

Using Card Sorting to Design Faceted Navigation Structures

Ed de Quincey
School of Computing and Mathematics
Keele University
e.de.quincey@keele.ac.uk

James Mitchell
School of Computing and Mathematics
Keele University
j.a.mitchell@keele.ac.uk

This paper describes how a variation of card sorting- ‘repeated single-criterion sorting’, can be applied to the information architecture design of digital music services. 52 respondents were asked to sort, using their own choice of criteria, 12 popular songs using an online card sorting tool. Once respondents had chosen a construct for a particular sort, e.g. “Genre”, they placed each card into a named category, e.g. “Rock”, “Pop”, and were encouraged to repeat this process until they could think of no more constructs. High levels of agreement were found for a small number of constructs such as “genre”, “gender”, and “speed of song” but the remaining constructs were individual to each respondent, e.g. “songs that make me cry”. The results highlighted differences with current approaches to music categorisation, as well as the potential for repeated single-criterion sorting to be used to design faceted navigation structures.

Requirements Elicitation, Card Sorting, Information Architecture, Faceted Navigation, User-Centered Design (UCD), User Experience (UXD, UX).

1. INTRODUCTION

User experience design (UXD) and usability evaluation are supported by a range of tools and techniques. However, a number of these methods utilise attributes predetermined by the researcher, such as heuristic assessment (Nielsen & Molich, 1990) and cognitive walkthroughs (Wharton *et al.*, 1994), which may not be relevant to the intended audience. To tackle this, methods that elicit criteria from the users themselves, such as think-aloud protocol (Nielsen, 1992), card sorts (Rugg & McGeorge, 1992), and laddering (Gammack, 1987) have been developed.

A popularised version of card sorting, termed “all-in-one” sorting, has become one of the standard techniques in the User-Centered Design (UCD) process (Usability.gov. N.D). “All-in-one” sorting improves “findability” within a system and has been successfully used to determine the Information Architecture (IA) of websites (Spencer & Warfel, 2009) (e.g. Frederickson-Mele, 1997; Tullis, 2003; Tullis & Wood, 2004) and software menu systems (Tullis, 1985). In addition, a variation of card sorting, ‘repeated single-criterion sorting’, has been highlighted (Maiden, 2009) for its potential in requirements elicitation, but few empirical studies investigate how this variation of card sorting relates to UXD and how it can be integrated into the UCD process.

In order to address this, this study has applied repeated single-criterion sorting to the problem of music categorisation and the design of digital music services.

This paper is structured as follows. Section 2 gives an overview of different card sorting techniques and work related to music categorisation. Section 3 outlines the method used, respondents and materials. Section 4 presents the results and Section 5 the analysis of these results. Section 6 discusses the implications of the study and Section 7 presents overall conclusions.

2. BACKGROUND

This section describes the different card sorting methods and current academic and commercial approaches to music categorisation.

2.1 Card Sorting

All-in-one sorting

This category of sorting covers a wide range of methodologies. The consistent feature is that the respondent(s) sorts the set of entities once. The respondent(s) are given a set of cards which represent products, pages, or functionality, within the site or application. The number of cards depends on the entity (Courage & Baxter, 2005). Up to five hundred cards however have been sorted in

previous studies (Tullis, 1985). The respondent sorts the entities into a set of pre-defined categories, or categories of the respondent's choosing and name them. In the case of closed sorting the category that each entity has been sorted into is recorded; for open sorting the names of the categories are recorded along with the category position of each entity. Open card sorting "is useful as input to information structures in new or existing sites and products" whereas "closed sorting is useful when adding new content to an existing structure, or for gaining additional feedback after an open card sort" (Spencer & Warfel, 2009). The advantages include savings in time and cost, and as it is user-centered, it is therefore not as susceptible to "gut-feel" biases (Spencer & Warfel, 2009). Disadvantages include that the methodology is content-centric and fails to take in to account the users' task i.e. how the user interacts with content on a site and that analysis can be time-consuming- especially with large numbers of cards and/or respondents (Spencer & Warfel, 2009).

"All-in-one" sorting is generally used to determine IA and uses variations of cluster analysis to determine the navigational structure of a website or application. It has been used successfully in the design of large-scale websites e.g. the Google AdWords Help Center (Nakhimovsky et al., 2006) and is now an integral part of the analysis and design stages of a UCD lifecycle (Bevan, 2003).

Repeated single-criterion sorts

Repeated single-criterion sorting (or open card-sorting) involves asking the respondent to sort entities into groups of their own choosing; then to sort again, using a different criterion of their own choosing, until they run out of criteria. For example, if the entities are shoes, "colour" might be the first criterion, with categories such as "brown", "black" and "white". They may be further sorted by "material" as the criterion, and into categories such as "leather" and "canvas". Empirical research has found no statistical difference between the types of criteria and categories elicited when using different types of entity (Rugg et al., 1992).

