ONLINE DATABASE OF SCORES IN THE HUMDRUM FILE FORMAT

Craig Stuart Sapp

Center for Computer Assisted Research in the Humanities
Stanford University
Centre for the History and Analysis of Recorded Music
Royal Holloway, University of London
`craig@ccrma.stanford.edu`

ABSTRACT

KernScores, an online library of musical data currently consisting of over 5 million notes, has been created to assist projects dealing with the computational analysis of musical scores. The online scores are in a format suitable for processing with the Humdrum Toolkit for Music Research, but the website also provides automatic translations into several other popular data formats for digital musical scores.

Keywords: Musical Data, Digital Scores, Humdrum Toolkit, Computational Music Analysis.

1 INTRODUCTION

The KernScores website (http://kern.humdrum.org) was developed to organize musical scores in the Humdrum **kern data format originating from a variety of musical sources. The website was designed to allow easy access to the data for students doing projects and coursework in CCARH courses about musical information and computational music analysis at Stanford University (http://ccarh.org/courses). Students work on projects during the second quarter in these classes, primarily facilitated by the Humdrum Toolkit for Music Research [1].

The Humdrum file format is text-based, allowing for easy viewing and editing in any standard text editor. Each staff/part in a work with more than one instrument forms a column similar to cells in a spreadsheet with musical time progressing by successive rows, and with all elements on the same row occurring at the same performance time. In addition to musical data, other forms of data can be included in their own columns (called *spines* in Humdrum terminology) that facilitate analytic mark-up of the musical data. Here is an example measure of music in the Humdrum file format:

```
**kern  **kern  **kern  **kern  **harm
4BB     4G      8d      4g      IVb
.       .       8c#     .       .
4AA     4A      4d      4f#     Ic
4GG     4B      4d      4e      ii7b
4AA     8E      4c#     4a      V
.       8A      .       .       .
*-      *-      *-      *-      *-
```

In this particular example, there is an extra column in the data for harmonic analysis labels. This example measure of Humdrum data represents the following graphical musical information:
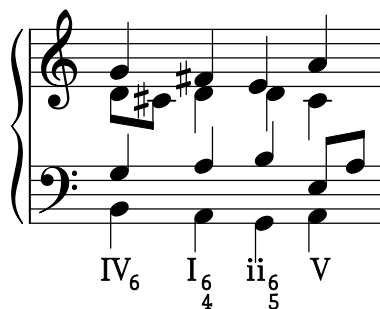
2 DATA ORGANIZATION

Musical scores on the website are accessible either by searching for bibliographic information such as composer and title on the main page, or by browsing through the database holdings in a hierarchical tree structure based on individual collections' organization. In addition, a list of most composers and genres available at the website are found on the main webpage. No monophonic or polyphonic searching of the actual musical data has yet been implemented.

Online access to music is controlled by a CGI interface to the data. Music currently under copyright restrictions is available only to users with login privileges. Access to all public-domain music is available to any website user without the need to log in.

Each data file represents one movement in multi-movement pieces, or an entire composition for single-movement works. Every data file has an information page that contains three sections: (1) bibliographic information stored as reference records in the data file, (2) automatic data translation links, and (3) basic online analysis links. These data pages are accessible directly from search results pages, or from the browsing pages that list either single works, or larger collections of works.

Browsing and search-results pages usually display three icons to the left of each work/movement title for quick data access. Clicking on the "H" icon yields the actual Humdrum data file; clicking on the "M" icon yields an automatic conversion of the Humdrum data into a MIDI file, and clicking on the "S" icon will download a PDF file with a graphical score for the work (when available). Clicking on the work titles takes a

user to the work information page. For power users, all browsing pages have a "Z" icon near the top of the page that allows download of the ZIP-compressed data contents of each page, and certain pages also allow downloading of a zipped file containing the recursive contents of data from all sub-pages.

