Usability of Musical Digital Libraries: a Multimodal Analysis

Ann Blandford
UCL Interaction Centre (UCLIC)
University College London
26 Bedford Way, London, WC1H 0AB, U.K.
+44 20 7679 7557
A.Blandford@ucl.ac.uk

Hanna Stelmaszewska
UCL Interaction Centre (UCLIC)
University College London
26 Bedford Way, London, WC1H 0AB, U.K.
+44 20 7679 7557
H.Stelmaszewska@ucl.ac.uk

ABSTRACT
There has been substantial research on technical aspects of musical digital libraries, but comparatively little on usability aspects. We have evaluated four web-accessible music libraries, focusing particularly on features that are particular to music libraries, such as music retrieval mechanisms. Although the original focus of the work was on how modalities are combined within the interactions with such libraries, that was not where the main difficulties were found. Libraries were generally well designed for use of different modalities. The main challenges identified relate to the details of melody matching and to simplifying the choices of file format. These issues are discussed in detail.

1. INTRODUCTION
As digital libraries become more widely available, particularly via the Internet, and as the capability of that network to transmit large volumes of data within reasonable times increases, so more collections of music are becoming available through web-based digital libraries. These music collections are stored in, and can be retrieved in, various formats, and can also be accessed by various mechanisms. To be genuinely useful, such collections need to be easily accessed by web users across the globe, with differing levels of musical and information retrieval expertise. There is very little literature on usability issues for such systems; one of the few examples we have found in an evaluation of a Digital Music Library (DML) project conducted by Indiana University [10]. The usability studies conducted for that project examined users' performance on pre-determined tasks and gathered their reactions to an early prototype of the DML interface; the study focused on general usability issues and user satisfaction, rather than issues that are particular to music libraries. In contrast, Cunningham [9] discusses how potential users of music information retrieval systems might be identified, and their needs ascertained. The work reported here takes a different approach: it investigates usability issues that pertain specifically to existing digital libraries containing music collections, including music stored in various representational formats. The primary focus is on usability issues relating to the modalities employed in the interaction between user and system. As others (e.g. [14]) have noted, there is a 'medium mismatch problem' when documents and queries are expressed in different media; our work investigates this problem from a usability perspective, considering both the use of different media and also mismatches within one medium.

The method employed in this work has been to apply a novel theory-based usability evaluation technique, Evaluating Multimodal Usability (EMU: [11]), to four different web-accessible music digital libraries. These four libraries have been chosen to provide a reasonably broad representation of the capabilities of existing libraries. They include different retrieval mechanisms, from the user simply typing the title of a target tune to the user entering a sound file that represents the target melody or entering a representation of the melody using a text-based tune contour notation (described below). They also include different media for the retrieved tune: musical score, "ABC" notation, lyrics, or various formats of sound file. In one case, the retrieved tunes are accompanied by video clips. These different retrieval mechanisms and media formats are described in more detail below.

1.1 Modality: A definition
The focus is on the use of different modalities within the interaction between user and digital library. There exist various definitions of a 'modality' (e.g. [1, 3, 4, 8, 17]). Here, we take the definition derived by Hyde [11] as a "temporally based instance of information perceived by a particular sensory channel". This definition encompasses three dimensions:

- Time, which may be discrete, continuous or dynamic;
- Information form, which may be lexical, symbolic or concrete; and
- Sensory channel, which may be acoustic, visual or haptic.

Here, 'discrete' means that the information is communicated over a very short period of time, as a discrete event, rather than repeating the same information over an extended period of time ('continuous') or varying the information content over time ('dynamic').

The information may be communicated in words, or other linguistic form ('lexical') or may be non-verbal, in which case it is classed as either 'symbolic' (meaning that the representation has some meaning to the user over and above being a picture, sound or sensation) or 'concrete' (simply 'being' – for example, the sight and sound of rain falling).

Information is communicated through sensory channels that correspond to people hearing ('acoustic'), seeing ('visual') and feeling ('haptic').

The various means of representing music can be classified according to this modality definition. For example, a traditional musical score would be classified as visual–symbolic–continuous, while the heard melody would be acoustic–concrete–dynamic and the lyrics (displayed on the screen) would be visual–lexical–continuous. Each main modality may also have dependent modalities: for example, the volume or dynamics of a melody may impart additional information, which would add additional acoustic–symbolic–modalities to the basic sound. In addition, modalities may be used together; for example, in music videos the auditory and visual modalities complement each other, and in some cases lyrics are displayed as the corresponding melody is played.

