Was Parsons right? An experiment in usability of music representations for melody-based music retrieval

Alexandra L. Uitdenbogerd and Yaw Wah Yap
Department of Computer Science, RMIT University
GPO Box 2476V, Melbourne 3001, Australia
+613 9925 4115
alu@cs.rmit.edu.au*

Abstract

In 1975 Parsons developed his dictionary of musical themes based on a simple contour representation. The motivation was that people with little training in music would be able to identify pieces of music. We decided to test whether people of various levels of musical skill could indeed make use of a text representation to describe a simple melody query. The results indicate that the task is beyond those who are unmusical, and that a scale numeric representation is easier than a contour one for those of moderate musical skill. Further, a common error when using the scale representation still yields a more accurate contour representation than if a user is asked to enter a contour query. We observed an average query length of about seven symbols for the retrieval task.

1 Introduction

Despite the recent increase in interest in music information retrieval we still know very little about the usability of such systems. For sung queries, we can be certain that the contour is fairly accurate, despite the intervals between notes being enlarged or reduced, and key drift that may occur (Lindsay, 1996; McNab *et al.*, 1996). However, many of the systems currently available in libraries or on the Internet use input in the form of a text string.

In 1975 Parsons published his "The Directory of Tunes" (Parsons, 1975) — an extension of Barlow and Morganstern's earlier "A Dictionary of Musical Themes" (Barlow and Morganstern, 1948). Parsons's aim was to produce a solution to locating music by theme for non-musicians. Each melody in the volume was represented by a contour string consisting of symbols: "U"representing up, "D" representing down, and "R" representing repeat or same note. For example, the theme from Beethoven's Fifth Symphony would be encoded as "*RRDURRD", with the asterisk representing the first note. The user merely had to work out the contour of the theme they wished to search for and could then look it up in the dictionary.

The question we wished to address was "Can novice musicians formulate a melody query in a text format?". We conducted experiments using a prototype melody-based music retrieval system, requiring users to enter a simple query using one of three text representations of queries. We found that the task was virtually impossible for music novices. Those with some music training had some success, and only the most musically skilled users were able to create queries reliably. In this paper we first discuss related work in music psychology, usability engineering methodology, and music retrieval user interfaces. We then describe the music representations used in our experiments in the context of music representations in general. This is followed by a report on our experiments.

2 Related Work

When it comes to musical perception, psychologists have clearly identified different classes of people: unskilled, musically skilled, and those with a professional level of music skill (Francès, 1958). Unskilled or beginner musicians have a holistic perception of music, whereas those with musical skill use cognitive processes in listening to music. Highly skilled musicians use both forms of perception.

For users of music retrieval systems several different behaviours and skills are required, including the ability to recall a melody, reproduce it in some way — for example, singing it — and also compare musical examples with the piece they are searching for. There is not a lot published that examines the capabilites of users with respect to the these tasks. We summarise the work that we know about below.

There have been two sets of experiments that examine the behaviour of listeners and singers. Their aim was to discover the types of errors people make when singing a given melody. McNab *et al.* (1996) studied people's recall of familiar melodies. Large intervals were reduced in size and small intervals were increased. Errors were greater with unusual changes in tonality. Accuracy depended more on singing experience than musical training. Some songs were commenced at the chorus. Lindsay (1996) examined how well people sang short unknown melodies. His study agrees with McNab's regarding interval size. He also notes that intervals were more accurate than the absolute pitch, and that singers didn't increase their error as they

*Contact author for all enquiries

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. ©2003 Johns Hopkins University.

progressed but that variance remained fairly constant throughout.

Uitdenbogerd *et al.* (????) discovered that it was a very difficult task for users to compare unknown pieces of music, however if the piece being looked for was known, then a rapid identification of whether an answer was relevant or not could be made.

More recent work on aspects of usability of music retrieval systems includes Blandford and Stelmaszewska's analysis of several on-line systems (Blandford and Stelmaszewska, 2002). They concluded that there were many problems regarding usability, particularly the dissonance between a user's representation of their information need and that of the system.

