Hundreds of Thousands of Pieces in MEI: Encoding Tablatures at Scale

Introduction

Tablature is one of the most significant forms of notation of western music, if measured by sheer number of pieces. Nearly 50,000 pieces for plucked instruments were notated in tablature up to the 18th century, and a contemporary revival of tablature in the online guitar community has led to the production of hundreds of thousands of new tablatures. Until now, this form of notation and the huge repertoire associated with it has largely been inaccessible to digital musicologists due to a lack of tools.

The present paper reports on recent efforts to capture and encode the vast online repertory of popular music as well as historical tablatures in a common MEI format. We demonstrate how this can be used with a new Python tablature module built using music21 to enable large-scale processing and analysis of this music [1].

Tablature in the wild

Tablatures were first used as a form of notation for plucked and bowed string instruments starting in the late 15th century, and it is this form that is most familiar to historical musicologists. It indicates where each string of the instrument should be stopped and when it should be sounded, rather than pitches and durations as in staff notation. Largely because of this specialised notation, although the lute was the preeminent solo instrument for much of the 16th and 17th centuries, this repertoire has taken a backseat to vocal music in our view of this period. By making the music accessible in music21, we support the creation of digital tools to address this by automatically transcribing tablature into standard notation, and our robust MEI specification allows the creation of flexible documents that blend tablature and standard notation for musicological study.

Today, a huge number of guitarists use tablature to share music online. Tablature is easily shared over the internet, as its essential features can be transcribed in plain text, using a monospaced font and the ASCII character set, viewable in a standard web browser, mail

client, or text editor. These are 'born digital' resources, presenting no digital transcription problem, and easily ingested into a toolchain for digital musicologists. UltimateGuitar.com, a user-contributed repository, contains over 150,000 tablature transcriptions for guitar and bass guitar. For scholars of 20th and 21st century popular music, this substantial and important dataset presents music in the primary genres for which the most popular instrument of all time continues to be used.

Past work: existing data models and tools

'ASCII' tablature is by far the dominant format for online tablature distribution. Significant but smaller collections exist in TabCode [2] and abctab [3], and a variety of less-open formats that are specific to particular software packages. Although HUMDRUM includes a tablature format, **fret, this is little used [4].

MEI contains a tablature module, but this schema can only encode the most basic tablatures and contains inconsistencies that need addressing [5]. Work done in the Transforming Musicology project produced first steps towards a more robust MEI schema for tablature [6][7][8]. music21 at present contains a modest tablature module, with a representation of chord symbols but no provision for actual music notated in tablature. Our work builds that module out into a robust set of tools for representing and manipulating music in tablature.