In interaction, user and computer system both express (output) modalities and receive (input) modalities; one source of difficulties will be incompatibilities between the two – for example, a user singing to a computer system that has no means of receiving auditory input.

The EMU evaluation technique involves defining user goals and tasks while working with a library and then stepping through the task identifying what input and output modalities are involved in each step, and checking for any mismatches or clashes. In practice, for any new system, this first involves an exploratory study in order to identify what user goals the system supports, and what the corresponding task structures are.

Hyde's [11] modality definition and her approach to assessing multimodal usability were used to structure the work reported here.

1.2 The collections

In order to investigate a range of the usability issues posed by web-accessible music collections, four collections that make use of different representational formats and retrieval mechanisms were selected for study. Two of the collections are available in the New Zealand Digital Library (NZDL) because we are collaborating with the developers at the University of Waikato [12, 2], so that usability evaluations are informing ongoing development work. Another two collections were identified as providing interesting contrasts to further explore multimodal usability issues. Each of these four collections is described below. These descriptions are based on the systems as they were in December 2001, and do not reflect any changes made more recently.

1.2.1 The NZDL Music Library

The NZDL Music Library[1] consists of several separate collections that have been developed at different times, and with different research objectives. All are based on the same retrieval mechanisms, of which the key features are as follows.

- Both search and browse facilities are available.
- Browsing is performed by tune title only.
- Searching can be performed by text entry of tune title, or by defining a melody file that contains the melody to be matched. Both 'and' and 'or' searches are supported.
- Any melody file has to be created 'outside' the NZDL system; a variety of sound formats are supported. Melody files are typically created by singing a tune clearly, using any available recording software. The melody files used in this study were created using an AIFF recorder on a Macintosh computer.
- When using melody files, the user can play back:
    - 'what you sang' (the user's original sound recording);
    - 'what I heard' (the sound file created by the melody indexing system, Meldex, after processing the input file, which is then compared against files in the NZDL music collection); and
    - any retrieved melodies.
- Tunes can be retrieved as melodies (played back using a browser plug-in such as QuickTime on the Macintosh) or as scores. In some collections, additional information about tunes can also be retrieved, but we do not consider that further here.

[1] accessed via http://www.nzdl.org/

1.2.2 JC's ABC Tunefinder

JC's ABC Tunefinder[2] supports retrieval of music as 'ABC' notation, standard musical scores and MIDI (sound) files. The site contains an index of files from over 200 separate online ABC collections, and the various output formats are generated automatically from the ABC files. Consequently, this site supports searching but not browsing. The user can specify a tune by text entry of either the title or a specification of the sound contour. The latter is generated by typing 'u' if a note is higher than the previous one, 'd' if it is lower, and 's' if it is the same (so, for example, the opening bars of 'Three Blind Mice' would be written as 'ddudd') [16].

The ABC notation is an established text-based notation that is both human and machine readable, and that expresses the main features of a tune as it might otherwise appear in standard music notation[3].

The site provides other facilities, such as a capability to convert ABC files into other formats, that we do not consider further here.

1.2.3 The Folk Music Collection

The Folk Music Collection[4] consists of early (prior to 1927) folk music. Users can search by text only; that text may be in the title, lyrics or other information about the song. Songs can also be browsed by various categories – for example, by country of origin or by type. The user can retrieve lyrics, together with any additional information about the song that has been made available, and can also download a midi file of the tune, which will play using the relevant browser plug-in. On most pages in this system, an appropriate tune is played as background music whenever the page is displayed.

1.2.4 The NZDL Music Videos Collection

The NZDL music videos collection[5] consists of short pop and rock video clips, with accompanying sound tracks. Tunes are retrieved by title or artist, and are played back as audiovisual clips using an appropriate browser plug-in, such as QuickTime on the Macintosh. The videos collection can also be browsed by title or artist. (In each category, items are ordered alphabetically; within artist, items by the same artist are grouped together).

2. AIMS AND METHOD

The aim of the work reported here is, as noted above, to develop an understanding of the usability issues that apply to collections of musical information, and particularly to retrieval of musical documents from a digital library. The four collections described above were selected as being representative of the current state of the art, in terms of what is publicly available for music retrieval.

For each collection, analysis proceeded as follows:

1) A familiarisation phase: the collection was studied informally, to identify key features and interaction possibilities, and to define representative tasks to be used for the EMU analysis.
2) For the representative tasks defined in step 1, a full EMU analysis was conducted.