This approach works well with nominal categories, and typically elicits group names and criteria consisting of short phrases. This method is described by (Gammack, 1987) and is described in detail in a tutorial paper by (Rugg & McGeorge, 1997). The technique has been applied to a wide range of topics, including web page quality metrics (Upchurch et al., 2001); quantification of copyright infringement (Martine & Rugg, 2005), and assessment of differences between expert and student programmers (Sanders et al, 2005). It is supported by a range of statistical analyses, including co-occurrence matrices (Martine & Rugg, 2005) and minimum edit distances (Deibel et al., 2005).

Rugg & McGeorge (1997) recommend the use of repeated single-criterion sorting for requirements elicitation due to its flexibility and a stronger grounding with the relevant theoretical foundation i.e. Personal Construct Theory (PCT). Maiden (2009) has also highlighted the value of this type of sort for requirements elicitation for similar reasons. Repeated single-criterion sorting has been used to study perceptions, such as identifying the features of web pages that users are interested in (Upchurch et al., 2001). However, this method is not formally aligned with the UCD process in the same way that "all-in-one" sorting is.

Music Categorisation

The ease of purchasing and streaming music online and the shift towards storing and organising music digitally has shaped music preference behaviour (Greasley & Lamont, 2006). Studies investigating people's use of music have indicated that people listen to music for specific reasons and that their motivations for listening to music depends on context (DeNora, 2000; North et al., 2000; Sloboda et al. 2001). However, the ubiquitous nature and ease of access to music presents problems: How to organise music so that it is accessible, and convenient? How to discover more songs, similar to those they enjoy?

Musical genre is a widely used standard for categorising music (Aucouturier & Pachet, 2003; Pachet & Cazaly, 2000) and often the preferred technique. However, the definition of a music genre is subjective since it is influenced by extrinsic factors (Lippens, 2004; Aucouturier & Pachet, 2003). This leads to undefined boundaries of genres and as a consequence there is a lack of a precise method of classifying music to genres. Online stores such as iTunes categorise music by standard music industry decided taxonomies (similar layout of traditional bricks and mortar retail stores) with genre being the primary method for users of the software to find songs that they like. The user has to have a definite idea of what genres they like, what genre a song fits into and for this to match with the categorisation used by the online store.

One of the most commercially popular attempts at improving musical classification with the intention of creating automated playlists for streaming radio is Pandora, based upon the Music Genome Project (Joyce, J., 2006). The Music Genome project attempts to describe music with vectors consisting of hundreds of genes or musical attributes describing each song (McKay, 2010). It is unclear what the complete list of attributes are, but a partial list (that is now not publicly available) suggested that the following are included: Structures; Roots; Tonality; Instrumentation; Feel; Musical qualities; Leanings/styling; Recording techniques; Influence; Instruments; Lyrical content and Vocals.

A song is represented by a vector of these attributes, with up to five hundred genes/attributes forming the vector. Each attribute/gene is assigned a number between 1 and 5 in half-integer increments (see Music Genome Project US Patent: No. 7,003,515). Given a single song or a group of songs, a distance measure is then calculated from this vector to produce a list of related songs. Though the retrieval process is automated, the scaling and classification of the songs is an entirely manual, subjective process.

There have been attempts at creating flat taxonomies of music using folksonomies or tagging e.g. Last.fm, which utilise the user generated attributes to categorise songs, increasing the number of attributes that can then be used to describe a song and potentially produce a richer categorisation schema. These rely solely on the individual's perceptions of a song and depend on users using the same tag or set of tags for the same songs if they can be then used to recommend similar songs.

For the majority of digital music services, there is a reliance on genre-based systems which are reliant on inconsistent, intrinsic features of a song. There is little research into appropriate feature sets for classifying different types of song, and the relationship between objective and subjective attributes that users classify songs by. There is also little exploration into which of these features can be supported by current technologies or by extending pre-existing functionality as opposed to creating entirely new systems. Card sorting offers a potential solution to these problems as it has been previously used to determine user perceptions of a range of media (e.g. Upchurch et al., 2001) and also similarity measures using a range of user identified attributes (Martine & Rugg, 2005).

The following sections describe the use of repeated single-criterion card sorting to identify users' perceptions of a range of popular songs.

3. METHODS

As part of a related study (de Quincey, 2010), an online card sorting tool has been developed. The application includes functionality to support a range of multimedia such as pictures, music and videos along with analysis techniques such as co-occurrence matrices.

Figure 1 shows the sorting interface where respondents are presented with the entities on the left-hand side (in this case artist names) and are asked to input the sort criteria (in this case "Gender of Artist") and sort the entities into groups of their choosing (in this case the user has created 2 groups, "Male" and "Female"). Users can then drag the entity into the appropriate group (see Figure 2).

For multimedia entities, users double-click the card to view/hear the video/song.

