Music encoding can link disparate types of musical data for the purposes of archiving and search. The encoding of human response data explicitly in relation to musical notes facilitates the study of the ways humans engage with music as performers and listeners. This paper reflects on the developments and trends in formal music encoding systems as well as the types of data representations used in corpora released by researchers working on expert music analyses, musical performances, and listener responses. It argues that while the specificity (and often simplicity) afforded by project-specific encoding formats may be useful for individual research projects, larger-scale interdisciplinary research would be better served by explicit, formalized linking of data to specific musical elements. The paper concludes by offering some concrete suggestions for how to achieve this goal.

Keywords: music encoding; musical notes; musical elements; music analysis; music performance; listener responses

1. Introduction

Standardized music encoding has the potential to facilitate deep interdisciplinary engagement and research, where data can be shared between researchers to assess comparative research questions on how humans engage with music. Specifically, by consistently encoding human response data (such as harmonic analyses, tuning, or timing data) and reported and physiological listener responses to musical elements (such as chords or notes) researchers can draw on otherwise disparate data describing how people analyze, perform, and react to music. As the music information retrieval (MIR) community becomes more and more committed to publicly licensed data (McFee et al., 2018), it is important to consider how the formats and data models used to encode these data may limit subsequent research. The design decisions about project-specific encoding formats are typically limited by the research questions being asked in a particular project. Linking the project-specific data to music elements in a standardized encoding format allows other researchers to more easily use the data for their research questions, thus broadening the applicability of the data.

A significant motivation for encoding data at the note- and/or chord-level is to provide direct linkages between symbolic and audio representations of music with human response data at a level that corresponds to objects that humans perceive when listening to music (as opposed to, for example, specific timing points in milliseconds). This paper considers the acts of analysis, performance, and listening to be related but distinct types of human responses to music. Figure 1 provides a visual guide to the ways data derived from these human responses to music connect with symbolic and audio representations of music. The inner paths illustrate how representations relate to the human responses, shown on the outside of the figure. For example, an expert analyst may look at a score and/or listen to a recording of a piece when engaging in musical analysis. Similarly, a performer may or may not be performing notated music in the creation of a recorded performance, while listeners typically are only engaged with a recording and not reading a notated score. The dotted lines around the figure between the three types of human musical experiences emphasize that we can use data from each to better understand particular phenomena occurring in the others. For example, an expert analysis may inform performance practice (e.g. Schmalfeldt, 1985) and, although less robustly theorized, listener responses may do the same.

This paper begins by surveying the developments in music encoding research, with a particular focus on developments related to the MIR community in the past 20 years (Section 2). Throughout this paper, encoding refers to not only the encoding of score-based elements but a wider-range of musical data (including annotations and data extracted from audio analysis). It then presents a
2. A Brief History of Music Encoding

The need for digital music encoding standards to advance computational musicology was recognized as early as the 1960s (Erickson, 1970). One of the earliest and most widely used encoding languages was the ASCII-based Plain and Easie code (Brook, 1965), used for encoding incipits in the Répertoire International des Sources Musicales (RISM) collection. Another was the Ford-Columbia Music Representation, subsequently known as DARMS (Erickson, 1975), which was designed to facilitate both engraving and computer-aided music research. An important issue that was raised in these early years, and that we continue to grapple with today, is striking a balance between simple encodings that fulfill the needs of a single project and more complex encodings that can be used to generate a complete musical score (Wolff, 1977). One of the most significant developments in the 1980s was the establishment of the Musical Instrument Digital Interface (MIDI) protocol (Loy, 1985; Moog, 1986). Though originally designed for sending control messages between hardware music devices, it has also been used extensively by music researchers. Another significant development was the increased popularity of text-based music encoding languages designed explicitly for research, rather than engraving or sound generation, such as MuseData (Hewlett, 1997). Humdrum (Huron, 1994) falls into this tradition of human-readable encoding languages. Thorough histories of music representations before 2000 can be found by Dannenberg (1993), Selfridge-Field (1997), and Brown (1999).