[2] http://trillian.mit.edu/~jc/music/abc/FindTune.html

[3] for more details see http://www.gre.ac.uk/~c.walshaw/abc/

[4] http://www.contemplator.com/folk.html

[5] accessed via http://www.nzdl.org/

3) In addition to the EMU analysis, which focuses on mismatches and clashes, other usability difficulties, relating specifically to the type of data being retrieved and the retrieval mechanisms made available, were noted. Although these were not identified explicitly through the EMU notation, the use of EMU provided a structure for the analysis. More specifically, the need to understand the systems well enough to conduct the EMU analysis meant that thorough exploratory walkthroughs were conducted, through which a range of usability challenges were identified.

The results of these two kinds of analysis are summarised below. Extended EMU analyses for all collections are presented by Blandford and Stelmaszewska [5].

3. RESULTS

The detailed EMU analyses were conducted for canonical tasks – that is, tasks in which the user makes no errors and that are achievable. Since each collection supports different user goals, which correspondingly different task structures, a different task was used for each collection. The EMU analyses *per se* identified very few usability difficulties, the main one being a physical clash in the Folk Music collection, which we describe briefly here. As noted above, on many pages, the relevant tune is played as soon as the page has loaded in the user's browser window. However, the user is also presented with an option to download the tune. In some browsers, if the user does so, the tune will start playing a second time, 'over' the first version. Although the computer system is capable of transmitting two streams of audio data, the user is not able to separate them into two separate streams, and therefore hears a cacophony of sound. Whether or not this happens depends on which browser is being used.

The fact that this was the only substantive usability problem relating to modalities that was found, and that this situation is easily avoided by a user once it has been discovered, indicates that, in term of modalities employed in the interaction, all four collections analysed were usable. The discussion that follows considers broader issues involved in the specification, matching and presentation of musical information.

We structure the discussion of results by the kinds of activity involved in retrieval of a tune: browsing; text-based searching; tune matching; how retrieved scores, melodies and other information is presented to the user; and how the collections interface with other systems.

3.1 Browsing

The two NZDL collections and the Folk Music collection support full browsing, and all collections support browsing within a search results set. In the case of the NZDL Music Collection, browsing is by title only, which supports user familiarisation with the overall contents of the collection. In the case of the NZDL Music Videos Collection, browsing can be by title or artist. Within the Folk Music Collection, browsing is by genre.

All four collections support browsing within search results, but present results in different orders. In the case of the NZDL collections, the order is by calculated quality of the match, with the best match first. In the case of the Folk Music Collection, search results can be browsed in alphabetical order of file name (although filenames ending in '.html' are listed before ones ending in '.htm'), which does not guarantee that tunes are listed in an order that might seem natural to the user. JC's ABC Tunefinder also lists results alphabetically. In all cases except the music videos, the fact that tunes are collected from disparate sources means that there may be multiple copies of the same tune, with the same or slightly different titles.

For a music collection, an alternative browsing technique might be auditory. However, auditory browsing is difficult to implement effectively due to the dynamic nature of auditory information, which does not fit well with the user-controlled activity of skimming results quickly, which is much more easily achieved for continuous information. In the context of the Internet, where sound files are stored remotely and download times (even for small sound files) may be significant, implementing auditory browsing would be very difficult. Therefore, it is hardly surprising that browsing has been limited to test only in existing music libraries.

3.2 Text-based searching

Text-based searching by title matching posed only minor difficulties. In some collections, there is ambiguity about what text is being matched; for instance, in the NZDL Music Collection, the user is invited to enter 'text', but the only text that will be matched against is that from the titles of tunes. Conversely, in the Folk Collection, text is matched against any text in the file, which may be title, lyrics or other information. In most cases, the user is given insufficient information about what is being matched, and no opportunity to define the scope of the matching. This is likely to be a consequence of the ways these kinds of collections have been amassed, since most have a structure that was not designed for the current purpose. The exception is the NZDL Music Videos collection, where users can select to search by title, artist or both.

The NZDL Music Collection search engine can be set to 'case sensitive' or 'case insensitive', but defaults to 'case sensitive'. Since some collections have tunes indexed by titles in upper case and some in title case, it is difficult for the user to anticipate which case is appropriate; it is also not immediately clear to the user that matching is case sensitive. This is one example (more are discussed in the following section) of a situation where the matching algorithm may not return the expected or intended results.

Due to the way folk tunes are archived, with titles being passed on by word of mouth, there can be variations in the ways tune titles are spelt, resulting in erratic retrieval of tunes. This is reported (Chambers, personal communication) to cause difficulties sometimes, but we did not experience any particular problems with text based searching.