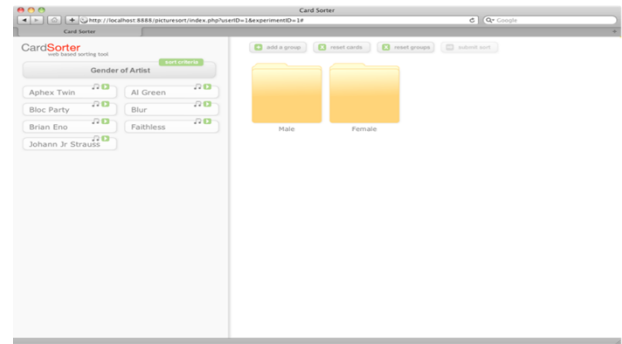


Figure 1: Screenshot of online card sorting tool



Figure 2: Screenshot showing sorting behaviour

Once the entities are sorted, the user is prompted to either perform another sort using a different criterion or to end the sort. The results are recorded automatically in a database

Respondents were asked to use the card sorting tool to sort, using their own choice of criteria, a number of popular songs. Once respondents had chosen a criterion for a particular sort, they placed each card into a named group and were encouraged to repeat this process until they could think of no more criteria. The researcher used a dyadic elicitation technique which involved playing two random music clips and asking the respondent whether they could think of any differences between the two songs which could form another sort criterion. The sessions were carried out under controlled conditions in the same room, using the same computer to remove unforeseen technical issues that may occur e.g. user not able to hear songs, slow internet connection etc. The sessions were undertaken between 2006 and 2007.

3.1 Respondents

There were a total of 52 respondents, 42 from the School of Psychology student pool and 10 from within the School of Computing and Mathematics. 32 females and 14 males participated (plus 6 participants who did not provide information on their gender). The participants' age range was 18 years to 37 years.

3.2 Materials

Twelve popular songs were chosen for the entities by researchers in the School of Psychology to

complement research that was being undertaken by (Greasley & Lamont, 2006). Songs were from popular artists and represented a range of genre (see Table 1).

Table 1. Songs and artists used as entities for sorting

ID	Artist - Song
1	Coldplay - Yellow
2	Eminem – Without Me
3	Misteeq – Why?
4	Rage Against the Machine - Wake Up
5	Maroon 5 - This is Love
6	UB40 - Red Red Wine
7	De La Soul - Three
8	Hard Fi - Living for the Weekend
9	Madonna – Hung Up
10	Chemical Brothers - Galvanise
11	Tracy Chapman – Fast Car
12	Mary J Blige – Family Affair

Each song was cropped to the first thirty seconds of the song. The screen representation of each song (see Figure 1) that users double click to play was labelled with an arbitrary number between one and twelve. Using the title or artist as the card label was considered but this may have prompted criteria related to the song title or the artist, not the song itself.

4. RESULTS

This section presents the results from the 52 card sorting sessions, outlining the constructs and categories used and their distribution between respondents.

4.1 Number of constructs and categories

A total number of 295 constructs¹ were elicited from 52 respondents. The number of constructs per session ranged from 2 to 11. Respondents used between 2 and 9 categories for each sort with the majority of sorts comprising of dyadic (2 categories used) and triadic (3 categories used) sorts.

4.2 Commonality of constructs

From the 295 constructs, there was direct verbatim agreement i.e. two or more respondents using the exact same phrase, for 28 constructs. The most frequently used verbatim constructs are shown in Table 2. When scrutinising the verbatim constructs it was apparent that different respondents used different words for similar constructs e.g. “Music Type” and “Type of music”, “Tempo” and “Speed”. In line with previous research (e.g. Gerrard &

Dickinson, 2005), an independent judge was used to group the constructs into superordinate constructs, giving an indication of commonality between respondents.

Table 2. Commonly used verbatim constructs

Verbatim Construct	Number of Subjects
Genre	23
Tempo	12
Gender of singer	10
Gender	9
Speed	7
Type of music	7
Artist(s)	5
Music type(s)	5
Style	5
Audience	4
Era	4
Group or solo	4

The judge was given a set of standard instructions and a full set of results including the constructs, category names and card groupings. A number of previous studies provided a list of grouped verbatim constructs to the independent judge. As identified in a previous study (de Quincey, 2010) this should remove constructs that use the same words but are related to different attributes (category names).

Table 3. Commonly used superordinate constructs

Superordinate Construct	No of constructs	% of Respondents
Genre of Music	45	88
Gender of Artist	34	67
Speed of song	27	53
Solo or group	18	35
Year music produced/released	16	31
Likeability of song	13	25
Main instrument	9	18
Audience	8	16
Emotion	7	14
Nationality of artist	7	14

When grouped into superordinate constructs, the number of constructs was reduced from 289 to 78. Table 3 shows the number of constructs included for each superordinate construct and the percentage of respondents that they were elicited from for the top 10 most used constructs. Following this grouping, agreement was found amongst respondents for 26 superordinates out of the 78 e.g. 88% of respondents used the “Genre of Music” as a construct and 67% of respondents used “Gender of Artist”. This shows a high level of commonality for a

¹ Criteria elicited during card sorting sessions of this type are normally known as constructs due to the link with PCT.

small number of the constructs with 52 constructs out of the 78 being generated by single users. Examples of these unique constructs included "Volume of drums", "Complexity of music", "Make you sad", "Is it relaxing" and "Music to work to".