3 DATA ENTRY

During the past year, input of music into the Humdrum data format has been greatly facilitated by the SharpEye program for optical music recognition [2]. SharpEye (only available for the Microsoft Windows operating-system) allows for output into the MusicXML data format, which is then translated into the Humdrum file format with a program, named *xml2hum*, written by the author[1].

During the past year, over 800 works/movements consisting of approximately one million notes have been processed by the author from scanned images of musical scores using SharpEye. Most music input using the SharpEye program has been for piano solo or string quartet. Using SharpEye as data input for orchestral scores is currently being evaluated at Stanford University and the Ohio State University.

Music entered into the Finale music typesetting program can be converted into MusicXML files using the Dolet plug-in[2] (for both Windows and Macintosh OS X) which can also be converted into Humdrum with xml2hum.

4 DATA TRANSLATIONS

The native storage format for musical scores in Kern-Scores is the Humdrum file format; however, six other digital musical formats can be converted on demand from the source format in real-time:

1. *Standard MIDI Files* – useful for listening to the musical data and for inputting into analysis programs that require only basic musical descriptions of scores.
2. *MusicXML* – An XML-based musical score interchange format that is used in many popular commercial programs such as Finale, SharpEye, and ScoreMaker[3] [3].
3. *Guido Music Notation* – A data format which can be used to generate graphical music at http://noteserver.org and can be processed with other programs listed at http://www.salieri.org/guido.
4. Note lists for the *Melisma Music Analyzer* which is a set of programs for harmonic analysis [4].
5. Input data for *Director Musices* which is a program for computer-performance of digital scores using performance rules [5].
6. *SKINI* – A text-based format that is similar to MIDI and used for physical modeling software synthesis in the Synthesis Toolkit [6].

If available, a printable score for the work is also available in the data translation section of each work description page. If the music was scanned from out-of-copyright musical scores, then the scanned music is often available in PDF format for reference, and is also useful for identifying and correcting errors in the Humdrum data file derived from the scanned source.

5 ONLINE ANALYSES

A few basic online analyses are available at the bottom of each data file description page. These links provide CGI interfaces to Humdrum Toolkit programs as well as other programs written by the author that are run in real-time on the Humdrum data files when a user accesses an analysis hyperlink.

For example, here is the opening of Chopin's Prelude Op. 24, No. 10 in C# minor in piano-roll notation available from the analyses section:
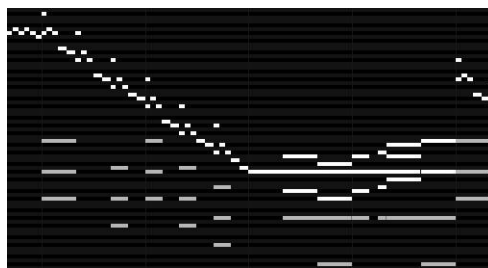
The analyses section for the prelude (and all other work pages) also has links to programs that will count pitch-classes in the file, generate MIDI files of the prelude with mixed-up pitch content, and allows for an on-demand harmonic analysis of the prelude using the Melisma Music Analyzer.

REFERENCES

[1] Huron, David. The Humdrum Toolkit: Software for Music Research. http://dactyl.som.ohio-state.edu/Humdrum.

[2] Jones, Graham. SharpEye Music Reader. http://www.visiv.co.uk.

[3] Good, Michael. MusicXML Definition. http://www.recordare.com/xml.html.

[4] Sleator, Daniel and Davy Temperley. The Melisma Music Analyzer. http://www.link.cs.cmu.edu/music-analysis.

[5] Friberg, Anders. Director Musices. http://www.speech.kth.se/music/performance/download.

[6] Cook, Perry and Gary Scavone. The Synthesis Toolkit in C++ (STK). http://ccrma.stanford.edu/software/stk.

---

[1] http://extras.humdrum.org/man/xml2hum

[2] http://store.recordare.com/doletfin2.html

[3] http://www.kawai.co.jp/cmusic/products/scomwin