The experiments reported in our paper are partly based on the methodology of usability engineering. Part of the process of usability engineering involves measuring aspects of a user's interaction with a system. The five main measurable attributes are *learnability*, *efficiency*, *memorability*, *error rate*, and *satisfaction* (for a brief introduction to usability, see Ferré and N. Juristo (2001)). These attributes are mostly quantified by the number of tasks performed by the user per time unit, and the number of errors per task or per time unit. For example, learnability is measured by recording how long it takes for a user to reach a defined level of proficiency. Satisfaction is usually measured by surveying users.

3 Music Representation

There are many ways in which music can be represented. Our focus in this work is representing a melody only. In terms of pitch a melody can be represented using absolute pitch, pitch relative to a reference note, or relative to the previous note in the melody. For rhythm, the same types of representation apply, that is, absolute durations, relative to a standard duration, or relative to the previous duration. In addition to the pitch and rhythm *relativity*, another factor that forms part of its representation is its *precision*. For example, a contour representation has very low precision, consisting of only three different symbols, whereas a representation that defines each pitch in terms of the exact number of semitones has high precision. Some representations also include a third element, that of *stress*. Again this can be represented in absolute or relative terms, and can have different levels of precision. In the work reported here we restricted ourselves to representations of low to medium precision in pitch, and ignored rhythm and stress. The three representations used were contour, an extended contour, and a simple numeric representation using the numbers from one to eight to signify the eight notes of the major scale. This last representation was incompletely defined, in that users were not told how to write queries that extended beyond an octave or how to define notes outside of the scale, such as flattened notes. We now describe the representations in more detail.

Contour

A melody's contour represents a melody's shape in terms of pitch direction only. The contour representation of the melody fragment shown in Figure 1, using U for up, D for down and S (or sometimes R is used) for same pitch, is:

SUSUSDDSDSDSD

Contour representation has the advantage that singers usually get the contour of a melody right but usually don't sing the intervals accurately, an important consideration when queries are sung. A melody query would need to be quite long for relevant answers to be found, however. In particular, two melodies can be represented by the same contour string yet have no perceived similarity. For example, the first phrase of Twinkle Twinkle Little Star has the same contour (and rhythm) as that of the second movement of Haydn's Surprise Symphony, yet the melodies are quite different. What hasn't been reported before, however, is whether people can write contour queries.

Extended Contour

Extended contour is a more fine-grained approach to melody representation. In this version of extended contour, we distinguish between large and small intervals with a different symbol. For example, we could use U for a large interval upwards, u for a small one, and similarly use D and d for large and small downward intervals respectively. A decision needs to be made regarding the classification of intervals as large and small. It is clear that step intervals of one or two semitones are small and intervals that are greater than five semitones are large. If we use the musical concept of steps and leaps, then all intervals of three or more semitones would be classed as large intervals. If we used the entropy-maximising classification approach of Downie (1998), then the same decision would be reached, as intervals of one or two semitones are the most frequently occurring intervals in melodies. In this case, the example melody would be represented as:

SUSuSddSdSdSd

The advantage of this technique is that it still allows for inaccurate singing, but would be more discriminating when searching a database of melodies. When considering pitch errors of singers, however, there may still be some mis-classification of intervals as singers often have an error of up to one semitone (Lindsay, 1996).

Numeric Scale

Unlike the two contour-based representations described above, the numeric scale representation is relative to a key note. The key note is represented by the number one, and the notes of the major scale are numbered contiguously from one to eight. This is a method of describing the notes of a scale that is often used when learning western music. It is also the basis of a music notation form that is frequently used by choirs from several Asian countries, and is closely related to the *sol-fa* system.

The example melody would be represented as:

1 1 5 5 6 6 5 4 4 3 3 2 2 1

These representations have been used in some form for melody matching systems, either directly (Blackburn and Roure, 1998; Parsons, 1975), or indirectly (Ghias *et al.*, 1995; Prechelt and Typke, 2001). However, little is known about users' ability to use such representations.