4.3 Distribution of Items

As described by Martine & Rugg (2005), card sorts data can be used to produce co-occurrence matrices that give an indication of similarity between the entities represented by the cards and the distribution of entities for similar constructs. The matrix is produced by summing the number of times one card appears in the same category as another card for all respondents and criteria.

	1	2	3	4	5	6	7	8	9	10	11	12
1		85	50	119	169	150	110	151	60	92	151	65
2			115	102	104	87	119	118	129	167	63	134
3				58	80	61	79	97	181	132	61	174
4					114	104	121	147	52	116	89	65
5						112	130	176	108	109	94	92
6							127	95	49	97	155	76
7								123	76	136	98	93
8									100	131	81	83
9										116	75	146
10											61	126
11												79
12												

Figure 3. Co-occurrence matrix for the 12 songs

Figure 3 shows the matrix for the twelve songs. With the total number of sorts being 289, the maximum number of times for two cards to be placed in the same category is also 289. The highest amount of co-occurrence was 181 (for songs 9 and 3). The lowest of amount of co-occurrence was 49 (for songs 9 and 6).

To determine levels of agreement within superordinates i.e. whether respondents have placed the same songs in the same groups for similar criteria, co-occurrence matrices were calculated for the sorts where a number of respondents had used similar criteria as determined by the independent judge. The matrices in Figures 4 and 5 show the percentage of times that two songs were placed in the same group to allow for comparison between matrices. For example, in Figure 2, songs 1 and 4 were placed in the same group 94% of the time for sorts related to "Gender".

For these matrices, if respondents agreed on the criterion, the matrix would at best case contain a selection of very high numbers (close to 100%) and very low numbers (close to 0%). This would indicate that certain songs would always be placed in the same group for that criterion and others would never be placed in the same group. Plotting the histogram of this data should demonstrate bimodal distribution.

For the superordinate "Gender" as shown in Figure 5, this is mostly the case. For example, songs 1 and 2 are placed in the same group 91% percent of the time when the criterion for the sort is related to "Gender" but songs 1 and 3 are never placed in the same group.

	1	2	3	4	5	6	7	8	9	10	11	12
1		91	0	94	94	94	94	94	0	94	41	0
2			0	88	91	94	94	91	3	94	38	3
3				0	0	0	0	3	97	0	53	97
4					91	88	91	91	0	91	35	0
5						88	94	94	0	94	35	0
6							94	88	0	94	44	0
7								94	0	100	41	0
8									3	94	41	3
9										0	53	100
10											41	0
11												53
12												

Figure 4. Co-occurrence matrix for "Gender" (values represented as %)

When respondents were sorting based on gender, then male vocalists would always appear in the same group and never appear with songs with female vocalists and vice versa. For the majority of songs this was the case, but song 11 was placed with most of the songs between 35% and 53% of the time. It seems that respondents were unable to consistently group the song into a specific gender. This may be due to the vocalist, Tracy Chapman, having an undeterminable voice.

For the other superordinates levels of agreement are less consistent. "Genre" related constructs were used by 88% of respondents, but the level of agreement between the respondents indicated by the matrix shown below are low. Songs 5 and 8 were sorted into the same group 60% of the time (which was the highest) but the majority of songs had low levels of co-occurrence.

	1	2	3	4	5	6	7	8	9	10	11	12
1		0	2	20	44	18	13	53	20	2	40	2
2			16	2	0	2	22	0	2	33	2	38
3				4	7	7	11	9	31	27	9	53
4					7	9	9	24	2	4	16	0
5						11	22	60	42	9	24	7
6							24	7	11	16	40	2
7								20	11	36	29	11
8									11	9	24	4
9										27	11	13
10											16	31
11												0
12												

Figure 5: Co-occurrence matrix for "Genre" (values represented as %)

The matrix for "Solo or Group" showed higher levels of agreement with a number of songs (5 and 1, 9 and 2, 2 and 11, 8 and 1) co-occurring over

90% of the time. The matrix for “Speed of Song” highlighted some songs that are perceived to be similar (9 and 3, 11 and 1) but the majority show low levels of co-occurrence. For this construct it would be expected that songs would form clear groups based on the time signature that they were written in e.g. 3/4 time compared to 4/4 time but that does not seem to be the case. Matrices for the constructs “Likeability of Song” and “Year Produced” show similar distributions to “Speed of Song” with some songs with high levels of similarity, but the majority having low levels of co-occurrence.