Music encoding also has a long history within the MIR community. The original ISMIR symposium in 2000 included an invited talk by Bonardi (2000) on how MIR techniques could serve musicologists, which in part reflected on the encoding and representation of musical data, as well as one by Chen (2000), who discussed representation and retrieval issues in the Muse project. Two posters dealt with encoding issues: The MuTaTeD’II project (MacLellan and Boehm, 2000) for rendering the Standard Music Description Language (SMDL) and the Notation Interchange File Format (NIFF) into a more easily computer-processable representation, and a description of an early version of MusicXML (Good, 2000). At ISMIR 2002, Peeters and Assayag (2002) gave a tutorial on digital representations for retrieval, composition, performance, and recommendation tasks. Futrelle and Downie (2003) addressed the issue of music representation in their survey of interdisciplinary issues in the early years of ISMIR (2000–2002). They highlighted some of the early work in symbolic representations for non-Western classical music, such as the work by Politis and Linardis (2001) on Byzantine music, the work by Lee et al. (2002) on Korean music, and Geekie (2002) on Carnatic ragas. Also in 2003 was the first paper on a digital edition project (Olson and Downie, 2003). These projects have also motivated the development of music encoding formats.

There have been several extensions to Humdrum proposed at ISMIR over the years (e.g., Kuuskankare and Laurson, 2004; Knopke, 2008; Kuuskankare and Sapp, 2013): Note and rhythm data is typically encoded in Humdrum using the **kern representation. **kern has been used to encode datasets (e.g., Sapp, 2005) and is supported by a number of tools that can interpret Humdrum’s representation formats (e.g., Kornstädt, 2001; Cuthbert and Ariza, 2010). MEI is an XML-based schema initially developed by Roland (2002) that allows for explicit separation of the presentation and content aspects of musical documents (Pugin et al., 2012). In recent years it has been extended to allow for full-scale document encoding (Hankinson et al., 2011), as well as integrated into Verovio (Pugin et al., 2014), a project for engraving musical scores in web browsers. MusicXML and MuseScore have also been used by researchers within the MIR community, but the development of the encoding formats themselves have taken place outside of the MIR community and largely with the goal of facilitating robust transfer between different music notation programs or encoding formats. At ISMIR 2008, a tutorial on the strengths and weaknesses of commonly used music encoding formats was presented by Selfridge-Field and Sapp (2008), but, beyond this, consideration of this issue at ISMIR has been limited.

One major advancement with encoding in recent years is the development of tools to integrate encoding formats...
into linked data representations of different types of musical information. Historically, however, much of the work on linked data has developed methods for linking different musical documents (Weigl et al., 2019b). An exception is Fields et al. (2011), who explored integrating human annotations of musical form into a linked-data framework, and more recent work has begun to incorporate note-level encodings into these frameworks, facilitating content-based search and indexing. An example of this is the Music Encoding and Linked Data (MELD) framework (Weigl and Page, 2017; De Roure et al., 2018; Lewis et al., 2018; Page et al., 2019), which uses MEI encoding and is integrated into the large-scale Towards Richer Online Music Public-domain Archives (TROMPA) (Weigl et al., 2019a) project. The work by Meroño-Peñuela et al. (2018) integrating MIDI encodings within a linked-data framework also provides the opportunity for linking data to musical elements. These projects point towards a solution to some of the issues that arise with the current datasets, although currently this work has predominantly produced protocols and tools as well as datasets with only document-level linkages (e.g., Meroño-Peñuela et al., 2017).

3. Expert Analyses
Expert music analyses, such as those published in music theory textbooks (Devaney and Shanahan, 2014), can help researchers gain insight into a range of musical questions. With respect to the notated music itself, it can help us define what is characteristic about the work of a single composer or a group of composers within a particular time period or examine how compositional practices evolve. It can illuminate elements of the cognitive processes of the analysts themselves and help researchers understand what trained listeners attend to in musical works and/or which musical elements are salient in various musical contexts. Studying expert analyses also gives us a window into the pedagogical practices that are used to train musicians, as these analyses form the basis of much of the conservatory curriculum in the Western art music tradition. This paper focuses on written-out harmonic analysis and chord labeling in musical scores and audio and is distinct from statistical analysis of unannotated music scores (e.g. Rohrmeier and Cross, 2008; Condit-Schultz et al., 2018; White and Quinn, 2018). Specifically, the focus for this section is on analytic annotations of notated Western art music/jazz lead-sheets and from popular music audio, as these have been the focus of the work published in MIR-related communities.