With these provisos, text-based searching is cognitively undemanding and generally successful. The user enters lexical items that can be easily matched against corresponding database items.

3.3 Tune matching

Tune matching presents much greater challenges to systems developers and users alike. While it offers great promise as a means of retrieving information (such as title and score) about tunes for which only the melody is known, there are substantial technical and usability difficulties to be overcome.

Only two of the four collections analysed support tune-based retrieval: the NZDL Music Collection and JC's ABC Tune Finder. We start by considering difficulties with the 'contour' feature of the ABC Tune Finder.

As discussed above, 'contour' search allows the user to enter letters (u, d and s) representing the direction in which the notes go (i.e. up, down and same). The contour does not allow the user to identify how much higher or lower each note is than the preceding one, or to represent note durations. If trying to retrieve a tune that is in their head, the user has to sing the tune to themselves, noting whether each note is higher, lower or the same as the previous one. For 'ABC novices' this is a cognitively demanding task as it involves converting from relative pitch (an acoustic, concrete, dynamic modality) to a textual representation on each (discrete) note-change in a dynamic melody (visual, lexical, discrete modality).

Furthermore, there are features of the ABC tune finder that might put a user off track. The first is that the system matches the user's entry against only the first 16 (or fewer) notes of the stored tune (i.e. 15 intervals); if the user enters more intervals that this then the system returns no matches, but with no explanation: the more precise the user tries to make the results set, the greater the chances of them getting no results at all. The second is that the system uses the ABC notation as an intermediate representation, so that heard notes that are a semi-tone apart may be represented as the 'same' (e.g. C and C#), and a single heard note that is internally represented as two or more slurred notes is also represented as a 'same' interval. A third difficulty is that notes that are not part of the main melody line may be included in the file. These phenomena are illustrated in Figures 1 and 2. Figure 1 shows the opening bars of "Yesterday", which includes a D-flat and D-natural in the second bar, which are on the same note of the stave: this interval sounds 'up' but has to be represented as 'same'. A second difficulty for individuals 'singing to themselves' is that the three notes at the end of the first bar and beginning of second are part of the accompaniment, and do not have lyrics attached, so someone singing the tune based on the lyrics would miss out these three notes (and the corresponding intervals), and hence the tune would not match. The contour of the song 'Yesterday' is represented as 'dssuduuusuudds', but the user might reasonably enter 'dsuuuuuudds' for the same musical phrase; the latter returns 'no matches'.

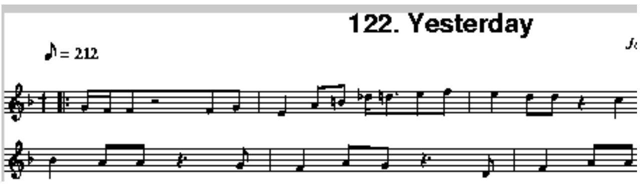

Figure 1: Extract from the score of "Yesterday"

The difficulty with slurred notes is illustrated by "Layla", as shown in Figure 2. This has a contour representation of 'usuddduuudduddu'. Here, notes (e.g. second and third of first bar) are slurred, so that the later of the two notes is not sounded separately. Again, the interval between these notes has to be represented as 'same' even though the second is not (audibly) a separate note from the first. In future, the developers are experimenting with removing 'same' from the matching, to assess whether more satisfactory results are returned.
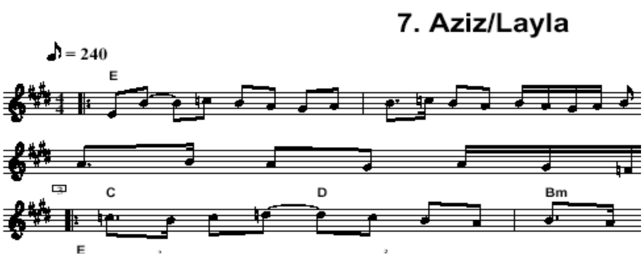
Figure 2: Extract from the score of "Layla"

For the user who already has the musical score and is simply trying to access the ABC notation or a MIDI file, these problems are surmountable, but the user who is 'playing by ear' is likely to experience tune retrieval as a very 'hit or miss' affair.

In summary, although conceptually simple, the ABC Tune Finding technique presents usability difficulties for many of the intended user population because the underlying data representation is precise, but does not necessarily match the user's internal representation of the tune accurately.