5. ANALYSIS

5.1 Constructs used

Respondents generated a large range of constructs (2 to 11) and categories (2 to 9). The large number of constructs generated by some respondents suggests that they have expert knowledge in the domain (Rugg & McGeorge, 1997). This might be expected from the student population that the respondents were recruited from as some were studying music or involved with university orchestras etc.

High commonality (>50%) was found for a small number of superordinate constructs such as “genre”, “gender” and “speed of song” but the remaining 75 constructs showed little agreement. This gives an indication that once “genre” and “gender” have been used, further constructs are individual to each respondent. Of these remaining constructs, there is a mixture of objective, such as “age of artist” and subjective criteria such as “would I pay to see them in concert”.

5.2 Sorting behavior

There was little agreement between respondents in how they sorted the songs into categories for all criteria. The maximum number of times that two songs were placed in the same group was 181 (for songs 9 and 3) out of a possible 289 (62% of the time). This may be due to the songs being entirely different, different perceptions of the group a song fits into, or respondents using different criteria. The co-occurrence matrices demonstrate a range in the levels of agreement between respondents when sorting using the same/similar criteria.

Genre: Almost all of the respondents used “genre” as a criterion but there was little agreement between which songs fitted into the same genres or what those genres should be called. This indicates that due to “genre” being the default index method in music retail, people use what they are accustomed to despite there being little agreement into which categories particular songs fall into.

Gender: When sorting using “Gender” as the criterion, respondents were consistent with the majority of songs, finding it easy to determine the gender of the vocalist in all but one case, but it is sometimes unclear as to what “gender” refers to. For the majority of constructs in this study the gender refers to the gender of the vocalist, but the distribution of songs may become inconsistent if there is more than one vocalist for a song.

Solo or Group: Surprising the co-occurrence levels for “solo or group” were again low, especially considering that the criterion is highly objective. This could indicate a respondent’s lack of knowledge of the song artists or the inability to determine from a thirty second clip whether it is a solo or group artist.

Speed of Song: The co-occurrence levels for “speed of song” were also low indicating that respondents are inconsistent in their perceptions of what constitutes speed, even though the majority used the word “tempo” as the criterion name and “fast”, “slow” and “medium” as categories. This echoes research by Scheirer (1998).

5.3 Summary

Some agreement was found in the criteria used for sorting but there were many constructs that were unique. Within the categories, agreement was only found for the “gender” construct, with some agreement for certain songs when using certain criteria. The differing levels of agreement regarding song categorisation have implications for digital music services that are described in the following section.

6. DISCUSSION

6.1 Implications for Digital Music Services

Having identified that there is some level of agreement between respondents for the criteria used to sort music, one possible use of these constructs would be to then integrate them into music library/streaming software such as iTunes, to improve the ability of users to organise and navigate music libraries. The following table (Table 4) details constructs that could potentially be used as attributes for categorising music that are already utilised by iTunes, Spotify or the ID3 specification. ID3 frames are a popular “audio file data tagging format” that are used by a number of popular music players (O’Neill, 2013). An ID3 tag is a data container within an MP3 file that is stored in a predefined format allowing users and artists/vendors to encode additional information into an MP3 file such as text or pictures. Currently iTunes, and similar software, do not use all of the frames that are defined within the standard and therefore specialised ID3 tag editors have to be used to edit the majority of the ID3 frames (although iTunes does

allow the user to edit certain fields and automatically populates them via the iTunes store or via the Internet).

Table 4. Constructs used in Digital Music Services

Construct	iTunes	Spotify	ID3
Genre of Music	Yes	Yes	Yes
Speed of song	Yes "BPM"	No	Yes: "Exact tempo codes"
Year music produced/ released	Yes	Partial: "Decades"	Yes "Date/Year of recording"
Likeability of song	Yes "iTunes Rating"	Partial: "Add to My Music"	Yes "Pupularimeter"
Emotion	No	Yes: "Moods"	No
Place to listen to music	No	Yes: "Focus", "Travel", "Dinner", "Sleep", "Workout"	No
Chart position	Yes "iTunes Chart"	Yes "Charts"	No
Familiarity with song	Yes "Play Count"	Partial "Your Music"	Yes: "Play counter"
Mainstream or alternative	No	Partial "Genres"	No
Popularity of music	Yes "iTunes Chart"	Yes "Charts", "Plays" and "Trending"	Yes "Pupularimeter"

This table illustrates that of the 78 constructs, only 14 are utilised for organising or finding music. Spotify, has recently started to support more subjective ways of browsing music such as "Emotion" and "Place to listen to music" via their own curated playlists in the "Genres & Moods" area of their desktop application.

From the commonly used constructs shown in Table 3, "Gender of Artist", "Solo or group", "Main instrument", "Audience" and "Nationality of artist" are not currently supported by either iTunes, Spotify or ID3 tags.