The harmonic analysis datasets released in the past two decades can be divided into two groups, those that only provide chord annotations (sometimes linked to time points in corresponding audio files), shown in Figure 2 and discussed in Section 3.1, and those that link symbolic note data to a range of analytical annotations (including chord annotations or structurally significant notes), shown in Figure 3 and discussed in Section 3.2. Datasets with linked note data have been released more recently than the ones that consist only of chord annotations. This suggests a trend of increased interest in linking multiple levels of musical data in the datasets that researchers are producing. There is also an increase in researchers using established encoding formats, with newly defined extensions as needed, rather than defining new encoding protocols.

![Figure 2: Example encodings from five chord-only datasets: (a) Kostka-Payne (Temperley, 2009), (b) Real World Computing [RWC] Functional Harmony (Kaneko et al., 2010), (c) iRealB Corpus of Jazz Standards (Broze and Shanahan, 2013), (d) Beatles (Harte et al., 2005), and (e) the McGill Billboard Dataset (Burgoyne et al., 2011).](image-url)
Figure 3: Example encodings from five note-level harmonic analysis datasets: (a) Generative Theory of Tonal Music [GTTM] (Hamanaka et al., 2014), (b) schenker41 (Kirlin, 2014), (c) Theme And Variations Encodings with Roman Numerals [TAVERN] (Devaney et al., 2015), (d) Annotated Beethoven Corpus [ABC] (Neuwirth et al., 2018), and (e) Beethoven Piano Sonatas with Functional Harmony [BPS-FH] (Chen and Su, 2018).
3.1 Chord-only annotations

Figure 2 shows examples from datasets that encode chord-level data separate from related note-data. The two Western art music datasets consist of transcriptions of the chord annotations in Kostka-Payne's *Tonal Harmony* textbook (Temperley, 2009) and functional harmony annotations of the Real World Computing (RWC) database (Goto et al., 2002; Kaneko et al., 2010). Examples of these two encoding formats are shown in Figure 2 in subplots (a) and (b). The Kostka-Payne annotations are linked to MIDI renditions of the original scores from the textbook and indicate the starting and ending time in MIDI frames of key and chord changes, with chord tones represented as pitch classes in relation to the key. The RWC annotations indicate the measure each chord change occurs in but do not explicitly indicate chord duration, although this can be assumed for measures with a single chord and inferred based on typical harmonic rhythm practices for measures with multiple chords. Subplot (c) in Figure 2 shows an example of lead-sheet based chord annotations. The iRealB Corpus of Jazz Standards (Broze and Shanahan, 2013) consists of chord progressions from a digital repository of jazz standard lead sheets using an extension to the Humdrum format that accounts for jazz chords. The two subplots in the bottom of Figure 2 show examples of two of the most widely used chord datasets in the MIR community based on audio annotations (d) the Beatles dataset (Harte et al., 2005) and (e) the McGill Billboard dataset (Burgoyne et al., 2011). Both of these datasets provide timing information for audio recordings corresponding to the chordal analyses. The Beatles dataset provides onset and offset times for each chord label, while the Billboard dataset provides onset times for the beginning of each measure. With the development of robust automatic music transcription (AMT) systems (Benetos et al., 2013), transcribed note data could be linked to the chord annotations via the encoded timing information.

3.2 Linked chord annotations

More recently, several datasets have been released that provide harmonic analyses explicitly linked to note encodings. Examples from five such datasets are shown in Figure 3. In subplot (a) is the GTTM Database (Hamanaka et al., 2014), based on the Generative Theory of Tonal Music proposed by Lerdahl and Jackendoff (1983) and encoded in separate linked XML files based on MusicXML for note data and five different analytic components of GTMM (prolongational trees, time-span trees, grouping structures, metrical structures, and harmonic analyses). Another example of a prolongational dataset is the schenker41 dataset (Kirlin, 2014), shown in subplot (b), where the analytic data is encoded in a system devised by the authors and the score data is encoded in MusicXML. Recently, Rizo and Marsden (2019) proposed an MEI extension for hierarchical analysis. This extension combines standard MEI encoding with tree-based structures from the Text Encoding Initiative (TEI), which MEI grew out of, to encode hierarchical musical relationships.