The NZDL Music Collection approach of 'singing to your computer' is, conceptually, even simpler than that of describing contours textually. However, it suffers from the same tension between the precision of the database matching and the accuracy (relative to the user's "song in the head"). Even though the matching is approximate [13], it can appear too precise for the user. Here, we consider only contour matching within the collection (where the system derives a contour from the sung melody for matching purposes).

Because the files in the database have been gathered in various ways, the quality is variable: the files in one collection ('Fake Book') were collected by applying optical musical recognition to printed music, while those in another ('MIDIMax') were gathered by trawling the web. Consequently, some melodies in the database have missing notes or the occasional inaccurate note, so that contour matching fails if the user sings the melody correctly. Preference settings (e.g. metronome beats) may also affect the way a melody is interpreted (the difference between 'what you sang' and 'what I heard') in ways that are difficult for non-expert musicians to anticipate. The user is given feedback, both visual and auditory, on how their input was processed, as shown in Figure 3. In principle this gives appropriate feedback, at least to the user who can read the standard music notation. One case where this can break down for new users occurs when they sing indistinctly, so that the interpretation mechanism fails to identify any notes, resulting in a display that shows just the musical stave with no notes showing. To the new user, this can look like decoration rather than feedback. The difficulty of poor database entries is illustrated in Figure 3 by the absence of 'AULD LANG SYNE' (non-Christmas version), which is in the database, but was not matched because of missing notes.
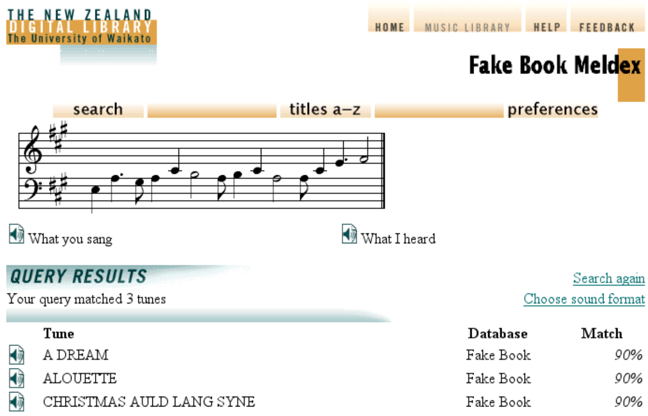
Figure 3: Search results, including feedback to user on 'what I heard'

In summary, in both contour-matching systems, there are user difficulties caused by differences between a user's mental representation of a tune, which will typically include approximate pitch intervals and note durations in the melody, and the precise but abstract representation of the melody in the database, which abstracts over pitch intervals (leaving just a contour) and durations (leaving just temporal sequence). Additional user difficulties lie in the means of translating the tune in the head into an external representation that can be entered into the computer system (whether that be an ABC contour or a sung melody) and then checking that the external representation is correct.