4 Experiments

Our aim was to determine the ability of a range of users in formulating a melody query for a music search engine. We wanted

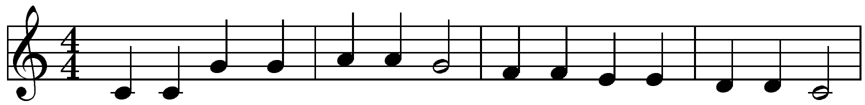

Figure 1: *The first two phrases of Twinkle Twinkle Little Star in Common Music Notation (CMN)*

to learn more about the usability of music search engines by focusing on the concepts of learnability, user error rate and satisfaction (Ferré and N. Juristo, 2001). Due to limited availability of volunteers we chose not to measure how memorable the user interface was for repeated use.

We developed a prototype search engine that made use of the collection of 10,466 MIDI files used in our earlier work (Uitdenbogerd, 2002; Uitdenbogerd *et al.*, ????; Uitdenbogerd and Zobel, 1999, 2002) plus four simple monophonic melodies that were used as targets for queries. We used the directed modulo-12 melody representation and 5-gram coordinate matching as the melody similarity measurement technique (Uitdenbogerd and Zobel, 2002). A simple web-based front-end was developed, with instructions to the user including examples of each of the three query representations. Users were asked to formulate a query for Twinkle Twinkle Little Star and enter this into the search engine. A query was considered a success if the song appeared in the ranked list of answers presented to the user.

There were 36 participants in the experiment, most of whom were members of the university community. Significant cohorts within the group were members of the university choir club, and computer science research students. Participants were asked to self-assess whether they were non-musicians, novices, intermediate or professional level of music skill. The number of participants in each category is shown in Table 1. Participants also provided other general information about themselves including their computer skill level. Some questions were also asked about the users' perception of the system, including overall satisfaction. During the experiment a further measure of users' satisfaction or frustration was taken by observing their facial expressions during the task.

If users didn't succeed in retrieving the target answer, they were permitted to create a second query. To allow feedback from users regarding the perceived difficulty of the different representations users were asked to try each representation for the same task. However, as participation was voluntary, we only have a small number of results for users trying all representations.

Results

Most results are shown in tables 1 to 6 and some are described below. We had a reasonable spread of abilities across the set of users. Users generally felt satisfied with the system, with the most skilled users being more satisfied than those with no musical skill (Table 6). They were generally happy with the screen navigation (Table 5) but some non-expert users found the system hard to learn (Table 4).

Users who tried more than one representation were fairly consistent in their preferences. Thirteen users tried contour representation first and then later used scale. Ninety-two percent of these users preferred scale. Similarly, of the twelve users who tried scale representation before contour, 75% preferred scale. The only group of users that preferred contour to scale were novice users, but the results are too close to be significant given the number of participants in the survey.

Users made a few different types of error when entering queries into this system. The most common type of error was to use the wrong representation type. For example, after choosing to use extended contour the user might enter a contour string as a query by mistake.

Different types of transcription error occurred for the different representation types. With contour, about 41% of queries were entered with an extra symbol at the start. Typically these queries looked like SSUSUSD instead of SUSUSD. This is due to the need for users to enter a symbol for the first note, despite the contour method representing the transitions between notes. Parsons attempted to address this problem by requiring users to first write down an asterisk to represent the first note. This would probably improve written queries to some extent. Despite the extra symbol at the start, the music search engine retrieves the answer, as the matching method counts the number of n-grams that are identical between the query and potential answers but doesn't penalise for those that do not occur.

We discovered that 31% of first *scale* queries were incorrect scale representations, but were correct when converted to contour representations. For example, instead of entering (1 1 5 5 6 6 5), some users entered (1 1 2 2 3 3 2), or (1 1 3 3 5 5 3), which still converts to SUSUSD.

Remarkably, all 36 query attempts by users with no musical knowledge failed — a clear result despite the sample size of 13 people. All groups except for experts expressed frustration when attempting the task and failed more often than they succeeded. Despite failing consistently, the musically unskilled group showed considerable interest in music, with over 60% of these users stating that they listened to music more than five times a week. However, we can see from the results that querying music search engines via a text representation is clearly something for experts only.