There are also a number of datasets that build on Roman numeral analysis, rather than prolongational analysis, although these also provide varying degrees of hierarchical modelling. The Theme And Variations Encodings with Roman Numerals (TAVERN) dataset (Devaney et al., 2015), in subplot (c) combines existing Humdrum **kern** and **harm** encoding specifications with a newly defined **func** specification for an additional level of hierarchical analysis: three-part harmonic function (tonic, predominant, and dominant) at the level of the musical phrase. In contrast, the Annotated Beethoven Corpus (ABC) (Neuwirth et al., 2018), shown in subplot (d) of Figure 3, uses a newly-designed format that links harmonic analysis in a tabular format (shown in the left side of the subplot) with a MuseScore XML-based encoding of the note data (shown in the right side of the subplot). The Beethoven Piano Sonatas with Functional Harmony (BPS-FH) dataset (Chen and Su, 2018), shown in subplot (e), also provides linked representations, but with an underspecified score using selected MIDI values to encode note, chord, and phrase information.

3.3 Summary and Discussion

While some of the datasets integrate and/or extend existing encoding systems, there is a great amount of variability in which existing encoding system is used and how much it is modified to suit the needs of the specific research project. This includes the use of some aspects of the MIDI protocol in the Kostka-Payne and BPS-FH datasets, the use of Humdrum and Humdrum-expanded syntax in the iRealB and TAVERN datasets, and the linking of project-specific annotation syntax with either MusicXML or MuseScore for encoding score data. While the project-specific modifications and enhancements are more compact, they can limit the accessibility of the data and the types of further analysis that can be performed. Furthermore, different encoding formats have different levels of interoperability. Humdrum and MEI have open development that encourages contributions and expansions from community members, meaning that a dataset like iRealB could be linked to relevant note data in **kern** format, should that become available, and the harmonically annotated note data in TAVERN can be more easily processed with slight modifications to existing software tools like the Humdrum (Huron, 1994), jRing (Kornstädt, 2001), or music21 (Cuthbert and Ariza, 2010). This would also be true for prolongational analyses encoded using Rizo and Marsden's MEI analytic extension, where the annotations and note data are explicitly linked, much more easily than with the separation of analytic encoding from the MusicXML or MuseScore data in the GTTM, schenker41, or ABC datasets. Audio datasets, like the Beatles or Billboard, could also be processed with these software tools, once automatic music transcription methods (Benetos et al., 2013) are more fully developed and able to output into a readable format. The explicit encoding of note data in a standard format also allows for easier and more reliable linking between datasets. Another alternative is linking these audio-based annotations to
Note-level encodings also offer the advantage of the annotations not being limited to Roman numerals or popular music chord names and symbols. For example, as an alternative to chord label annotations, an encoding could specify which notes are members of the chord and which are not. Harmonic analyses represented in this way (Devaney and Arthur, 2015; Kaliakatsos-Papakostas et al., 2015) facilitate alternative interpretations and do not require handling expansive chord vocabularies (McFee and Bello, 2017). Note-level encoding also allows for a broader range of empirical analytic theories to be implemented, as the GTTM and schenker41 datasets demonstrate. These types of annotations can also be made and recorded in real-time during a performance with the appropriate visualization and annotation technology (Page et al., 2015).

Expert harmonic analyses facilitate the modelling of the relationship between lower-level musical elements and higher-scale musical structures. This can be done either for corpus analysis purposes, such as the XML-based method proposed by Gotham and Ireland (2019), or for visualization of large-scale structures, such as the JSON-based system described by Giraud et al. (2018). These analyses can also inform the analysis of data extracted from performances or collected from listener responses, with note-level representations being particularly useful for linking with performance data as we shall see in the next section.

