Notes on the Music

A social data infrastructure for music annotation

David M. Weigl
Werner Goebl
{lastname}@mdw.ac.at
University of Music and Performing Arts Vienna
Vienna, Austria

David J. Baker
Tim Crawford
Federico Zubani
david.baker@gold.ac.uk
tim.crawford@gold.ac.uk
fzuba001@gold.ac.uk
Goldsmiths, University of London
London, United Kingdom

Aggelos Gkiokas
Nicolas F. Gutierrez
Alastair Porter
Patricia Santos
{firstname.lastname}@upf.edu
Universitat Pompeu Fabra
Barcelona, Spain

ABSTRACT

Beside transmitting musical meaning from composer to reader, symbolic music notation affords the dynamic addition of layers of information by annotation. This allows music scores to serve as rudimentary communication frameworks. Music encodings bring these affordances into the digital realm; though annotations may be represented as digital pen-strokes upon a score image, they must be captured using machine-interpretable semantics to fully benefit from this transformation. This is challenging, as annotators' requirements are heterogeneous, varying both across different types of user (e.g., musician, scholar) and within these groups, depending on the specific use-case. A hypothetical all-encompassing tool catering to every conceivable annotation type, even if it were possible to build, would vastly complicate user interaction. This additional complexity would significantly increase cognitive load and impair usability, particularly in dynamic real-time usage contexts, e.g., live annotation during music rehearsal or performance. To address this challenge, we present a social data infrastructure that facilitates the creation of use-case specific annotation toolkits. Its components include a selectable-score module that supports customisable click-and-drag selection of score elements (e.g., notes, measures, directives); the Web Annotations data model, extended to support the creation of custom, Web-addressable annotation types supporting the specification and (re-)use of annotation palettes; and the Music Encoding and Linked Data (MELD) Javascript client library, used to build interfaces that map annotation types to rendering and interaction handlers. We have extended MELD to support the Solid platform for social Linked Data, allowing annotations to be privately stored in user-controlled Personal Online Datastores (Pods), or selectively shared or published. To demonstrate the feasibility of our proposed approach, we present annotation interfaces employing the outlined infrastructure in three distinct use-cases: scholarly communication; music rehearsal; and rating during music listening.

DLfM2021, July 28–30, 2021, Virtual Conference, GA, USA

ACM ISBN 978-1-4503-8429-2/21/07.
https://doi.org/10.1145/3469013.3469017

CCS CONCEPTS

• **Applied computing** → **Annotation**; • **Information systems** → *Multimedia content creation.*

KEYWORDS

Annotation, Music Encoding, Linked Data, Digital Musicology

ACM Reference Format:
David M. Weigl, Werner Goebl, David J. Baker, Tim Crawford, Federico Zubani, Aggelos Gkiokas, Nicolas F. Gutierrez, Alastair Porter, and Patricia Santos. 2021. Notes on the Music: A social data infrastructure for music annotation. In *8th International Conference on Digital Libraries for Musicology (DLfM2021), July 28–30, 2021, Virtual Conference, GA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3469013.3469017

1 INTRODUCTION

Musical scores communicate information, serving varied uses as boundary objects, *"artifacts ... that help people from different communities build a shared understanding"* [31, pg.1890]. For performers, scores primarily prescribe the musical actions to be undertaken in order to perform a (part of a) musical composition. For scholars, scores provide a trace of the creative thoughts that resulted in either the creation or performance of a musical work. And for enthusiasts, scores provide a common touch point for engaging with a work through the use of non-domain-specific annotations, such as ratings and comments.

Though the use of scores is not limited to these three groups, considering the needs of each demonstrates the heterogeneity of ways an individual or group might interact with notated music. For example, while both performers and scholars effortlessly scribble down notes on their own scores, the reminders of an upcoming difficult page turn that needs to be marked in real-time during a rehearsal for the collaborative pianist are markedly different than a conductor contemplating how they will conduct the opening trumpet solo of a symphony before the first rehearsal. The former highlights annotations that need to be made in the immediate context of performance; those in the latter may be created in a calmer environment. Both are traces of musical thought that might be of interest to a musicologist.

Traditionally these annotations are made on physical sheet music, with a pen or pencil freely traversing the musical score to suit the needs of the moment. Though straightforward and functional in this traditional medium, retaining this ease of capture while incorporating the benefits of digital music scores presents a unique set of challenges and opportunities.

Scores are ubiquitous in the Western tradition, and individuals who digitally interact with a musical score do so expecting the same affordances they are used to with pen and paper. Performers need direct access to be able to make the markings they need, free of technological obstructions, so when a collaborator reminds them of their tendency to go flat at the modulation, this thought is easily externalized to the score without interfering with a rehearsal. Likewise, scholars need the ability to clearly mark points of analytical interest that help to build the desired narrative argument. Such annotations might highlight the augmentation of an important motive hidden in the orchestration between several instruments, or simply demarcate a formal structure in the music.

Working with digital scores opens up possibilities beyond those offered by traditional sheet music. For example, the ability to link one's annotations to an externally-available digital score enables sharing and publishing of musical insight, forming the basis of a community of knowledge [4]. These annotations may include semantic information that can further be extended and linked to other forms of accessible, digitized knowledge.

A primary affordance of music score is performance, and interactions between scores and multimedia performance recordings particularly benefit from digitisation of both types of material. The study of performance recordings increasingly plays a focal role in music scholarship [1, 12]; and they carry a particular priority in music information retrieval research, the majority of which is carried out on recorded, rather than notated, music. Here too, annotations have an important role to play; scholarly arguments are built on the basis of referenced (annotated) time intervals in specified recordings; musicians refer to their own and others' recordings during rehearsal, collaborative discussion, and self reflection; and, music information retrieval researchers rely on listener-provided annotations to act as 'ground-truth' reference data in the training and improvement of machine-learning models.

In this paper, we introduce a social data infrastructure for music annotation which seeks to serve the heterogeneity of needs of the many people who annotate music, while opening novel avenues for communicating information about music available with digital score encodings. The data infrastructure, detailed in Section 4, was developed in conjunction with the Towards Richer Online Music Public-domain Archives (TROMPA) project[1] [27], which aims to address music information needs of music scholars, performers, and enthusiasts. We provide an overview of prior work that has addressed the annotation requirements of these groups in Section 2 to contextualise specific requirements identified for the use-case scenarios within TROMPA in Section 3. We demonstrate the feasibility of our ideas through implemented prototype applications for each scenario in Section 5, before concluding in Section 6.