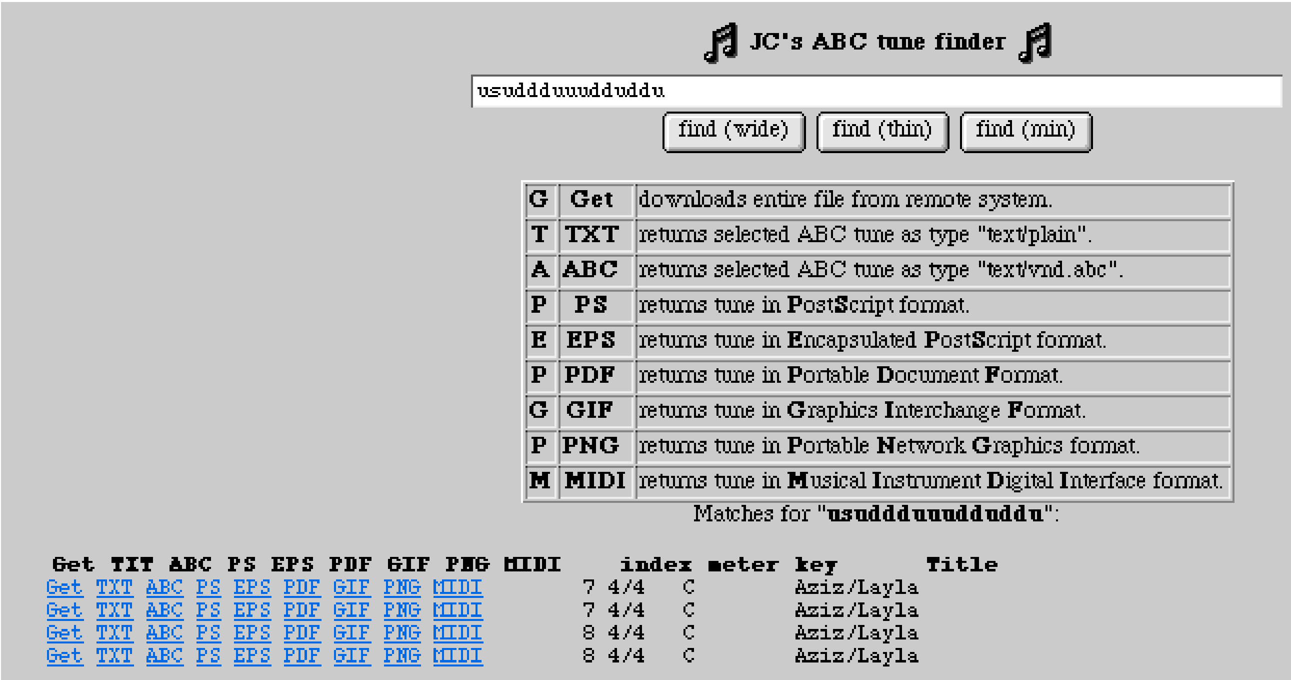

Figure 4: Search results for contour search, including the various output formats in JC's ABC Tune Finder

3.4 Representation of output possibilities

The developers of music collections are faced with a real challenge in determining appropriate output formats. There are both multiple modalities (the musical score, the sound file, and other textual representations such as ABC notation, lyrics or other information) and also multiple formats for each modality – typically, that have been developed for different platforms or different user populations.

The outputs available for JC's ABC Tune Finder are shown in Figure 4. In summary, there is one option to 'GET' a whole file (rather than just one tune representation), two textual options ('TXT' and "ABC') that return the ABC notation, five graphical options that return the musical score, in formats from PS to PNG (which return files that are visually very similar if the user has the necessary file readers for each format) and, finally, one sound option for playing the MIDI sound file (using a browser plug-in such as QuickTime). As shown in Figure 4, while the different file formats are clearly represented at the interface, the different modalities of the file contents are not, even though this is likely to be the most important file feature for a user. ABC, PS and EPS cause a file to be downloaded, to be further interpreted by suitable software on the user's computer, whereas the remaining formats appear without further user intervention in the browser window.

Figure 5: Sound output formats in NZDL Music Collection

Chambers (personal communication) argues that the difference between the various graphical formats is important: that PS is much higher quality than GIF or PNG, and that in folk circles lower quality is generally preferred. In terms of modalities, as well as the primary modality (visual–symbolic–continuous) through which the tune is represented, there is a dependent modality, also visual–symbolic–continuous, that communicates the 'folkiness' of the paper representation. Clearly, for some – arguably more sophisticated – users, this property of the visual representation is important, while for others it is not. However, this understanding is only available to people who are experts in both folk music culture and computer music representational formats; for others, additional explanation would be helpful, or the choice may be bemusing.

Whereas in JC's ABC Tune Finder, the largest choice of formats is made available for the musical score, within NZDL Music Collection, the greatest choice is of sound formats, as shown in Figure 5.

In practice, a different subset of these formats is likely to be available on any particular computer system. To most users, MIDI Types 0 and 1 are indistinguishable, and most first-time users of this kind of system are likely to have difficulty ascertaining which other sound formats are actually available to them. If the user chooses a non-available format, the system response is unpredictable; for instance, when we selected 'Ulaw' on one computer system, we waited 12 second while the file downloaded, at which point a 'broken' QuickTime icon was displayed, with no output sound; whereas when we selected Real Audio, an error message was immediately displayed stating that the application 'Real Player' was not available.

Other than the two MIDI formats, the different sound formats that are available on a particular computer system typically have audibly different qualities – notably of pitch and pace. The user who does not have a background in audio technology can only discover these differences through a process of trial and error. Also, 'what I sang' can only be played back in the sound format in which it was recorded, so giving the user a choice is liable to lead to user errors. While this choice of formats may be appropriate for experts, it would ideally be avoided for novices, by setting suitable defaults.

Within the NZDL Music Collection, the different types of output are represented by icons next to the tune names, as shown in Figure 6. In this example, there are three alternatives: to display the musical score (the icon with the treble clef), to hear a MIDI sound file (the speaker icon) or to see textual information about the tune. There is a notational difference between MIDI files (which can only be played back in MIDI format) and other sound files (see for example the speaker icon next to 'what you sang'), which users are expected to understand.