The majority of the remaining unsupported constructs are highly subjective e.g. "Music to work to", "Times to listen to" suggests that automation of these parameters may be unrealistic and it would therefore be more practical to include functionality within the software that allowed users the freedom to sort and define songs with relevant constructs and attributes. Playlists provide this functionality to a certain extent but specific ID3 tags for certain attributes that could be saved within the file (as opposed to within the software such as Playlists)

would be preferred due to the potential for standardising and sharing this information, particularly for objective attributes such as "Gender of Artist" and "Main instrument".

6.2 Automation of constructs/categorization

With large collections of music, manual curation of songs and playlists is non-trivial and automated methods of categorisation are appealing.

Over 50% of the respondents from this study used "Speed of song" as a criterion, suggesting that tempo is a widespread construct in perception of music and would be a suitable attribute to include in automated music retrieval and classification systems. Tempo is one of the basic attributes of music and has been used previously as a parameter for automatic information retrieval (Scheirer, 1998). The current support in music players relies on the artist/user manually including/editing the relevant "BPM (beats per minute)" ID3 frame or using software such as Media Center 9 to try and automatically detect the BPM (anecdotal evidence however suggests that this method is inaccurate). Automated rhythm and frequency methods have been previously used to identify the tempo and beat of a song and these methods have also been compared to people's perceptions of tempo (e.g. Lippens et al., 2004).

Respondents from this experiment used the following criteria related to "speed of song": Song speed, Speed, Slow and fast, BPS, Pace, Type of dance, Tempo, Music speed, Fast paced and Beat speed. The co-occurrence matrix for sorts related to "speed of song" indicated that there was a high level of disagreement between respondents even though they used similar category names e.g. "slow", "fast", "quick" etc. This indicates that even if an automated method for determining tempo is used, users themselves are using different attributes or measures for what constitutes the speed of a song.

6.3 Folksonomies and tagging systems

There has been increased interest in the idea of using folksonomies and tagging as a way of categorising and exploring music. A number of studies into the use of tagging and the related field of social bookmarking (Kipp, 2007; Kipp & Campbell, 2006; Golder & Huberman, 2005) suggest that tagging and bookmarking share similar features to more traditional indexing systems (Kipp & Campbell, 2006) but also contain extra dimensions such as tags related to time e.g. "toread" and task or users' emotional responses to a document e.g. "cool", which conventional indexing systems and approaches do not support (Kipp & Campbell, 2006).

Last.fm is a website that builds profiles of musical listening habits and also allows users to tag songs

and artists with descriptive words or phrases. Comparing the constructs used by the respondents in this study with the tags generated by the users of last.fm highlights some interesting similarities and differences. The majority of tags used currently on last.fm (see Figure 19 in (de Quincey, 2010)) are genre based and are similar to the most common criteria used by respondents in this study e.g. “alternative”, “classis rock”, “electronic” etc.. There are also tags such as “female vocalists” and “male vocalists” that refer to the “gender of artist”, temporal tags e.g. “00s”, “80s” etc. that are similar to the “year music produced/released” and “favorites” “favourite”, “favourites” are similar to the “likeability of songs” construct.

It is interesting that “seen live” and “albums I own” are popular tags on last.fm but were not commonly used by respondents in this study (no respondents used “seen live” as a criterion and only one used a criterion related to “ownership”). Another point of interest is that the constructs “speed of song” and “solo or group” do not have equivalent last.fm tags.

6.4 Implications for Faceted Navigation

Amazon and Google make use of faceted navigation structures to allow users to further filter search results. The use of these structures can be linked to Facet Theory, originally devised as an improved way for categorising and indexing books by Ranganathan (Ranganathan, 1962). Generally, in these systems, once a user has performed a standard keyword search, as well as seeing the list of returned results, they are also given the option of searching/filtering within those results by various facets. This approach is often called “guided navigation” and although the term facet is not explicitly used, it is clear that providing users with options to search by format e.g. video, academic resources, images etc. or by geographical location or by time e.g. latest, past 24 hours, past year etc. the resources or products are being categorised by various facets.

Ranganathan’s approach to facets, deriving them systematically using Canons, Postulates, and Principles, meant that several high-level attributes or facets could be used to describe any entity (in his work, the entities were books). Ranganathan’s five initial facets were “Personality”, “Matter”, “Energy”, “Space” and “Time” but it is apparent that now, these terms, although appropriate for Ranganathan and librarians of the time, are not useful for all users of books. Automated methods for identifying facets are now being investigated (Ben-Yitzhak et al., 2008).

When comparing the results of this study with the faceted navigation structure utilised by Amazon for the “Music” section of its website, it can be seen that there are some similarities and significant differences.

The number of similarities demonstrates that Repeated single-criterion sorts are a potential method for eliciting these types of navigation structure. The search filters that cannot be mapped onto a specific construct can be explained by the choice of entities used in this study. The card sorting tool only provided respondents with the mp3 of the song itself. No information was given regarding the artist, edition, or format, so it is unlikely that these could have been used as criteria. Therefore, choice of the representation for a song needs further consideration and a mix of media may be more appropriate e.g. having the album cover representing the song instead of a number or a screenshot of the product description webpage (which could then also incorporate price and delivery options).