2 BACKGROUND

2.1 How do users annotate music?

2.1.1 Scholars. As most broadly documented by Inskip and Wiering [8], the needs and workflows of music scholars are variable and diverse. In a study of over 600 musicologists, Inskip and Wiering surveyed the needs of musicologists and noted: *"a common set of disciplinary values seem to emerge from the responses, emphasising qualities such as completeness, depth of analysis, accuracy, reliability, serendipity and the materiality of resources."* [8, pg.460] Despite these common values, an annotation infrastructure able to encompass the diverse areas of musicological study needs to handle not only different forms of media including images, sounds, and text, but must also handle the various ways that music can be notated.

Recent projects have developed bespoke interfaces collating the disparate forms of musical material scholars require to construct narratives for specific musicological scenarios [14]. For example, Dreyfus et al. [13] present a digitally encoded musicological narrative simultaneously combining video, text, a musical score, score visualisations, and audio recordings.

Similarly, music analysts and theorists employ workflows incorporating detailed annotations of a musical score alongside extra-musical material [9, 11, 26]. In these examples, the analyst publishes a text augmented by various types of media that are pulled from external sources into a unified (static) page of supplementary material. Further, in all the music theoretic examples (but unlike in the multimedia digital musicological narrative), there is no predefined semantic network for re-use of the common types of media employed in these contexts. Media is put together *ad hoc*, with no toolkit to recycle for similar projects.

While the above examples combine various forms of media with traditional Western scores, the diverse needs of musicologists extend beyond the traditional common practice era of music composition and need to be met by corresponding tooling for these domains. For example, scholars may need to work with early music tablature [29] which might somewhat resemble modern notation, but demands a quite different encoding to be represented digitally. Similarly, some musicological work requires the use of real-time, multimodal data capture as documented in [8] and [19].

2.1.2 Performers. Investigating performers' interactions with scores in the context of group rehearsal, Winget presents an extensive study of annotations made by musicians of varying skill levels and performance modes [31]. This study demonstrates the enduring nature of performers' rehearsal annotations generated within such a setting. Far from fleeting ephemera associated with a particular moment in a rehearsal session, such annotations are expected by their creators to have a persistent, social life: almost all participants said they would reuse their own annotations in the future, given the opportunity; most thought that other musicians would find their annotations useful; and almost all indicated a desire to gain insight into the annotations of experts – for varied purposes, including study, self-reflection, and interpretive stimulus.

Winget also demonstrates the varied range of annotation types employed by rehearsing performers. Where interactions are confined to printed scores, this variety poses no challenge to the traditional pen-and-paper interface; but (as is the case with scholars), it becomes prohibitive if annotations are to be digitally captured with semantic precision (beyond recording digital pen-strokes), as a digital interface that exposes menu items for each possible type of annotation would be so complex as to be unusable, even if it were possible to build at all.

[1] https://trompamusic.eu

Lewis et al. describe the trade-offs involved between precision and usability in building interfaces for score annotation [15]. They present the deployment of a custom-built interface for musicological live annotation during a string-quartet masterclass, capturing and reflecting the discussion of performative aspects between the musicians upon a digital score. The annotation activity took place *in-situ* with the performers upon a large screen, allowing the annotator's actions to figure into the ongoing discussion. The annotation types available within the interface for placement upon the score were assembled in anticipation of likely topics of discussion in this specific masterclass, comprising: bowing markers (*upbow* / *downbow*), a palette of fingering indications (numbers *1–5*), a palette of dynamic markings (*pp–ff*), hairpin markers for *crescendo* and *diminuendo*, and a *phrase* marker. Each annotation was captured with a timestamp, and an audiovisual recording of the masterclass was aligned with the clock used by the annotation interface. This enabled the captured annotations to serve as a semantic index into the recording of the discussion.

On review, the annotator noted that, though the foreseen annotation types all duly played a role in the discussion, long stretches of the event remained unannotated. Rather than due to an emptiness in conversation, this was because the discussion frequently strayed into relevant but varied topics that were not foreseen, including "*sections of music being brought to the attention of the players as an important point of resolution or change;"* and *"directions ... phrased in non-technical terms (e.g., 'jazzy' or 'workmanlike')"* [15, pg.5]. Modifying the annotation interface by simply adding each additional topic to the palette of annotations would further overfit the tool to the particular occurrence of the event – another masterclass would result in another set of relevant but non-recorded topics. Instead, Lewis et al. decided to add a more general set of annotation types to their tool: a *cursor*, allowing the specification (annotation) of a particular measure as the current topic of discussion without capturing the specifics of meaning, and an *important* modifier to the cursor, to prioritise the given measure at the given recorded instant as being particularly important to the discussion. A re-annotation effort undertaken by the same annotator on the video recording of the event yielded more than twice the number of annotations, demonstrating the applicability of the cursor in offering greater flexibility at the expense of more rigid *semantic precision.*

2.1.3 Enthusiasts. Unlike scholars and performers, music enthusiasts' modes of interacting with musical information do not necessarily depend on the use of a musical score. For example, previous work has investigated users of music discussion websites such as Genius [7], which provide online infrastructures that allows users to comment and interact with musically meaningful information. In the case of Genius, a website that allows users to "annotate and explain the meaning of segments of lyrics in music and other written works", Fields and Rhodes [7] were able to model community interest and expertise through a network of regular users of the website, providing the foundation of new conceptualisations of notions of genre and of categorical segmentation in musical communities. Previous work led by Fields [6] utilized a similar approach to demonstrate how annotations of musical material can also be used in recommendation system contexts, leading to meaningful measures of musical similarity. These examples demonstrate ecological use cases where researchers were interested in how spontaneously generated annotations can lead to new insights about musical behavior.

2.2 Reinterpretation in other contexts

As noted in Section 1, annotated scores serve varied uses to different groups of people. The annotations of notable players, conductors, and composers first serve these annotator's ongoing rehearsal, performance and creative practice, but later provide the foundation for archival musicological research. Similarly, the masterclass annotation exercise by Lewis et al. first helps to specify an on-going discussion between musicians, before subsequently providing a corpus for scholarly analysis.