We observed that most users entered a query representing the first phrase of the song only, and in some cases an incomplete phrase. The average query length was just under seven symbols, and this was fairly consistent across different user groups. This query length is considerably shorter than that found in Uitdenbogerd *et al.* (????), which was based on an expert musician formulating queries for specific pieces without the use of a music retrieval system. This tends to suggest that real queries are likely to be quite short.

5 Conclusions

Music search engines vary in their user interface and the type of query that users are required to enter. Many current systems

Music Background	Number of Users	Percentage (%)
Zero	13	36
Novice	10	28
Intermediate	8	22
Professional	5	14
Total	36	100

Table 1: Distribution of Users

Music Background	Contour		Extended Contour		Numeric Scale	
Zero	0/19	(0.00%)	0/3	(0.00%)	0/14	(0.00%)
Novice	1/16	(6.25%)	1/6	(16.67%)	3/12	(25.00%)
Intermediate	1/11	(9.09%)	1/3	(33.33%)	5/12	(41.67%)
Professional	4/07	(57.14%)	4/6	(66.67%)	4/06	(66.67%)

Table 2: Proportion of successes by type of users. The fi rst number in each cell represents the number of successful queries during the experiment for that type of user, the second is the total number of trials. This is followed by the proportion as a percentage.

Music Background	Excellent	Good	Moderate	Hard to Use
Zero	1 (07.69%)	7 (53.85%)	4 (30.77%)	1 (07.69%)
Novice	0 (00.00%)	7 (70.00%)	3 (30.00%)	0 (00.00%)
Intermediate	0 (00.00%)	6 (75.00%)	1 (12.50%)	1 (12.50%)
Professional	2 (40.00%)	3 (60.00%)	0 (00.00%)	0 (00.00%)
Total	3 (08.33%)	23 (63.89%)	8 (22.22%)	2 (05.56%)

Table 3: Result of the user survey question on “user friendliness”. The fi rst number in each entry represents the number of users, and the second is the percentage of that type of user giving the response.

Music Background	Easy	Moderate	Hard	Very Hard
Zero	4 (30.77%)	4 (30.77%)	4 (30.77%)	1 (7.69%)
Novice	2 (20.00%)	5 (50.00%)	3 (30.00%)	0 (0.00%)
Intermediate	1 (12.50%)	6 (75.00%)	1 (12.50%)	0 (0.00%)
Professional	2 (40.00%)	3 (60.00%)	0 (00.00%)	0 (0.00%)
Total	9 (25.00%)	18 (50.00%)	8 (22.22%)	1 (2.78%)

Table 4: Result of the user survey question on “learnability”.

Music Background	Excellent	Good	Confusing	Complicated
Zero	2 (15.38%)	11 (84.62%)	0 (00.00%)	0 (0.00%)
Novice	3 (30.00%)	7 (70.00%)	0 (00.00%)	0 (0.00%)
Intermediate	0 (00.00%)	7 (87.50%)	1 (12.50%)	0 (0.00%)
Professional	5 (100.00%)	0 (00.00%)	0 (00.00%)	0 (0.00%)
Total	10 (27.78%)	25 (69.44%)	1 (02.78%)	0 (0.00%)

Table 5: Result of the user survey question on “screen navigation”.

Music Background	Very Satisfi ed	Satisfi ed	Moderate	Poor	Very Poor
Zero	1 (07.69%)	4 (30.77%)	6 (46.15%)	2 (15.38%)	0 (0.00%)
Novice	1 (10.00%)	5 (50.00%)	4 (40.00%)	0 (00.00%)	0 (0.00%)
Intermediate	1 (12.50%)	5 (62.50%)	1 (12.50%)	1 (12.50%)	0 (0.00%)
Professional	2 (40.00%)	3 (60.00%)	0 (00.00%)	0 (00.00%)	0 (0.00%)
Total	5 (13.89%)	17 (47.22%)	11 (30.56%)	3 (08.33%)	0 (0.00%)

Table 6: Result of the user survey question on “overall satisfaction”.

on the Internet use a text representation of some kind. Similarly, those few music library systems that contain a searchable melodic representation also use a textual form. Our experiments show that these types of interfaces are beyond the average person, and require a musical expert to prepare queries. In addition we discovered that users are unable to use a simple contour representation despite this method — at least theoretically — not requiring musical knowledge to construct. It seems that the cognitive processes that are required must be developed by musical training or the user must possess considerable natural ability.