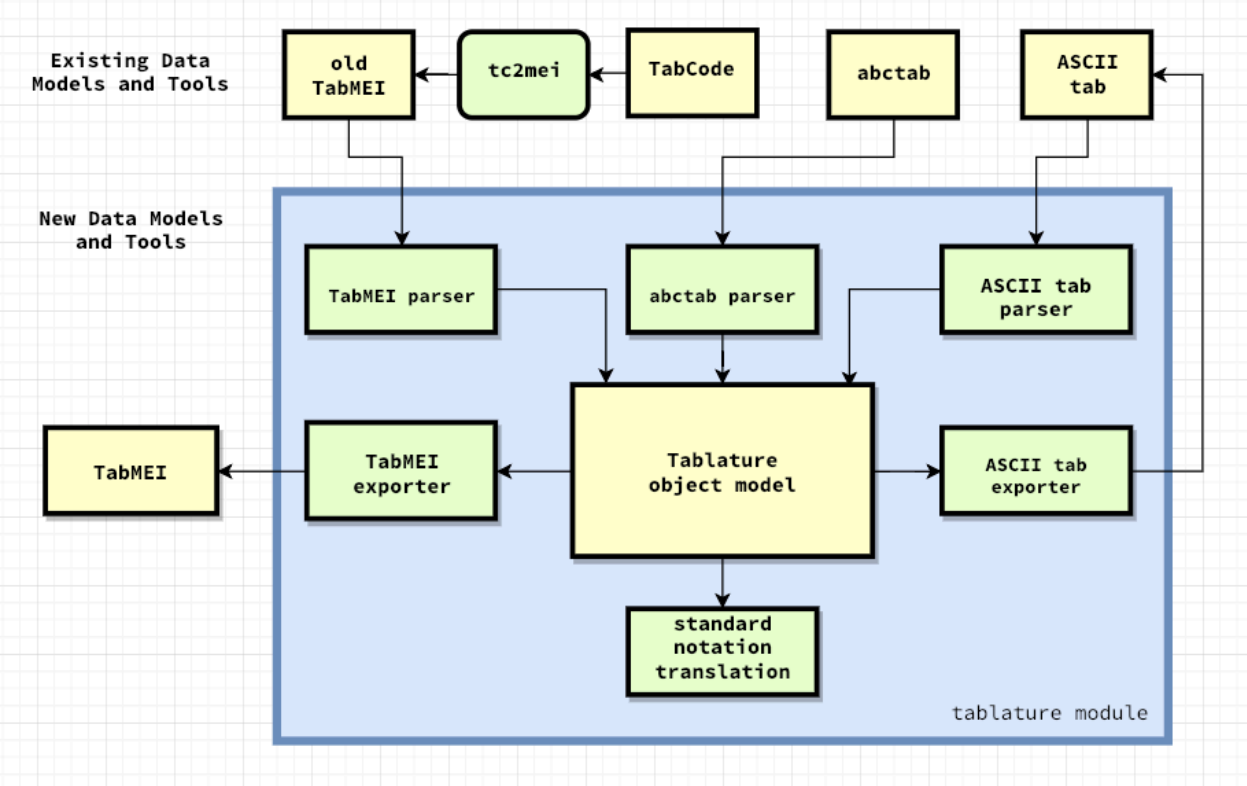
New data models and tools

In this paper, we propose a new MEI schema for tablature that addresses the shortcomings of the old module. In particular, we address the variety of forms of tablature and the diversity of fretted instruments while maintaining a common representation. Our schema separates the encoding of the tablature from a representation of the instrument, allowing encoders to specify important distinguishing details, while providing sensible defaults.

In conjunction with our work on this new MEI schema, we have developed a set of tools based on music21 in order to provide users with a flexible means to manipulate this data. We provide an object model for tablature and a set of parsers and exporters for common tablature formats. This allows scholars to use tablatures as data, yet remain effectively

agnostic to their original format and, for example, to view music originally encoded in TabCode and abctab side-by-side.

Furthermore, once tablature is imported into our model, we can create a rough transcription into standard notation, which allows users to harness the full power of the music21 toolkit to perform analysis on this data, or to place it alongside music transcribed in standard notation. In this way, tablature can be re-integrated into the world of notated music and analysed using tools that have years of development for use on standard notation.



We present a variety of proof-of-concept projects created using this new pipeline. Using the UltimateGuitar.com dataset, we can ask some basic questions that might be of interest to musicologists or music theorists such as: distribution of the most common chords; alignment between popularity and number of transcriptions; and automatic detection of pieces in common variant tunings. Using the Renaissance lute data set, we could show

analogous work: extraction of the most common cadential formulae or the most common embellishment formulae in 16th century lute music.

Future work

As the work presented here is a set of tools, future work in this area is limited only by the imagination and curiosity of scholars who take them up. Musicologists interested in pop/rock music could use the UltimateGuitar.com dataset to study questions that vary from sociological to music-theoretical. As part of our research on playability [9], we intend to use the ASCII tab dataset to widen our evidence base for the playability of music on plucked instruments, heretofore limited to historical lute tablatures. There are ample opportunities to use our tools to create MELD applications including, e.g., the alignment of ASCII tab with media such as YouTube videos. We could also create pedagogical materials; e.g., extracting the most common cadential or ornamental materials from a repertoire and grading them by difficulty enables the creation of smarter teaching materials.

References

[1] M. Cuthbert, Music21: a Toolkit for Computer-aided Musicology <http://web.mit.edu/music21/>

[2] http://www.ecolm.org/?page=TabCode

[3] C Dalitz, abctab2ps, <http://www.lautengesellschaft.de/cdmm/>

[4] D. Huron, 'Humdrum Representation for Tablature' <https://csml.som.ohio-state.edu/Humdrum/representations/fret.rep.html>

[5] Music Encoding Initiative, 'Tablature Notation' <https://music-encoding.org/guidelines/v4/content/tablature.html>

[6] D. Lewis, R. Lewis., T. Crawford, 'MEI for lute music', MEC 2015 (Florence)

[7] T. Crawford, J. Schwartz, D. Lewis, and R. Lewis, 'Encoding Music As People Play It: MEI and the role of tablatures in capturing musical performance', MEC 2016 (Montreal)

[8] R Lewis, 'Customising MEI for lute tablature' <https://tm.web.ox.ac.uk/blog/customising-mei-lute-tablature>

[9] [Reference removed for anonymity]