Music enthusiasts have also been used to help generate specific meaningful annotations for scholarly work. With more data collection opportunities made online, many researchers have turned to "citizen science" models of data collection where members of the general public volunteer their time and ability to help sort, classify, and identify various forms of data. This approach has been used in the sciences [24] and also adopted in more humanistic contexts such as in the mass validation and editing of optical music recognition outcomes [23]. Aside from digital musicology scholarship, such data also proves valuable to music information retrieval researchers as reference ('ground-truth') data used to train and improve machine-learning models [17]. In each of these cases, online infrastructures allow the pooling of human resources, supporting large-scale, data-intensive projects.

3 USE-CASE SCENARIOS

Having outlined prior work on the annotation behaviours of music scholars, performers, and enthusiasts in Section 2, we now outline requirements gathered for applications addressing specific use-case scenarios for these groups developed in the TROMPA project [27].

3.1 Music scholars

As discussed in Section 2.1.1, music scholars require the ability to incorporate various forms of multimedia material into their musicological workflows. Inskip and Wiering explicitly note: *"software should support online access to high quality digital resources (image, text, sound) which are comprehensive and discoverable, and can be shared, reused and manipulated at a micro and macro level."* [8, pg.460]

To address this established need, we sought to develop a prototype application that allowed musicologists to incorporate, annotate, and share music scores, images, texts, and audiovisual recordings in conducting and disseminating scholarly work. In collaboration with prominent Gustav Mahler scholar Paul Banks, we decided to focus on historical musicological work surrounding Mahler's Fourth Symphony.

Banks provided us with several annotations based on his investigations into the various sources for the opening of the symphony. In addition to comments regarding Mahler's idiosyncratic use of performance directives in the score, we further supplemented these textual annotations with several audio and video recordings of the opening passage, as well as photographs of the full score which was marked up in great detail by the Dutch conductor Willem Mengelberg while working closely with Mahler on a visit to conduct the work in Amsterdam in 1904. In this context, the score and the performance directives embedded within it serve as an initial point of interest for a Mahler scholar. Other users (who might also be professional or amateur musicologists, or even music-loving members of the public) could then offer their own replies to Banks' comments, or introduce their own ideas. We set out to implement an interface permitting such interactions with the material, in which audio clips linked to the directives would provide a chance for other users to hear for themselves how various historical conductors (including Mengelberg) have responded to the tempo and expression markings in the printed and hand-annotated scores, permitting fuller participation of users who may lack score-reading expertise. A prototype implementation of our interface is presented in Section 5.1.

3.2 Music performers

As noted by Winget [31], annotated scores are of enduring, persistent use to music performers, who build on, refer back to, and share them over a long-term period of rehearsal and performance.

In the digital realm, encoded scores can provide further uses to the rehearsal process: aside from conveying the musical intent of the composer, and providing a framework for annotations, they can be interlinked with the timelines of rehearsal or performance recordings through manual alignment or using automated score-performance alignment algorithms [2, 18]. This allows annotated score encodings to function as semantic indices into timed (e.g., audiovisual) media that are otherwise semantically undifferentiated to a machine agent [15]. Such indices enable users of digital interfaces to precisely navigate recorded media via the score; and further, it enables the visualisation of performance features obtained from such recordings using music information retrieval techniques – e.g., pertaining to tempo, dynamics, or performance error.

The TROMPA project set out to develop a digital rehearsal companion offering such capabilities, focusing on a target audience of highly skilled (amateur and professional) solo piano performers. Requirements were gathered through a small pilot study employing wireframe mock-up interfaces, and through feedback obtained at a workshop on music information user studies. An initial implemented prototype was evaluated in a series of structured interviews into the rehearsal process of 10 piano major students at the University of Music and Performing Arts Vienna, informing further development of the system described in Section 5.2.

Though employing a much smaller set of participants, these interviews pertaining to a solo rehearsal context nevertheless reveal interesting complements to Winget's results for score annotation in a group setting. While all but one participants described making annotations upon printed scores as a regular component of rehearsal practice, only two participants reported using annotations more complex than circling notes, inserting simple dynamic markings or other directives, or writing in fingerings; these participants placed a much stronger emphasis on annotations in rehearsal practice, incorporating their own sets of symbols as well as elaborate free-text directives, and tying meanings to different colours, both to serve in self-reflection and to provide memory cues during performance.

Most participants saw value in being able to share annotations, six mentioning pedagogical contexts (sharing annotations with their teacher or their own students), three seeing value in the prospect of sharing digital annotations with others outside of this setting – importing the annotations of expert pianists for review, or sharing with other musicians outside of the solo piano context (e.g., in chamber music). Particular value was placed on the prospect of digital information sharing in the current global situation – all but one of the participants who saw value in annotation sharing made direct mention of the potential usefulness of corresponding digital solutions in context of the Covid-19 pandemic.

3.3 Music enthusiasts

One aim of the TROMPA project is to allow enthusiast listeners to easily share insights with researchers as "citizen scientists" [25]. Though music notation literacy is not ubiquitous among such users, their insights and perspectives provide important objects of study, both as sources in reception studies and music scholarship (Section 2.1.3), and as 'ground-truth' reference data used to improve machine-learning models in music information retrieval. For example, features approximating the emotional connotations of music are valuable for music indexing and recommendation, and thus of interest in the development of music information systems catering to music enthusiasts [5, 10, 32]. Other seemingly more objective aspects of recorded music, such as the placement of beats, may be approximated by beat annotation algorithms [3]; but their success in matching human beat perception is grounded in the reference data used to train the machine model. Iterative cycles of validation and correction of such outputs incorporating human listeners' perspectives may be used to enrich and extend this data, and thus ultimately improve the performance of the model.

4 A SOCIAL DATA INFRASTRUCTURE FOR MUSIC ANNOTATION

We now detail an infrastructure for music annotation implemented to accommodate the diverse needs and use-cases served by the TROMPA project, as identified in Section 3. To demonstrate the feasibility of the proposed solution, we detail different annotation interfaces implemented using this infrastructure in Section 5.

