxford Music Online.
- [19] "Oxford Music Online." [Online]. Available: <http://www.oxfordmusiconline.com>. [Accessed: 22-Sep-2011].
- [20] B. G. Glaser and A. L. Strauss, *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publ., 1977.