Figure 6: NZDL Music Collection (MidiMini) query results

In the Music Videos collection, the user is presented with a choice of video output format, as illustrated in Figure 7; sometimes there are only one or two choices, but sometimes – as shown here – there are three. When tested, selecting one of the leftmost icons resulted in the file being downloaded and then played using QuickTime; selecting the centre icons ('Mpeg') resulted in the file being downloaded but then 'disappearing'; selecting the rightmost icon ('Real Video') resulted in an immediate error message. Since these video files are large (typically 4 or 5 MB), they can take several minutes to download, making the user cost of downloading high, and particularly so if the resulting file cannot be played.

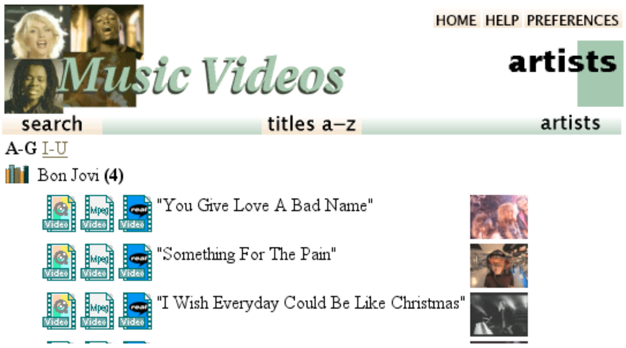

Figure 7: Music Videos browsing results

In the Folk Music collection, the user is presented with no choice of output format for the available output modalities. This lack of choice leaves less scope for user error.

In summary, for the broad range of web users that might access these systems, there has to be a clear way of presenting the different modalities of output that are possible (e.g. sound, score, text), but there should be a sensible default for file formats; for example, MIDI, pdf and html (respectively) would be sensible defaults for web-accessible music collections.

3.5 Interfacing with other systems

All the collections studied are implemented to be accessed via a web browser, and to make use of browser plug-ins such as QuickTime. In an informal study with novice users, we found that the boundary between the collection software and such plug-ins was not always obvious to them. With the exception of the Folk Collection, when music files are played, the user sees a window from which the only way to reach another page is via the browser's 'back' button. In many cases, the moving position indicator on the QuickTime audio control panel (Figure 8) might be considered redundant, as the user can also hear the music playing. However, it provides complementary information – to indicate how far through a tune we are – and also might alert a user to the fact that the sound is turned down on their computer (if they can see the indicator moving but not hear the melody).

Figure 8: QuickTime audio control panel

If a user is trying to retrieve tunes by melody matching in the NZDL Music Collection, this necessitates constant switching between NZDL and sound recording software. We get a cycle of activity in which the user, working with NZDL, has to exit to the operating system, initialises or re-selects the sound recording software, records some singing, saves that to file, then returns to the web browser (i.e. NZDL Music Collection) and browses the disk to find the file just saved (having to remember the name, of course) to initiate a new search. This places a high cognitive load on the user, having to remember the sequence of operations and the latest name of the file. Remote melody matching over the Internet places heavy constraints on the developer (the melody has to be saved in order to be transmitted), so this difficulty may be unavoidable, but developers should be aware of the problem.

4. CONCLUSIONS

We have seen that, in the web-accessible musical digital libraries studied, the greatest difficulty is not the 'medium mismatch problem' [14], whereby documents and queries are expressed in different media, but the data mismatch problem for melody matching. This problem arises both because of the way that many collections have been gathered (having variable quality) and because of intermediate representations used in the matching. Thus, there is a tension between notional ease of use and usefulness and actual ease of use. While the idea of melody matching is intuitive and appealing, the current state of the art is such that major usability difficulties still exist.

There have been rapid developments in web technologies. A range of technical formats for each medium (text, sound, graphics, video) have been developed, some of which are proprietary or platform-specific, others of which are now very widely available. Where system developers have catered for a variety of remote user systems, this can result in confusion for any individual user, who does not necessarily know what is available on their particular system. As standards converge, and as it becomes increasingly possible for the web server to identify features of the client (so that decisions about formats can be made by systems without user intervention), it should be possible to minimise or eliminate user involvement in format selection.

There is a tension between the widespread aim of catering for novice users and the need to demand a comparatively sophisticated understanding of the technology and concepts that underpin musical digital libraries. The various libraries studied here have demanded different levels of musical sophistication of their users: the more powerful retrieval mechanisms unavoidably demand a deeper understanding of underlying technologies. The design challenge is to minimise the understanding needed and to communicate effectively with users in the users' language.