More noteworthy is that there are a number of constructs elicited during this study that are not part of the search filters on the Amazon website, specifically “Gender of Artist”, “Speed of song”, “Solo or group”, “Main instrument”, “Audience”, “Emotion” and “Nationality of artist”. These are all potential methods of guiding a user through a set of search results and apart from “Emotion” are all objective characteristics of the entity. This is similar to the findings of Cassidy (Cassidy et al., 2013) who used “All-in-one” sorting to determine how children categorise games. When compared with existing categories in the Google Play Store, they found that “children used categorization criteria much more aligned to the goals of the game rather than more abstract categories currently found in mobile phone application stores” (Cassidy et al., 2013).

6.5 Implications for the UCD process

“All-in-one” card sorting is already used in the UCD process to determine the IA of websites. However, this is commonly used to determine single level hierarchies where an entity fits into one specific top-level category e.g. Books, Music, Games, Films.

Repeated single-criterion sorting is a complimentary method for eliciting faceted navigation structures for when a user has chosen a top-level category, such as Books, and is now looking for a particular item using relevant criteria such as Author, Publication Date, Genre etc. From this study it is clear that this method can elicit traditional criteria which are akin to those originally proposed by Ranganathan but can also provide additional, user-centered dimensions such as those seen in user generated tagging systems.

A combination of the two sorting variations would provide a methodology for developing IA’s that avoid the limitations of hierarchical structures where products/pages may fit into multiple sub-categories. Closed sorting could also be used with both sorting variants to evaluate how well the category and criteria labels work.

Repeated single-criterion sorting could also be used in the initial stages of user experience design to analyse the target users' perceptions of web sites in the particular field. For example, if a developer was creating a site for a theatre company, the tool could be populated with images of the homepages of other theatre company websites and a study into users' perceptions of those sites could be undertaken. The results from this study could then be fed directly into the development process, with attention then being directed at the attributes of the pages that were elicited.

6.6 Challenges and Limitations

The main challenge with repeated single-criterion sorting is the same as with any other user-centered activity; recruiting respondents. For this study, a pool of willing participants was fortunately available, but this in itself causes potential issues with bias and representativeness. It should also be noted that there is limited published evidence on how many respondents are needed for this type of sorting to create effective IA's and this is where future work is needed i.e. to take the results from a study such as this, build a website with faceted navigation and then evaluate it with users.

7. CONCLUSIONS

The results of this study have demonstrated that card sorts are an effective method for investigating people's perceptions of music. Repeated single-criterion music sorts have elicited criteria which have previously eluded music psychologists and digital music service designers, while providing explanations to inconsistent musical genre classifications seen in previous studies.

Current support for digital media organisation and discovery relies heavily on genre across a range of different product types e.g. Music, Books and Films. Although genre was used by the majority of respondents, there was little agreement between respondents in which songs should appear in the same genre. This study has shown that perception of music is highly subjective, and genre, although considered to be objective by music retailers, is no longer adequate. The results indicate that as the volume, and variety of music increases, categorising music becomes more difficult. Repeated single-criterion sorting provides a method to support the systematic elicitation of additional objective and subjective features for use in the design of digital music services. This variation of card sorting has previously been reliant on a number of time-consuming processes (recruiting respondents, conducting synchronous sessions etc.) and although low cost and high yield, has not been frequently used as a standard part of design and development

methodologies. The key contribution of this work therefore is a demonstration that this card sorting variation and tool described can support developers and designers at various stages in the development life cycle to determine IA's for websites and applications that go beyond traditional objective features. This study has also shown how this method could be used as part of the UCD process in parallel with the more common "All-in-one" version to create faceted navigation structures.

9. ACKNOWLEDGEMENTS

Thanks to Dr Gordon Rugg for his supervision during the data collection stages of this study.

10. REFERENCES

- Aucouturier, J.J. & Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research*, 32,1: 83-93.
- Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., Sheinwald, D., Shekita, E., Sznajder, B. & Yogev, S. (2008). Beyond basic faceted search. In *Proceedings of the international conference on Web search and web data mining*, 33- 44.
- Bevan, N. (2003). *UsabilityNet methods for user centred design. Human-Computer Interaction: theory and Practice 1*. APA.
- Cassidy, B., Saurabh Antani, S. & Read, J. (2013). Using an open card sort with children to categorize games in a mobile phone application store. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2287-2290.
- Courage, C. & Baxter, K. (2005). *Understanding your users: a practical guide to user requirements: methods, tools, and techniques*. San Francisco, CA: Morgan Kaufmann Publishers.
- de Quincey, E. (2010). *Software support for comparison of media across domains*. Ph.D. Dissertation, Keele University, UK.
- Deibel, K., Anderson, R. & Anderson, R. (2005). Using edit distance to analyze card sorts. *Expert Systems*, 22,3: 129-138.
- DeNora, T. (2000). *Music in everyday life*. Cambridge: Cambridge University Press.
- Frederickson-Mele, K. (1997). Usability Testing an Intranet Prototype Shell: A Case Study. In *Proceedings of SIGCHI 1997*.
- Gammack, J.G. (1987). Different Techniques and Different Aspects of Declarative Knowledge. In *Knowledge Acquisition for Expert Systems*. New York: Plenum Press, 137-163.