4.1 Data model

The Web Annotation Data Model[2] serves as the generic core representation of annotation data within the TROMPA application database. We have applied and extended this model to allow annotation types to be specified, grouped together, and shared for re-use, supporting the creation of data-driven annotation interfaces that adapt to specifications of particular use cases. Key components of Web Annotations include: a unique *identifier* (URI) representing the annotation; one or more *targets* referencing the (URIs of) web resources or resource fragments being annotated (e.g., elements of encoded score; timed offsets of music recordings); zero or more *bodies* containing information the annotation associates with its target(s), e.g., tags, descriptions, ratings; and, zero or more *motivations* describing the purpose of the annotation (termed *annotation*

[2]https://www.w3.org/TR/annotation-model

type in the preceding sections). Thirteen generic motivations are specified (e.g., oa:describing; oa:linking), and these may be further specialised with custom motivations using the SKOS[3] vocabulary. In addition, we introduced concepts from the schema.org[4] model to support "meta"-definition of annotations. This includes the following items:

4.1.1 Closed vocabularies. Annotations may apply closed vocabulary descriptors, e.g., a set of emotion labels to apply to a music recording (happy; sad; angry; exciting). schema:DefinedTermSet is used to define such vocabularies. Constituent terms, modelled using schema:DefinedTerm, can be referenced as an annotation body, e.g., within annotations with the oa:classifying motivation.

4.1.2 Rating definitions. The oa:assessing motivation is used to assign a numerical rating to an item. These are encoded using schema:Rating, specifying the minimum and maximum values of the rating scale alongside the rating value. To accommodate aggregations of ratings of an item (e.g., different users' perceived difficulty of performance of a given work), we introduce the custom concept of cust:RatingDefinition; these are abstract schema:Rating templates with set maximum and minimum values, but without a rating value. Individual users' schema:Ratings are then linked to this definition using the prov:wasDerivedFrom property of the PROV Ontology[5]. Beside the aggregation of ratings, these templates also inform software agents (e.g., Web clients) rendering scales in graphical annotation interfaces.

4.1.3 Annotation palettes. Some annotation use cases are best served by grouping together related functionalities within annotation palettes (e.g., [15]). In our model these are represented as schema:DefinedTermSets containing one schema:DefinedTerm element corresponding to each item in the palette. Web clients presenting generic annotation interfaces may then be directed to incorporate given annotations palettes specified by corresponding schema:DefinedTermSet URIs.

4.1.4 Annotation toolkits. Annotation toolkits, schema:ItemLists containing references to the annotation components described above or to further (constituent) annotation toolkits, are provided as a further abstraction in the specification of annotation user interfaces. Client applications load such specifications in order to instantiate user interfaces, rendering appropriate graphical components in response to guide the user through the annotation process.

4.1.5 Annotation sessions. Annotations are not made in isolation; a single document may be referenced by multiple related annotations made by one or more people, and annotations may be generated at different times, in different contexts, or for different tasks. Annotation sessions logically group annotations into an aggregate object, allowing them to be displayed together, or to be referenced as a single item. This grouping is accomplished by reference to each annotation's URI, allowing annotations to be simultaneously incorporated within many different annotation sessions. These session objects are represented as ldp:Containers, as specified by the Web Annotation Protocol[6].

4.2 Public and private storage

The Solid [16] project applies a decentralised W3C standards-based Linked Data technology stack to enable online interactions between users that retain data ownership with each individual user. This permits each user fine-grained access control over their personal data, supports sharing of data with specified users, and simple integration by reference (URI) within external contexts, e.g., in digital libraries. This emphasis on Web standards, and on user choice and control of contributed data, provides a strong fit to research data regulations and best practices, including the EU General Data Protection Regulation (GDPR) and the FAIR guidelines for research data management [30]. Solid provides users with Personal Online Datastores (Solid Pods), which act as WebID[7] identity providers as well as user-controlled storage spaces. This allows users to log in to TROMPA's Web applications (alongside any other Solid-compatible applications) with their own WebID, and to privately share data with other users by reference to their WebIDs. The TROMPA project's Pod provider (Solid server) is available online. Due to the decentralised nature of Solid, users are free to choose this or any other Pod provider on the Web for interaction with our applications. Users with advanced technical expertise may additionally choose to self-host. We have developed a library of annotation tools capable of interacting both with Solid Pods and with the TROMPA application database. his means that annotations which are incomplete or which the user wishes to keep private can be stored with restricted access permissions within a Solid Pod. We also provide functionality to "publish" annotations by granting public read access and copying them to the TROMPA application database, either by reference to the annotation's URI within the Solid Pod, or by ingesting the contents of the annotation to the database alongside an oa:canonical reference to the original URI [28].

4.3 Client interface components

To accommodate the varied requirements of working with digital scores, we have implemented a selectable-score [8] Javascript (React) component wrapping the Music Encoding and Linked Data (MELD) client library [14] to abstract more complex functionalities of displaying, selecting, and annotating elements of music notation in an easily implemented API. MELD, also used to implement the masterclass annotation tooling [15] and the digitally encoded musicological narrative [13] discussed in Section 2, augments the Verovio MEI engraver's [20, 21] affordances for interaction with music scores with a traversal engine for Linked Data and a set of handler functions capable of incorporating such data within multimedia user interfaces. We have contributed extensions to the MELD client library to transparently handle WebID authentication and reading and writing of data within Solid Pods.

[3]http://www.w3.org/2004/02/skos/core#
[4]https://schema.org
[5]https://www.w3.org/TR/prov-o/
[6]https://www.w3.org/TR/annotation-protocol#annotation-containers
[7]https://solid.github.io/specification/protocol#webid
[8]https://github.com/trompamusic/selectable-score

Our annotation library[9] contains reusable software components allowing the construction of UIs offering annotation input and interaction: *score annotations* are authored using the selectable score component discussed above; *audio files* are annotated using a component that renders a representation of a waveform, permits playback, and allows the graphical selection of timed instants or intervals to serve as annotation targets; a component may be used to *view all annotations* in a session; a further component is used to load annotation toolkits and render each constituent; *annotation palette* are rendered and made interactive (allowing user selection), showing either textual labels or images associated with each item in the palette; a component for *closed vocabularies* is used to show constituent terms and support user selection; a component for *ratings* is used to show a graphical representation of a scale between the relevant minimum and maximum values; finally, text box components support *free-text input* for other textual annotation types.

