A TOOLKIT FOR LIVE ANNOTATION OF OPERA PERFORMANCE: EXPERIENCES CAPTURING WAGNER'S *RING* CYCLE

Kevin R. Page[†], Terhi Nurmikko-Fuller[†], Carolin Rindfleisch[‡], David M. Weigl[†]
Richard Lewis[#], Laurence Dreyfus[‡], David De Roure[†]

[†]Oxford e-Research Centre
University of Oxford
United Kingdom
first.last@oerc.ox.ac.uk

[‡]Faculty of Music
University of Oxford
United Kingdom
first.last@music.ox.ac.uk

[#]Department of Computing
Goldsmiths, University of London
United Kingdom
first.last@gold.ac.uk

ABSTRACT

Performance of a musical work potentially provides a rich source of multimedia material for future investigation, both for musicologists' study of reception and perception, and in improvement of computational methods applied to its analysis. This is particularly true of music theatre, where a traditional recording cannot sufficiently capture the ephemeral phenomena unique to each staging. In this paper we introduce a toolkit developed with, and used by, a musicologist throughout a complete multi-day production of Richard Wagner's *Der Ring des Nibelungen*. The toolkit is centred on a tablet-based score interface through which the scholar makes notes on the scenic setting of the performance as it unfolds, supplemented by a variety of digital data gathered to structure and index the annotations. We report on our experience developing a system suitable for real-time use by the musicologist, structuring the data for reuse and further investigation using semantic web technologies, and of the practical challenges and compromises of fieldwork within a working theatre. Finally we consider the utility of our tooling from both a user perspective and through an initial quantitative investigation of the data gathered.

1. INTRODUCTION AND MOTIVATION

The performance of a fully staged opera is perhaps the richest form of production when considering the potential for a wide diversity of music information united around a single body of work. Its study provides both opportunity and challenges for gathering, organising, retrieving, and analysing data and artefacts from and about the event. Thanks to a willing partnership with the Birmingham Hippodrome and the Mariinsky Opera under the baton of Valery Gergiev, their performance of all four operas comprising Richard Wagner's *Der Ring des Nibelungen* (henceforth *Ring*) over five days in November 2014 presented a unique opportunity to develop and trial a musical performance annotation kit providing a structured frame of reference for interpreting collections of multimedia data.

In this paper we report on the design and implementation of the annotation software and supporting tools, which were co-designed with a musicologist to provide maximal utility when deployed for fieldwork in a working theatre. We begin by considering motivations from the fields of musicology and Music Information Retrieval (MIR).

1.1 Musicological motivation

In recent decades, methodological shifts such as a 'performative turn', widely affecting research in the Arts, Humanities and Social Sciences, and reception theory questioned musicology's traditional focus on the work as an idealised concept and on the written score. Instead, music is considered as a continuous cultural practice, couched within the respective contexts in which it is perceived, which attaches an increased value both to performance as a general concept or ritual as well as to specific performance events [6]. The individual realisation of a work in performance, especially in music theatre, differs significantly from the abstract aesthetic concept captured in the score: while the musical dimension may be treated with a high degree of 'faithfulness', scenic interpretation is created afresh in every new staging. Even in cases such as Wagner's music dramas, in which music and scenic events are coordinated down to the smallest detail, the degree to which his scenic instructions are followed varies considerably, and the reality of individual stagings goes far beyond the concept in the score. This raises the question of how a music-dramatic performance, as an ephemeral phenomenon, can be captured [14]. Analyses of recorded performances are almost as old as the respective technologies themselves [9]; but as the recording often assumes the status of an aesthetic text in the process, ephemeral phenomena are again overlooked [5]. As an audiovisual recording is neither an objective nor an exhaustive documentation, the investigation of new ways of capturing different kinds of performance data is a worthwhile undertaking. Live annotation of a performance helps to overcome the 'recording bias' by en-

abling researchers to document events and gather information which cannot be reified in audiovisual recordings. The method of documentation and the resulting record is moreover of a significantly different quality: live-annotation allows for a selective, focused and structured record-keeping, where different annotation schemes can be tailored to a specific research question, thus integrating documentation with on-the-fly analysis. While digital technologies ease gathering of this information, this comes at a scale greater than could be recorded 'by hand'. The ability to semantically structure gathered data for publishing and reuse, and to undertake computationally assisted analysis, provides further breadth to the study of performances.

1.2 MIR motivation

A second motivation is the utility of a well-described and structured multimedia dataset, annotated by an expert musicologist, and tooling to create such corpora, to inform, refine, and test MIR algorithms. A comprehensive data source could act as an authoritative ground truth for a variety of MIR tasks including: automated identification of musicological facets and melodic phrase recognition (e.g. leitmotif detection); tempo prediction and score following (based on page turns and annotations). It holds prospects for hypothesis-driven exploration of bio-sensed data measured from audience members, and for calibration of automated prediction of listener arousal from scores, and potentially that of musical expectation and mood.