- Gerrard, S. & Dickinson, J. (2005). Women's working wardrobes: a study using card sorts. *Expert Systems*, 22,3: 108-114.
- Golder, S. & Huberman, B. (2005). *The structure of collaborative tagging systems*. HP Labs Technical Report.
- Greasley, A. & Lamont, A. (2006). Music preference in adulthood: Why do we like the music we do. In *Proceedings of the 9th International Conference on Music Perception and Cognition*.
- Hudson, W. (2014). Card sorting for navigation design. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 1007-1008.
- Joyce, J. (2006). Pandora and the Music Genome Project. *Scientific Computing*, 23,10: 40-1.
- Kipp, M. (2007). @toread and cool: Tagging for time, task and emotion. In *Proceedings Information Architecture Summit 2007*.
- Kipp, M. & Campbell, D.G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In *Proceedings of the American Society for Information Science and Technology*, 43,1.
- Lippens, S., Martens, J-P., & De Mulder, T. (2004). A comparison of human and automatic musical genre classification. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP'04)*, 4.
- Maiden, N. (2009). Card Sorts to Acquire Requirements. *IEEE Software*, 26,3: 85-86.
- Martine, G. & Rugg, G. (2005). That site looks 88.46% familiar: quantifying similarity of Web page design. *Expert Systems*, 22,3: 115-120.
- McKay, C. (2010). *Automatic music classification with jMIR*. Ph.D. Dissertation, McGill University, Canada.
- Nakhimovsky, Y., Schusteritsch, R. & Rodden, K. (2006). Scaling the card sort method to over 500 items: restructuring the Google AdWords help center. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, New York, NY, USA, 183-188.
- Nielsen, J. (1992). The usability engineering life cycle. *Computer*, 25,3: 12-22.
- Nielsen, J. & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People (CHI '90)*. ACM, New York, NY, USA, 249-256.
- North, A.C., Hargreaves, D.J. & O'Neill, S.A. (2000). The importance of music to adolescents. *British Journal of Educational Psychology*, 70,2: 255-272.
- O'Neill, D. (2013). Retrieved September 16, 2016, from <http://id3.org/>
- Pachet, F. & Cazaly, D. (2000). A taxonomy of musical genres. In *Proceedings of Content-Based Multimedia Information Access (RIAO)*, 1238-1245.
- Ranganathan, S.R. (1962). *Elements of library classification*. Bombay: Asia Publishing House.
- Rugg, G. Corbridge, C., Major, N. & Burton, A. (1992). A comparison of sorting techniques in knowledge acquisition. *Knowledge Acquisition*, 4,3: 279-291.
- Rugg, G. & McGeorge, P. (1997). The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14,2: 80-93.
- Sanders, K., Fincher, S. Bouvier, D., Lewandowski, G., Morrison, B., Murphy, L., Petre, M. et al. (2005). A multi-institutional, multinational study of programming concepts using card sort data. *Expert Systems*, 22,3: 121-128.
- Scheirer, E. (1998). Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103,1: 588-601.
- Sloboda, J., O'Neill, S.A. & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory using the experience sampling method. *Musicae Scientiae*, 5,1: 9- 32.
- Spencer, D. & Warfel, T. (2009). Card Sorting: A Definitive Guide. Retrieved January 18, 2016, from http://www.boxesandarrows.com/view/card_sorting_a_definitive_guide.
- Tullis, T. (1985). Designing a menu-based interface to an operating system. *ACM SIGCHI Bulletin*, 16,4: 79-84.
- Tullis, T. (2003). Using Card-sorting Techniques to Organize Your Intranet. *Intranet Journal of Strategy and Management*, March 2003.
- Tullis, T. & Wood, L. (2004). How Many Users Are Enough for a Card-Sorting Study? In *Proceedings of the UPA 2004 Conference*, Minneapolis, MN, USA. Jun-2004.
- Upchurch, L., Rugg, G. & Kitchenham, B. (2001). Using card sorts to elicit web page quality attributes. *IEEE Software*, 18,4: 84-9.
- Usability.gov. (N.D.). *User-Centered Design Process Map*. Retrieved September 16, 2016, from <http://www.usability.gov/how-to-and-tools/resources/ucd-map.html>.
- Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In *Usability Inspection Methods*, J. Nielsen and R. L. Mack, eds. New York, NY: John Wiley & Sons, 105-140