5 FEASIBILITY DEMONSTRATION

The data infrastructure presented in Section 4 flexibly addresses diverse music annotation needs arising in different use cases and among different user groups, while adhering to regulations and best practices around research data. To demonstrate the feasibility of the proposed infrastructure, we now present prototype applications using it to fulfill the requirements outlined in Section 3. Beside serving the purposes of their intended primary audience, their shared technological backing also allows annotations generated with these systems to be reused and reinterpreted in other scholarly and analytical contexts, along the lines discussed in Section 2.2.

5.1 Scholars annotator tool

Our digital score annotation tool prototype[10] addresses the requirements of the scholarly use case outlined in Section 3.1. Figure 1 illustrates the annotation interface of this prototype. To generate annotations and participate in scholarly discourse using the system, users must log in using their WebID (Figure 1, *A*) and select an `ldp:Container` within their associated Solid Pod (*B*; see also Section 4.2). The contents of this container form an annotation session (i.e., a collection of annotations available for interaction within the application), and may include references (by URI) to content generated by the user within their own Solid Pod, as well as references to shared content housed within other users' Pods.

Users specify a combination of 'note', 'measure', 'dynamic', and 'directive' selectors to generate new annotations upon the score (*C*); these are used to configure the behaviour of the underlying selectable-score component (Section 4.3) in enabling the corresponding types of MEI elements to be selected. It should be noted that this selection mechanism is not limited to the CMN music scores relevant in the narrow context of this use case, but is widely applicable to different types of music notation by virtue of the comprehensive capture of music semantics provided by the MEI schema, and the wide-ranging capabilities of the underlying Verovio engraving tool. The interface could thus be easily adapted for use on different types of musical material, such as lute tablature [29].

The interface presents an annotation palette exposed as radio buttons serving as motivation selectors (*D*). The deployed palette includes *describing* (`oa:describing`) for textual annotations; *linking* (`oa:linking`) for the specification of hyperlinks to web resources; *cue media*, a custom specialisation of `oa:linking` addressing timed intervals within audiovisual recordings; *image*, another custom specialisation of `oa:linking` used to reference image resources to be rendered; and *reply* (`oa:replying`), used to reply to the annotations previously generated by the user or by other users. The input field (*E*) dynamically adapts to the selected input type in order to allow the user to specify appropriate annotation bodies.

Annotation bodies of currently relevant annotations (see below) are displayed as a scrolling list (*F*), using MELD (Section 4.3) to map annotations types to corresponding rendering and interaction handlers and to display the content. Where the user has the necessary authorisation (e.g., for their own annotations), sharing permissions may be set to grant or retract public access to the given annotation; additional metadata pertaining to annotation provenance is available through a small information icon. Users may issue replies (i.e., generate new `oa:replying` annotations) in response to any listed annotation by clicking on the "reply" button, which causes the input control elements (*D*) and (*E*) to adapt accordingly.

The selectable score (*G*; see Section 4.3) is configured according to the selection type criteria specified in (*C*); layout modifications are easily accommodated (e.g. to change scale to 'zoom' in or out using a simple interface) while allowing the notated content to reflow across page boundaries. Previously annotated elements are highlighted at a measure level, a convenient grouping that strikes a compromise between *anchor precision* (per [15]) and ease of use. Measures containing a target of one or more annotations in the collection are highlighted using a transparent bounding rectangle (*I*). Clicking on such measures declares the associated annotations as currently relevant, signalled by modifying the selected measure's highlight to an outline (*H*). Relevant annotations are loaded into the annotation list (*F*); these annotations may be clicked to colour-in the corresponding target score elements, restoring fully granular *anchor precision* where this doesn't interfere with usability.

Though serving as a prototype catering to a specific use-case (Section 3.1), the underlying data infrastructure (Section 4) permits the same user environment to handle any MEI-encoded music score, and to be easily adapted to incorporate other annotation palettes, toolkits, and rendering and interaction handlers. Indeed, work is ongoing on an alternate deployment in the context of a score search engine developed in the related F-TEMPO project[11]. This search engine maps an input query comprising fragments of encoded music to matching (and partially matching) results from a large database of scanned score images with associated (OMR-generated) MEI encodings. Our alternate deployment of the interface presented here will allow inputs to be specified by graphical selection upon the rendered score, and query results to be saved, shared, and published as `oa:linking` annotations connecting the selected fragments with the resulting scores. This will for the first time provide a usable graphical interface wrapping F-TEMPO's query pipeline[12].

[9] https://github.com/trompamusic/trompa-annotation-component

[10] Code available at https://github.com/trompamusic/music-scholars-annotator; Application available at https://trompamusic.github.io/music-scholars-annotator/

[11] https://f-tempo.org

[12] A poster on this integration has been accepted for presentation at the 8th International Conference on Digital Libraries for Musicology (DLfM 2021).