Contour representaion, while not useful as a direct method of querying a system, may still be used indirectly for retrieval. This practice already occurs for query-by-humming systems, but can also apply to other types of user interface. If a query using a numeric representation is unsuccessful, the contour of the query can be used to retrieve some results. However, as seen elsewhere (Uitdenbogerd and Zobel, 1999), such queries will need to be longer to retrieve relevant results.

In order for an ordinary user to retrieve from a music search engine based on melody, they will need to either sing, or use some other form of audio input. And what of Parsons? While an interesting concept, and a curiosity for music library users, unfortunately the directory is beyond the capabilities of its target audience.

Acknowledgements

We thank the volunteers that helped with our experiments.

References

Barlow and Morganstern (1948). *A dictionary of Musical Themes*.

Blackburn, S. and Roure, D. D. (1998). A tool for content-based navigation of music. In *Proc. ACM International Multimedia Conference*. ACM.

Blandford, A. and Stelmaszewska, H. (2002). Usability of musical digital libraries: a multimodal analysis. In M. Fingerhut, editor, *International Symposium on Music Information Retrieval*, volume 3, Paris, France.

Downie, J. S. (1998). Informetrics and music information retrieval: an informetric examination of a folksong database. In *Proceedings of the Canadian Association for Information Science, 1998 Annual Conference*, Ottawa, Ontario. CAIS.

Ferré, X. and N. Juristo, H. Windl, L. C. (2001). Usability basics for software developers. *IEEE Software*, **18**(1), 22–29.

Francès, R. (1958). *La Perception de la Musique*. L. Erlbaum, Hillsdale, New Jersey. Translated by W. J. Dowling (1988).

Ghias, A., Logan, J., Chamberlin, D., and Smith, B. (1995). Query by humming — musical information retrieval in an audio database. In *Proc. ACM International Multimedia Conference*.

Lindsay, A. T. (1996). *Using contour as a mid-level representation of melody*. Master's thesis, MIT, Massachusetts.

McNab, R. J., Smith, L. A., Witten, I. H., Henderson, C. L., and Cunningham, S. J. (1996). Towards the digital music library: Tune retrieval from acoustic input. In *Proc. ACM Digital Libraries*.

Parsons, D. (1975). *The Directory of Tunes*. Spencer Brown and Co., Cambridge, England.

Prechelt, L. and Typke, R. (2001). An interface for melody input. *ACM Transactions on Computer-Human Interaction*, **8**(2), 133–149.

Uitdenbogerd, A. L. (2002). *Music Information Retrieval Technology*. Ph.D. thesis, School of Computer Science and Information Technology, RMIT University, Melbourne, Victoria, Australia.

Uitdenbogerd, A. L. and Zobel, J. (1999). Melodic matching techniques for large music databases. In D. Bulterman, K. Jeffay, and H. J. Zhang, editors, *Proc. ACM International Multimedia Conference*, pages 57–66, Orlando Florida, USA. ACM, ACM Press.

Uitdenbogerd, A. L. and Zobel, J. (2002). Music ranking techniques evaluated. In M. Oudshoorn, editor, *Proc. Australasian Computer Science Conference*, Melbourne, Australia.

Uitdenbogerd, A. L., Chattaraj, A., and Zobel, J. (????). Music information retrieval: Past, present and future. In D. Byrd, J. S. Downie, and T. Crawford, editors, *Current Research in Music Information Retrieval: Searching Audio, MIDI, and Notation*. Kluwer. (originally presented at ISMIR 2000), to appear.