4. Performance

In the Western art music tradition, music scholars have historically focused on musical scores, but it is musical performances that both convey musicians’ interpretations and are what listeners actually hear. Recently performance has been more deeply considered by musicologists (Cook, 2013). Thus, linking performance data with symbolic note data through encodings allows researchers to more fully explore the relationship between musical materials and listener experiences. For non-notated music, linkages can be made between performance data and other musical objects, such as chords, beats and downbeats. Studying performance can help us gain insight into a range of questions. For example, historically, we can consider how performance practices evolved across time and distance (e.g., Timmers, 2007) and computationally, it can help us to develop models of “expressive” performance (e.g., Kirke and Miranda, 2012; Cancino-Chacón et al., 2018).

In order to link performance data from audio with symbolic data, the low-level audio features must be summarized into note- or higher-level descriptors. Ideally, these descriptors should encompass a range of performance parameters, including timing, dynamics, tuning, and timbre. For score-based music, these descriptors can be estimated using performance-score alignment methods on both symbolic (Gingras and McAdams, 2011; Nakamura et al., 2017) and audio (Mayor et al., 2009; Devaney et al., 2012; Huang and Lerch, 2019) music data. Higher-level descriptors of performance allow us to map directly between the symbolic representation and the performance and consider differences within and across performers/performances. Most performance datasets encode data in tables at either the beat- or note-level, with beat-level encodings typically not including any note information and note-level encoding typically not including any metrical information. Thus researchers who want to use these data in relation to the broader musical context available in encoded scores must generate their own links between these data tables and the score unless the performance data is explicitly linked to score data in the dataset.

4.1 Beat-level encodings

An example of beat-level encoding is the Mazurka Dataset (Sapp, 2007), which was created as part of the research undertaken at the AHRC Research Centre for the History and Analysis of Recorded Music (CHARM). The Mazurka dataset was generated by using Sonic Visualiser (Cannam et al., 2006) to extract tempo and dynamics data based on tapped time points in piano performances. The tempo and dynamics data were encoded at the beat-level in an Excel spreadsheet, as shown in subplot (a) of Figure 4. This encoding protocol was recently employed by Kosta et al. (2018) for their expanded version of the Chopin Mazurka dataset. In the absence of robust transcription methods for polyphonic instruments, beat-level encodings for timing and dynamics data are useful for characterizing piano performance. For monophonic instruments, including ensembles instruments where each monophonic part can be recorded individually, however, note-level encoding is both more feasible and typically preferable, as the linkages between the musical performance data and the musical objects are more explicit.

4.2 Note-level encodings

The Ensemble Expressive Performance Dataset (Marchini et al., 2014) is an example of a note-level encoding for individual monophonic instruments within an ensemble. The encoding format for these data is similar in structure to the Mazurka Dataset but here each line represents a note with the estimated onset and offset times and note name, as shown in subplot (b) of Figure 4. The QMUL Singing Dataset (Dai et al., 2015) is another example of a note-level dataset, and one with more performance features encoded, shown in subplot (c) of the figure. In addition to estimated onset and duration information, the QMUL dataset provides estimated tuning information alongside nominal duration and pitch information derived from a score. What neither of these datasets provides, however, is an explicit link to the musical score data that the performers are performing from.

An example of explicit linking between score and performance data is the performance data encoding formats used in the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT) (Devaney and Gauvin, 2019). These encoding formats allow for both reconstruction of the musical score and linking with other musical data encoded with the musical score in other datasets. AMPACT extends two existing music encoding
Devaney: Note-Level Encodings for Research on Human Engagement

Devaney: Note-Level Encodings for Research on Human Engagement

211

formats, Humdrum and MEI, and provides a means of encoding score-linked performance data for a range of note-level descriptors of timing-, pitch-, dynamics-, and timbre-related parameters. Subplot (d) in Figure 4 shows examples of the Humdrum (left) and MEI (right) extensions. Both allow for note-level encoding of a range of parameters. The MEI extension also allows for encoding of the continuous data used to estimate the note-level parameters while the Humdrum extension allows for encoding at the beat or other metrical level.