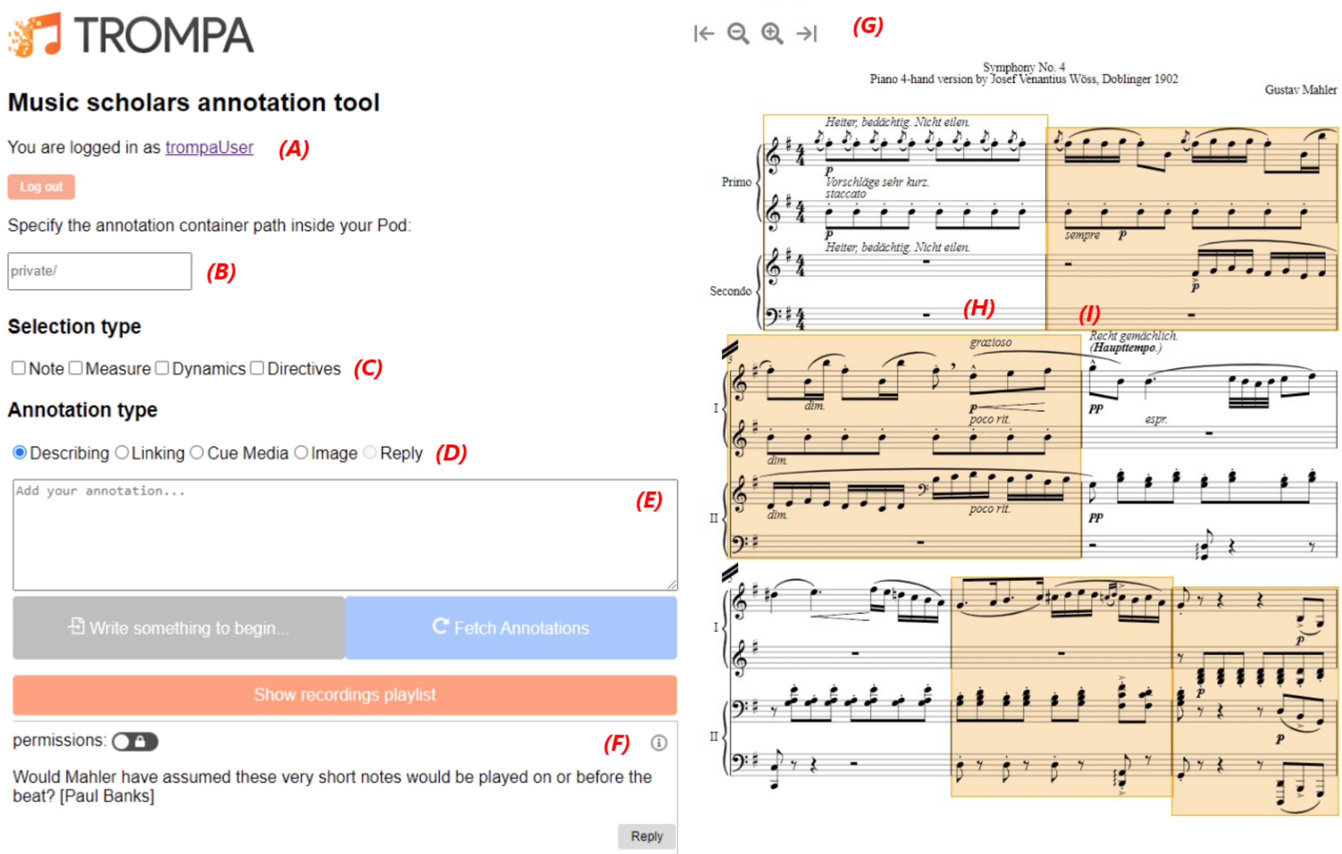

Figure 1: The music scholars annotation tool allows users to view, generate, and share multimedia annotations of digital scores.

5.2 Piano rehearsal companion

In setting out to implement a digital rehearsal companion meeting the user requirements outlined in Section 3.2, we noted that requirements for circling of score elements, annotation of fingerings, and insertion of simple free-form directives (e.g., “slow down”, “loose wrist”) were raised by all but one participant; and that participants universally identified lack of complexity and ease of use among the most important qualities a digital rehearsal tool should exhibit.

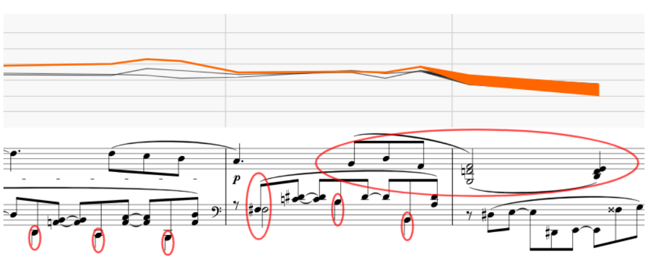

Figure 2: Score annotations within the rehearsal companion tool, linked to performance features (here: dynamic range in upper staff notes performed in three rehearsal recordings).

The annotation tooling deployed in the implemented prototype[13] is scaled back accordingly, allowing users to click-and-drag (or tap-and-swipe) to select notes, dynamic markings, or directives. A small annotation toolkit comprises a *circle* button used to generate `oa:highlighting` annotations, rendered as an ellipse enclosing the selected elements; a *fingering* palette used to generate annotations with the `oa:classifying` motivation, rendered by placing the number *1–5* from a corresponding `schema:DefinedTermSet` above the center of the selected region; and, a *text* button prompting the user for a textual description rendered in the same way, encoded using the `oa:describing` motivation type. In each case the generated annotations target the selected score elements, which are precisely aligned with timed instants within rehearsal recordings captured ‘live’ during rehearsal sessions[14]. This maximises both *anchor precision* and *associative precision* per Lewis et al. [15], allowing users to review their rehearsal progress – and inspect visualisations of performance features (e.g., tempo curves, dynamics, performance errors) – by means of the annotated score in near-real time (Figure 2). *Semantic precision* (per Lewis et al.) is deemphasised in favour

[13] Code available at https://github.com/trompamusic/clara/; Application available at https://trompa.mdw.ac.at/

[14] Rehearsal recordings are obtained using the Web-MIDI API in an automated workflow reconciling the output of the MIDI-alignment tool described in [18] with MEI identifiers. Code available at https://github.com/trompamusic/trompa-align

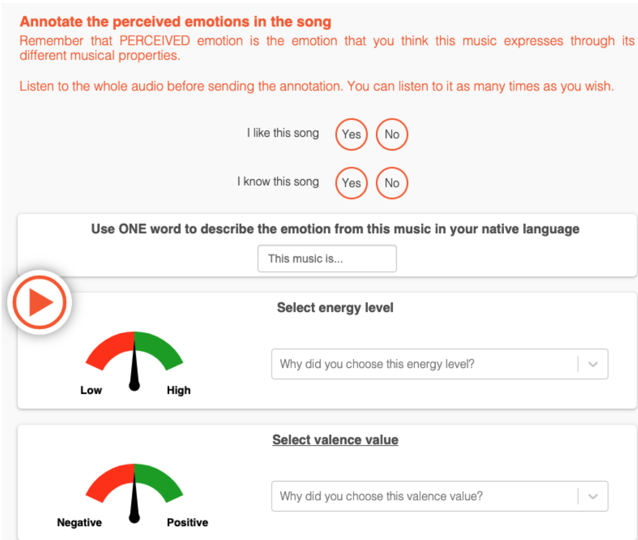

Figure 3: Emotion perception annotator interface.

of reduced interface complexity, with only the fingering annotations providing a high level of semantic precision. An alternative configuration providing more specified annotation palettes could be deployed, e.g., for the purpose of large-scale musicological analysis of rehearsal behaviour; though this would impose increasing limitations to usability due to interface complexity.