2. RELATED WORK

The complementary nature of *performance studies* and *empirical musicology* (§6) has been noted by Cook [6]. Kershaw [14] discusses performances as "site-specific spectacles", reporting research that largely confirms theatre reception as extensively influenced by idiosyncratic observer perspectives. The recording as the most accurate capturing of the live performance has been commented on by Trezise [24], while Doğantan Dack [8] used video recordings for a performer-centred study of chamber music, although not extending to theatrical aspects and staging.

In Section 3.3 we capture the scenic elements of performance through annotation. The extensive ethnographic study of *musical annotations* carried out by Winget [25] illustrates existing precedent for this approach, though from the perspective of musicians marking a score rather than a musicologist annotating a live operatic performance. Our technique is strongly guided by established rehearsal practice for opera, where the scenic aspects and stage directions that constitute a new staging are captured in an annotated score. These 'scripts' are not usually published, and no other works capturing timings of specific features of a live performance are known to the authors.

An overview of *digital technologies* in performance studies by Marsden [15] contends that research successfully bridging musicology with the digital is found within the domain of music information retrieval, rather than musical or performance analysis. For exploratory analysis, Dolan *et al.* refer to Sonic Visualizer [3, 11], but are exclusive of staging, theatrics, and actor dynamics, the digital annotation of which has little prior work. Okumura *et al.* [18] modelled ways to capture deviations from strict interpretations of the score during a performance – a potential use case for our dataset. A model for acquiring content, description of data, and subsequent evaluation that complements our work is been outlined by Repetto and Serra [21].

Reflecting on corpora containing live performance and annotations, Bainbridge *et al.* [1] list The Hathitrust Digital Library [4], and the International Music Score Library Project (IMSLP) [16], as examples of large-scale digital libraries for or including music and music-related data; Doerr *et al.* comment on the role of metadata for digital library resource retrieval [10]; cross-cultural approaches and models for resource discovery in music digital libraries have been examined by Hu *et al.* [13] and Porter *et al.* [20] respectively; and Smith *et al.* [23] designed and implemented a large database for structural annotation. These inform our ontological structures (Section 3.2).

3. DESIGN AND IMPLEMENTATION

The Musical Score Annotation Kit – 'MuSAK' [1] – was assembled from off-the-shelf hardware and applications combined with additional bespoke software, for recording the ephemera of live performance, as motivated in Section 1.1. It was designed to three primary requirements: *(i)* an interface sufficiently intuitive and fast enough to operate so that the musicologist could annotate under the pressure of a live performance, including turning pages to match activity on stage; *(ii)* for reconfigurability to incorporate changing annotation techniques and structures developed in the course of preparatory study prior to the performance events; and *(iii)* to be adaptable to the uncertainties of fieldwork in a working theatre environment, including potential changes to locations, power supply, access, etc. and extremely limited 'dress rehearsals' with a touring production.

3.1 Toolkit components

3.1.1 Annotation server and tablet interface

At the heart of MuSAK is an annotation system used by the musicologist during the performance. Initial designs called for a taxonomic palette of symbols that could be selected on an iPad tablet touchscreen and placed as annotations onto a digital copy of the score. This quickly raised three problems: *(i)* all proposed user interface sketches for selecting one of many annotations were complex and intrusive enough to interrupt score following and the performance observation; *(ii)* the operational cognitive load was judged high and different enough from traditional 'paper and pencil' marking to require a significant period of learning and training before use at a live event; *(iii)* predetermining an adequate set of music and scenic symbols required several weeks' precursory study, leaving limited time to add symbols to the system; furthermore, symbols might be created ad-hoc during use.

[1] http://www.transforming-musicology.org/tools/metaMuSAK

A pragmatic compromise was reached: short piano score pages from the IMSLP music library [2] were shown on a tablet, allowing freehand digital annotations according to a pictogram key of the musicologist's design. Desirable 'in content' semantics were lost, but a user experience strongly matching the traditional and familiar pattern of score marking was gained. It retained digital advantages including timestamped annotations and ease of saving, replacing, modifying, and deleting content. Image layers were 'flattened' to combine scores and existing annotations into new images, which could be redeployed for re-annotation.

A Union Platform [3] server with custom room module was run on a laptop deployed in the theatre, handling storage and communication of the annotation events. To simplify distribution and quick modification, each score page was served to the tablet as a JPEG resource from an HTTP daemon, alongside client HTML and Javascript communicating with Union from the regular Safari web browser.