4.3 Summary and Discussion

Unlike the expert analyses, where a range of standard encoding formats are linked to annotations, the majority of performance datasets only provide limited musical note data. This ranges from metrical information but no note data in the beat-level encodings of the CHARM dataset to note names but no metrical information in the EEP and QMUL Singing datasets. In contrast, performance data in the AMPACT format are linked to either Humdrum or MEI score data, which encodes note and metrical information. The encoding of both note and metrical data is important for comparing performance data between different pieces as well as for linking the performance data to expert annotations. Linking with expert annotations, which is trivial if both datasets are encoded at the note-level in a standard format, allows for a more nuanced interpretation of the performance practice based on the musical context.

There is a long tradition of qualitative close reading of scores combined with listening or spectrogram-visualization methods for analyzing audio recordings. The encoding of multiple expert analyses and multiple performances allows for quantitative comparative study across multiple analytic and performance interpretations that was previously difficult, if not impossible, to undertake. It can also facilitate experiments on how people may annotate different performances of the same musical material differently.

Another advantage of linking performance data with a corresponding music representation is that the performance data can be visualized while a recording is played to augment the listening experience. Fornotated music, when the data are encoded or linked with a symbolic representation, the musical score can be shown simultaneously with the performance data and audio.

Examples of this include the visualization of timing information in PerformScore (Jeong et al., 2017) and Peachnote’s Tutti Tempi. Other types of performance data can also be linked to score and audio data, such as the motion capture data from performers that is

Figure 4: The upper row of the figure shows example encodings of performance at the beat-level from the (a) CHARM Mazurka dataset (Sapp, 2007) and at the note-level from the (b) Ensemble Expressive Performance [EEP] (Marchini et al., 2014) and (c) Queen Mary University of London [QMUL] Singing (Dai et al., 2015) datasets. The lower row shows example encodings from the performance data extensions to Humdrum (left) and MEI (right) in the (d) Automatic Music Performance Analysis and Comparison Toolkit [AMPACT] (Devaney and Gauvin, 2019).
visualized in the PHENICX project (Schedl et al., 2016) and performers’ EEG data that can be visualized to provide feedback for performances during rehearsal (Blanco and Ramirez, 2019). These types of data are informative for understanding both the performer(s) and the listener(s). The datasets in the next section are examples of listener responses to music that can be measured and represented.

5. Listener Responses

Continuous responses or physiological measurements from listeners give us an indication of people’s moment-to-moment experience of listening to music. Earlier work on listener experience focused on post-hoc reports of participant’s overall preference of a musical experience (Brittin, 1995) or their emotional responses (Madsen, 1997), but recent work has captured and analyzed continuous data that provide insight into how people react to specific musical moments (Goebel et al., 2014). This area of research has traditionally fallen outside of mainstream MIR research, although related research was discussed by Carol Krumhansl in her 2010 ISMIR keynote. Also, a small number of ISMIR papers have used continuous response data as features to predict from musical audio, such as Han et al. (2009) and Schmidt and Kim (2010).

Broadly speaking, studying listener responses can help us gain insight into a range of questions, including similarities and differences in how people respond to music, differences between conscious and subconscious responses to music, and how people respond to specific musical events. These questions can be addressed through reported measures, such as arousal and valence, or by physiological ones, such as heart rate, blood volume pulse, respiration, skin conductance, temperature, and surface electromyography.

5.1 Datasets

Recently released databases of continuous measurements include the Database for Emotional Analysis of Music (Aljanaki et al., 2017) and the Repeated Responses Dataset (Upham, 2018). The Database for Emotional Analysis of Music, shown in subplot (a) of Figure 5, contains reported data from subjects on felt emotional arousal and valence, encoded at 500 ms intervals. The Repeated Responses Dataset, shown in subplot (b) of Figure 5, encodes six physiological responses at 100 ms intervals: blood volume pulse (BVP), electromyography over the left-side corrugator (Corr), the output of a respiratory belt (RESP), skin temperature (Skin Temp) measured on the finger, and the output of a surface electromyography sensor on the trapezius (Trap), electromyography over the left-side Zygomaticus (Zygo).