5.3 Listener annotation interfaces

We have built a number of interfaces to enable enthusiast listeners to contribute to music information retrieval research by sharing insights on their music perception (Section 3.3). As noted above, notation literacy should not be assumed in this target audience, and so these interfaces do not integrate the digital score tooling described in Section 4.3. The required functionalities can, however, be accommodated with the remaining components described in Section 4, demonstrating the flexibility of the proposed data infrastructure.

Figure 3 illustrates an interface used to gather music emotion annotations from listeners[15], incorporating an annotation toolkit with the following components: a free-text tag (motivated by `oa:tagging`) of the most salient emotion perceived within the excerpt; a rating (`oa:assessing`) of the perceived valence and arousal, used to identify one of the four quadrants of the circumplex (valence/arousal) model [22]; a closed-vocabulary (`oa:classifying`) of emotion types within a `schema:DefinedTermSet` corresponding to the specified quadrant; descriptions of the users' reasons for their choices (`oa:describing` annotations that either reference a predefined `schema:DefinedTerm` or allow a custom response), and finally, a binary indication (`oa:classifying`) of the annotator's familiarity with and liking or disliking of the excerpt. The annotations encoding the user's response to each of these components are aggregated within a custom higher-level annotation (specialising `oa:linking`), allowing a set of responses to be referenced as a coherent object.

Factors pertaining to the annotator, rather than the excerpt – including spoken languages, cultural background (approximated by birth place), and mood – are also collected to study whether and how such factors impact on users' behaviour in the task. While the (anonymised) responses pertaining to music excerpts are stored in the TROMPA application database, these personal data items are instead stored privately within each user's Solid Pod (Section 4.2), empowering them to grant or revoke access to this information.

Another example illustrating this approach is a semi-automated beat annotator tool comprising two components: a front-end interface for annotating beats on a graphical waveform, and a back-end algorithm capable of accepting user input in determining beat locations in an audio signal. A hybrid workflow combines human understanding of rhythmic structure, and the microtiming accuracy of computational methods. The workflow begins by the automatic estimation of candidate beat locations by the algorithm. These beat locations are visualised by the front-end component and can be (partially) edited/corrected by the user. Corrections are submitted in aggregate (i.e., as a response encoding all corrected beat locations for a given signal), referenced by an annotation with a custom motivation specialising `oa:assessing`. In response, the algorithm takes into consideration the corrections to re-estimate the beats. This loop can be repeated several times until the user approves the final annotations, which may then serve as 'ground truth' for down-stream music information retrieval research.

6 CONCLUSION

We have presented a social data infrastructure for music annotation capable of representing and handling annotation structures corresponding to varied annotation requirements. Based in social Linked Data and Web standards, the infrastructure allows meta-specifications of annotation toolkits, palettes, and motivations, as well as annotations themselves, to be shared, reused, and reinterpreted by reference to persistent, universal identifiers (URIs). We emphasise user-retained control over contributions, providing a strong fit to research data regulations and best practices. This infrastructure is motivated in diverse requirements gathered for different use-cases in the context of the TROMPA project. The feasibility of our proposed solution is demonstrated through implemented music annotation prototypes targeting music scholars, performers, and enthusiast listeners. We note that no claims are made here regarding the final *usability* of the interfaces at this stage, though their design is informed by requirements gathered in user research.

A motivating factor in our design lies in the ability to specify generic interface components that are easily adaptable for reuse in different contexts. While the interface designs presented here cater to the requirements of TROMPA's use-cases, the adaptation of the scholars' annotator as a graphical interface for a score search engine (Section 5.1) demonstrate the flexibility for adaptation of our components. We anticipate further concrete examples of such adaption to arise in future projects, and hope to have provided a useful contribution to facilitate interface development and improve research data management for annotation tasks in digital musicology research.

[15] Code available at https://github.com/trompamusic/music-enthusiast-front-end; Application available at https://ilde.upf.edu/trompa/

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme *H2020-EU.3.6.3.1. - Study European heritage, memory, identity, integration and cultural interaction and translation, including its representations in cultural and scientific collections, archives and museums, to better inform and understand the present by richer interpretations of the past* under grant agreement No. 770376. We gratefully acknowledge financial support by the Open Access Fund of the mdw - University of Music and Performing Arts Vienna, and the collaboration of our colleagues in the TROMPA consortium. Many thanks to Paul Banks for kindly sharing his insights on Mahler's Fourth Symphony in the context of our music scholars' annotation use-case; and, to all TROMPA user study participants for their contributions to our research.

REFERENCES

[1] Nicholas Cook. 2010. The ghost in the machine: Towards a musicology of recordings. *Musicae Scientiae* 14, 2 (2010), 3–21. https://doi.org/10.1177/102986491001400201

[2] Johanna Devaney. 2020. Using Note-Level Music Encodings to Facilitate Interdisciplinary Research on Human Engagement with Music. *Transactions of the International Society for Music Information Retrieval* 3, 1 (2020). http://doi.org/10.5334/tismir.56

[3] Jonathan Driedger, Hendrik Schreiber, W Bas de Haas, and Meinard Müller. 2019. Towards Automatically Correcting Tapped Beat Annotations for Music Recordings.. In *Proceedings of the 20th International Society for Music Information Retrieval Conference*. 200–207. https://archives.ismir.net/ismir2019/paper/000022.pdf

[4] Timothy Duguid, Maristella Feustle, Francesca Giannetti, and Elizabeth Grumbach. 2019. Music Scholarship Online (MuSO): a research environment for a more democratic digital musicology. *Digital Humanities Quarterly* 13, 1 (2019), 1 online resource (8 pages) : illustrations.

[5] Tuomas Eerola and Jonna K Vuoskoski. 2011. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music* 39, 1 (2011), 18–49. https://doi.org/10.1177/0305735610362821

[6] Ben Fields, Michael A Casey, Kurt Jacobson, Mark B Sandler, et al. 2008. Do you sound like your friends? Exploring artist similarity via artist social network relationships and audio signal processing. In *Proceedings of the International Computer Music Conference*.