Looking to the future, we can identify usability requirements that apply particularly to digital music libraries. One is that format options should be transparent to users – or should default to common standards for novices, with choices available to more sophisticated users – so that users can focus on modality and information content options. For example, novices should be able to choose between text, score, ABC and sound modalities, without having to choose between (say) PS and PDF or AIFF and MIDI1. Here, by 'novices', we mean people who are unfamiliar with computer music and the various computer music formats, although they may be expert musicians.

Designers need to pay attention to the real challenge of trading precision for accuracy. More seamless integration of technologies should gradually become easier to achieve. In this paper, we have not considered general web [15] or library [6] usability issues, although universal requirements such as systems being self-explanatory and giving appropriate and timely feedback apply as much to music digital libraries as to any other. The provision of truly usable web-accessible musical digital libraries represents a huge challenge; this study has provided some pointers towards areas that need more attention.

5. ACKNOWLEDGMENTS

This work is supported by EPSRC Grant GR/M71848. We are grateful to Joanne Hyde for guidance on the application of EMU, to George Buchanan for technical advice, and to David Bainbridge and John Chambers for feedback on the design of the NZDL Music Collection and JC's ABC TuneFinder (respectively) and for constructive comments on an earlier version of this paper. Figures are reproduced with permission.

6. REFERENCES

[1] Alty, J. L. (1991) Multi-media - what it is and how we exploit it. In Diaper, D. & Hammond, N. (Eds.): Proceedings of HCI '91. Cambridge University Press, pp. 31-46

[2] Bainbridge D., Nevill-Manning C., Witten I.H., Smith L.A. and McNab R.J. (1999) Towards a digital library of popular music In E.A. Fox and N. Rowe (Eds.) Proc Fourth ACM Conference on Digital Libraries,edited, pp 161-169. ACM.

[3] Bernsen, N. O. (1995a) A revised generation of the taxonomy of output modalities. In Bernsen, N. O., Jensager, F., Lu, S., Verjans, S. (eds): Information theory and information mapping, Amodeus project deliverable D15

[4] Bernsen, N. O. (1995b) A taxonomy of input modalities. In Bernsen, N. O., Jensager, F., Lu, S., Verjans, S. (eds): Information theory and information mapping, Amodeus project deliverable D15

[5] Blandford, A. E. & Stelmaszewska, H. E. (2002) Evaluating Multimodal Usability of Musical Digital Libraries: a Case Study. RIDL Technical Report. Available from http://www.cs.mdx.ac.uk/ridl/

[6] Blandford, A., Stelmaszewska, H. & Bryan-Kinns, N. (2001) Use of multiple digital libraries: a case study. In Proc. JCDL 2001. 179-188. ACM Press.

[7] Chambers, J. (personal communication) Email message dated 6th April 2002.

[8] Coutaz, J., May, J., Young, R., Blandford, A., Nigay, L., Salber, D. (1995) Integrating system and user modelling through abstraction: the CARE properties for reasoning about multimodality. In: Nordby, K., Helmersen, P., Gilmore, D. J., and Arnesen, S. (eds): Human-Computer Interaction: Interact'95. Chapman and Hall, pp. 115-120

[9] Cunningham, S.-J. (2002) User Studies: A First Step in Designing an MIR Testbed. In J. S. Downie (Ed.) The MIR/MDL Evaluation Project White Paper Collection Edition 1. pp. 19-21

[10] Fuhrman, M., Gauthier, D. & Dillon, A. (2001) Usability Test of VARIATIONS and DML Prototype. Available from www.dml.indiana.edu/pdf/VariationsTest.pdf

[11] Hyde, J. K. (2002) Multi-Modal Usability Evaluation. PhD thesis. Middlesex University

[12] McNab, R. J., Smith, L. A., Bainbridge, D. & Witten, I. H. (1997) The New Zealand Digital Library MELody inDEX, D-Lib Magazine, May 1997

[13] McNab R.J., Smith L.A., Witten I.H. and Henderson C.L. (2000) Tune retrieval in the multimedia library Multimedia-Tools and Applications 10,113-132. Kluwer Academic Publishers.

[14] Meghini, C., Sebastiani, F. & Straccia, U. (2001) A model of Multimedia Information Retrieval. Journal of the ACM. 48. 909-970.

[15] Nielsen, J. (2000) Designing Web Usability: The Practice of Simplicity. New Riders

[16] Parsons, D (1975) The Directory of Tunes and Musical Themes. Spencer Brown, Cambridge.

[17] Purchase, H. (1999) A semiotic definition of multimedia communication. In: Semiotica, vol 123-3/4, pp 247-259