5.2 Summary and Discussion

Regular time-stamps in these datasets allow for clear linkage between the listener response data and the corresponding audio. However, linking these data to note- or chord-level data would allow for empirical analysis of how listeners respond to specific musical events in musical time. Linking these data would also allow researchers to systematically analyze and interpret the listeners’ responses with respect to the organization of musical material (as derived from encoded expert analyses) and its performance (as available in the encoded performance data).

6. Conclusions and Next Steps

The datasets described above can facilitate empirical research in a variety of ways. The expert analyses can facilitate studies into musical structure in specific musical traditions. The performance data can help researchers model and understand how musicians craft musical materials in performance practices. And the listener response data can provide information about how listeners respond to localized musical events. Currently, however, researchers studying music analysis, performance, and listener responses are siloed both epistemologically and with regards to data practices. Thus, research into these types of questions does not yield integrative theories about how humans engage with music or produce data that can be used to work towards such theories. Some promising work in this direction is being done in the Towards Richer Online Music Public-domain Archives (TROMPA) project (Weigl et al., 2019a), but there remains a wealth of carefully curated research data that is not being utilized to its fullest potential.

Unified and/or linked encoding of these disparate data can facilitate the exploration of new research questions, but to do so would require addressing several issues related to how best to represent such data. For expert analyses, there are open questions regarding what are the most useful hierarchical levels at which to perform similarity analyses or retrieval tasks. The encoding of chord-level annotations in a format that also allows for simultaneous encoding of hierarchical analyses, such as the one used in the TROMPA project, can help researchers engage with musical materials in performance practices. And the listener response data can provide information about how listeners respond to specific musical events in musical time. Linking these data to note- or chord-level data would allow for empirical analysis of how listeners respond to specific musical events in musical time. Linking these data would also allow researchers to systematically analyze and interpret the listeners’ responses with respect to the organization of musical material (as derived from encoded expert analyses) and its performance (as available in the encoded performance data).

![Figure 5](image_url)

**Figure 5**: The left side of the figure (a) shows an example encoding of arousal and valence listener response data from Aljanaki et al. (2017) and the right side of the figure (b) shows an example encoding of physiological responses from Upham (2018).
as by Rizo and Marsden (2019), can provide enough flexibility for different representations of the musical material to be generated. This is ideal but often much more time-consuming than encoding only what is needed for the current research project. Tools are needed that assist researchers in encoding their annotations with rich music data.

Integration of disparate data does not require developing new encoding formats. The ones currently available, particularly the easily extensible ones like Humdrum and MEI, are more than sufficient for these tasks. However, there are challenges that need to be addressed when converting data between encoding formats, as demonstrated by Nápoles et al. (2018). Baraté et al. (2019) argue that adhering to the IEEE 1599 Format standards is a way to facilitate interoperability between datasets, but there is a not-insignificant amount of overhead required to comply with this which may hinder the widespread adoption necessary for this to be useful. Better tools are also needed for integrating complex symbolic information, such as notes or lead-sheet chords, with performer and listener data. Likely, a large part of the motivation for the bespoke annotation systems described above is the complexity of integrating the annotations with the full score data, even when it is available. The MELD framework is a promising approach for integrating not just annotations with encoding data but also audio and other relevant documents (Page et al., 2019).

Consistent and complete encoding of these data with musical objects could also benefit other MIR tasks. For example, automatic music transcription and alignment tasks could benefit from better models of performance and musical structure. Thus, beyond the primary benefit of facilitating a deeper understanding of how humans engage with music, there are also computational benefits to linking music-related data at the level of musical objects. This can be achieved by data creators by using a standard encoding format like MEI or Humdrum when building a new dataset and linking audio to symbolic representations where possible. For those interested in developing tools, it would be useful to build both pipelines that robustly convert from bespoke or proprietary formats to standard ones and software that can assist researchers in encoding their data in accessible and sharable ways.

Note

Acknowledgements
The author would like to thank the members of the music encoding community who gave her valuable feedback when she presented much of the content of this paper in a keynote at the 2019 Music Encoding conference. The author would also like to thank the TISMIR reviewers and editors for their insightful and thoughtful feedback during the review process.

Competing Interests
The author has no competing interests to declare.

References