[7] Ben Fields and Christophe Rhodes. 2016. Listen To Me – Don't Listen To Me: What Communities of Critics Tell Us About Music. In *Proceedings of the 17th International Society for Music Information Retrieval Conference*. 199–205. https://archives.ismir.net/ismir2016/paper/000173.pdf

[8] Charles Inskip and Frans Wiering. 2015. In their own words: Using text analysis to identify musicologists' attitudes towards technology. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*. 455–461. https://archives.ismir.net/ismir2015/paper/000171.pdf

[9] Marianne Kielian-Gilbert. 2020. Experiencing Chen Yi's Music: Local and Cosmopolitan Reciprocities in Ning for Pipa, Violin and Cello (2002). *Music Theory Online* 26, 3 (2020).

[10] Youngmoo E Kim, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull. 2010. Music emotion recognition: A state of the art review. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, Vol. 86. 937–952. http://archives.ismir.net/ismir2010/paper/000045.pdf

[11] Megan L Lavengood. 2020. The Cultural Significance of Timbre Analysis: A Case Study in 1980s Pop Music, Texture, and Narrative. *Music Theory Online* 26, 3 (2020).

[12] Daniel Leech-Wilkinson. 2009. *The Changing Sound of Music: Approaches to Studying Recorded Musical Performance*. CHARM. https://www.charm.rhul.ac.uk/studies/chapters/intro.html

[13] David Lewis, Laurence Dreyfus, and Kevin R. Page. 2021. Narratives and exploration in a musicology app: Supporting scholarly argument with the Lohengrin TimeMachine. In *Proceedings of the 7th International Conference on Digital Libraries for Musicology*. Accepted.

[14] David Lewis, David M Weigl, Joanna Bullivant, and Kevin R Page. 2018. Publishing musicology using multimedia digital libraries: creating interactive articles through a framework for linked data and MEI. In *Proceedings of the 5th International Conference on Digital Libraries for Musicology*. 21–25. https://doi.org/10.1145/3273024.3273038

[15] David Lewis, David M. Weigl, and Kevin R. Page. 2019. Musicological Observations During Rehearsal and Performance: a Linked Data Digital Library for Annotations. In *Proceedings of the 6th International Conference on Digital Libraries for Musicology*. 1–8. https://doi.org/10.1145/3358664.3358669

[16] Essam Mansour, Andrei Vlad Sambra, Sandro Hawke, Maged Zereba, Sarven Capadisli, Abdurrahman Ghanem, Ashraf Aboulnaga, and Tim Berners-Lee. 2016. A demonstration of the Solid platform for social web applications. In *Proceedings of the 25th International Conference Companion on World Wide Web*. 223–226. https://doi.org/10.1145/2872518.2890529

[17] Meinard Müller. 2015. *Fundamentals of music processing: Audio, analysis, algorithms, applications*. Springer.

[18] Eita Nakamura, Kazuyoshi Yoshii, and Haruhiro Katayose. 2017. Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment. In *Proceedings of the 18th International Society for Music Information Retrieval Conference*. 347–353. http://archives.ismir.net/ismir2017/paper/000035.pdf

[19] Kevin Page, Terhi Nurmikko-Fuller, Carolin Rindfleisch, David M. Weigl, Richard Lewis, Laurence Dreyfus, and David De Roure. 2015. A toolkit for live annotation of opera performance: Experiences capturing Wagner's Ring Cycle. In *Proceedings of the International Society for Music Information Retrieval*. 211–217. http://archives.ismir.net/ismir2015/paper/000311.pdf

[20] Laurent Pugin. 2018. Interaction perspectives for music notation applications. In *Proceedings of the 1st International Workshop on Semantic Applications for Audio and Music*. 54–58. https://doi.org/10.1145/3243907.3243911

[21] Laurent Pugin, Rodolfo Zitellini, and Perry Roland. 2014. Verovio: A library for Engraving MEI Music Notation into SVG. In *Proceedings of the 15th International Society for Music Information Retrieval Conference*. 107–12. https://archives.ismir.net/ismir2014/paper/000221.pdf

[22] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[23] Ioannis Petros Samiotis, Sihang Qiu, Andrea Mauri, Cynthia CS Liem, Christoph Lofi, and Alessandro Bozzon. 2020. Microtask crowdsourcing for music score Transcriptions: an experiment with error detection. In *Proceedings of the 21st International Society for Music Information Retrieval Conference*. https://program.ismir2020.net/static/final_papers/257.pdf

[24] Jonathan Silvertown. 2009. A new dawn for citizen science. *Trends in ecology & evolution* 24, 9 (2009), 467–471.

[25] Jack Stilgoe. 2009. *Citizen Scientists: reconnecting science with civil society*. Demos London.

[26] Caitlyn Trevor. 2018. Animated performance:'Better' music means larger movements. *Music Theory Online* 24, 4 (2018).

[27] David M. Weigl, Werner Goebl, Tim Crawford, Aggelos Gkiokas, Nicolas F. Gutierrez, Alastair Porter, Patricia Santos, Casper Karreman, Ingmar Vroomen, Cynthia C. S. Liem, Álvaro Sarasúa, and Marcel van Tilburg. 2019. Interweaving and Enriching Digital Music Collections for Scholarship, Performance, and Enjoyment. In *6th International Conference on Digital Libraries for Musicology* (The Hague, Netherlands) *(DLfM '19)*. Association for Computing Machinery, New York, NY, USA, 84–88. https://doi.org/10.1145/3358664.3358666

[28] David M. Weigl, Werner Goebl, Alex Hofmann, Tim Crawford, Federico Zubani, Cynthia C. S. Liem, and Alastair Porter. 2020. Read/Write Digital Libraries for Musicology. In *7th International Conference on Digital Libraries for Musicology* (Montréal, QC, Canada) *(DLfM 2020)*. Association for Computing Machinery, New York, NY, USA, 48–52. https://doi.org/10.1145/3424911.3425519

[29] Frans Wiering, Tim Crawford, and David Lewis. 2005. Creating an XML vocabulary for encoding lute music. In *Humanities, Computers and Cultural Heritage: Proceedings of the 16th International Conference of the Association for History and Computing*. 279–287.

[30] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9. https://doi.org/10.1038/sdata.2016.18

[31] Megan A Winget. 2008. Annotations on musical scores by performing musicians: Collaborative models, interactive methods, and music digital library tool development. *Journal of the American Society for Information Science and Technology* 59, 12 (2008), 1878–1897.

[32] Yi-Hsuan Yang and Homer H Chen. 2011. *Music emotion recognition*. CRC Press.