The web client implements buttons to turn pages and undo annotations; all annotations were recorded using Javascript event handlers to millisecond accuracy and stored both by the browser and by Union which logged to file [4]. The tablet and server were networked using a small battery powered wireless router with a private IP address space.

3.1.2 Digital pen

The tablet tool is, by necessity and design, reductionist. Inevitably some elements of the live performance are worthy of note, but either not preconceived within the symbolic key, or so unique as to require a longer form description. To accommodate this the kit includes a Livescribe Echo [5] digital 'smart pen'. It has a standard ballpoint pen tip, but when used in conjunction with 'Anoto' paper, captures a digital copy of all writing – the paper is printed with a faint non-repeating pattern, which is read by a small infra-red camera in the pen that ascertains nib position within and between pages. While 'in content' semantics are not automatically decoded, the use of the pen is similar to standard note taking, and thus minimises intrusiveness. The digital transcription is downloaded from the pen using USB.

A second feature is a microphone for timestamped audio recordings. When polite to speak aloud, short audio comments were taken in lieu of written notes; when silence was required, the microphone captured background noise for the duration of the performance. While low quality, the latter is sufficient to calibrate temporal synchronicity.

3.1.3 Score following and replay tool

We developed a second, simple, Web application for following and recording the page-turns during the performance by a second operator, independent to the annotator (who might skip forward and backwards between pages to add notes during quieter spells). The score-following page turns capture timings for the realization of the music contained on each page for this specific performance. Pages of the score are rendered one at a time, timestamped in a PostgreSQL database when the user advances to the next page.

An extension of this interface displays tablet annotations (§3.1.1) in real-time using an HTML canvas superimposed over the score. Data is converted from CSV to JSON to ease JavaScript working and a custom renderer calculates the appropriate time delta before drawing a stroke. JSON page-turn timestamps were also combined with the score, turning pages at the correct moment.

3.1.4 Audio and video

As is typical in commercial theatres, audio and video feeds of the performance were available within the venue (§4) but, also typically, limited distribution rights preclude their inclusion in public archives. It is desirable, and for some calculations essential, to reference their implicit existence, particularly when synchronising captured annotations for replay (§3.1.3 & §4) and structured data dissemination (§3.2) – or rather, to explicitly reference the timeline against which the notional recording was made [6]. For replay of annotations it is possible to include a *substitute* audio recording of an alternate performance (§4).

A second distinct video use was recording the annotation actions of the musicologist, providing a contextual reference for toolkit evaluation and, should the Union server fail, potential for reconstruction of annotation times.

3.2 Data publication

The use of semantic technologies to publish performance metadata from the Internet Archive Live Music Archive [7] is described by Bechhofer *et al.* [2], and in the context of diversifying and enriching music information retrieval by Page *et al.* [19]. Crawford *et al.* [7] examines the potential of Linked Data for early music corpora, and Bainbridge *et al.* [1] comments on the effect of musical content analysis and Linked Data in the context of digital libraries. Sebastien *et al.* [22] report on ontology creation for musical perfomance, forms and structures.

Adopting these motivations, and to provide a strong foundation for the further investigation and reuse for musicology and MIR, we have structured our data as RDF. This entails complex ontological structures to fully and explicitly represent the items and their relationships, illustrated in Fig.1 by the timeline patterns required to encode the apparently simple relationship between the annotation of score pages and their performance on stage [8].

A second benefit of web technology is fidelity of access at the resource level. For example, we might publish the overall structure and formal annotations, but restrict access to the video to individually registered ethnographers.

[2] http://imslp.org/

[3] http://www.unionplatform.com/

[4] 1 CSV file per score page with co-ordinate defined paths tracing the annotations, 1 row per straight line. Each annotation path may be described by multiple lines; paths within the same 'pen down'-'pen up' event are given the same timestamp. Pages without annotations are empty but still timestamped to record page turn times.

[5] http://www.livescribe.com/uk/smartpen/echo/

[6] While the recording is not technically required in addition to the *timeline* of the recording, its conceptual, if not actual, inclusion can simplify the metadata encoding structures and increase their comprehension.

[7] https://archive.org/details/etree

[8] See [17] for a detailed description of Linked Data generation.

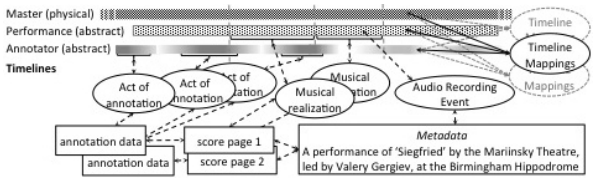

Figure 1. Simplified data modelling.

3.3 Score preparation and musicological annotations

Central to the annotation workflow, and used in several tools, the score page images required several iterations of processing [9] and annotation before distribution in the kit. Piano and vocal arrangements were chosen to reduce the frequency of page turns, converted from IMSLP PDFs to images [10]. Screen use was reduced by a semi-automated process: whitespace detection identified edges, and markers indicating proportionately smaller margins were accepted or adjusted in a simple editing view; saved geometries enacted the crop and enable later scaling of annotations for overlay on original scores. Artefacts from the pre-IMSLP scanning process were cleaned and score images sharpened. Scripts applied a consistent naming scheme for images, used later for inter- and intra-opera page ordering.

A two stage annotation process reduced the note-taking required during the live performance to a minimum. Annotations extracting certain musical points of interest in the score (such as leitmotifs and marked changes of tempo or dynamics) were made by hand, using symbols designed prior to the performance. Each opera score was marked over an average of three days [11] and scanned to images, creating the first derivative layer.

The musicologist used the same symbolic key to annotate this layer using the tablet during the live performance, making further notes where musical aspects differed from those expected and previously marked, and of 'stage directions' (such as lighting, use of props, actions and movements of the characters) which were not directly marked in the original score yet are critical for any interpretation of the performance. These live annotations were 'flattened' into a second derivative image layer of the score.

4. TOOLKIT DEPLOYMENTS

The kit was deployed by a musicologist and two technical assistants as part of a larger project team for the Mariinsky Opera's production of the *Ring* at the Birmingham Hippodrome. Installation in a working theatre hosting a large touring troupe [12] necessitated quick adaptations to the limitations of the available spaces and ad-hoc solutions as events unfolded – the majority beyond the control of the annotation team. Earlier design decisions to reduce the technical complexity of the components proved worthwhile – it is not an understatement to report that a more sophisticated version would have been insufficiently resilient to the challenges of this fieldwork.

For the first night's opera (*Das Rheingold*) the musicologist was located in a dressing room backstage with an audio and video feed from the stage; while the quality of this viewing was far from ideal, it enabled spoken Livescribe annotations [13]. On subsequent nights (*Die Walküre*, *Siegfried*, & *Götterdämmerung*) the 'audio describing' room was used, adjoining a lighting gallery rear of the circle and with an unobstructed view of the stage. In this improved location lights were dimmed and silence maintained; notes were written, not spoken. The annotation server and router were co-located with the musicologist every night and a video camera recorded the annotation process. The score-following annotation system was run from a laptop in a theatre office with an audio feed provided for the operator.

While the simplified design generally paid dividends, there were some malfunctions: we had not expected nor tested for the hour long *second* interval in *Die Walküre* and the connection between tablet and annotation server timed out. The most practicable solution was to restart both tablet and server, losing annotations for the first scene of the third act [14]. A second issue occurred when paging through tablet annotations after a performance, causing time-stamps to be rewritten – original times were reconstrcuted from page turn logs and intra-page timings.

The captured *Ring* totalled 15 hours, consisting four nights' performance over five days, with corresponding tablet activity of over 100,000 strokes making 8,216 annotations and almost 1,300 performance based page turns. The kit deployment and data capture generated 1,316 digital images, 104 pages of writing producing nearly 13 hours of digital pen replay, and 15 hours of video footage. While Network Time Protocol (NTP) clients were used to synchronise equipment clocks some drift was observed, due to differences in Operating Systems and many devices lacking a live connection to an NTP server; these offsets are crucial for data replay and thus explicitly recorded for data publication (§3.2).

A second deployment of the kit demonstrated its flexibility in reconfiguration: at a public engagement event, audience members used their mobile devices to provide annotations while listening to a live audio replay, either by annotating musical score, or "annotating" by placing marks on a simple image with zones for e.g. fast/slow, loud/soft. Both versions of the interface were provided using simultaneous client connections. Comparative visualisations were played to a substitute audio track, derived from a commercial recording using the MATCH Vamp plugin [15] and the rubberband audio time warping tool [16].

[9] The digital processing scripts described here were implemented using Image Magick and the Perl Image::Magick module.

[10] An image file per page; served by respective web servers in the kit.

[11] For context, the score for *Das Rheingold* (the shortest) is over 250 pages; *Götterdämmerung* (the longest) comprises 365 pages.

[12] Whose predominant language was Russian, compared to English for the annotation and Hippodrome teams!

[13] In German, the musicologist's native tongue.

[14] Which includes the section popularly known as the *Ride of the Valkyries*. The cause of this problem was not indicated in logs; rebooting may have destroyed debugging evidence.

[15] https://code.soundsoftware.ac.uk/projects/match-vamp

[16] http://breakfastquay.com/rubberband/

Opera	Shapes /page	S.D.	Duration	S.D.	Overhead	S.D.
Das Rheingold	5.46	4.6	34.92	15.78	-0.028	5.83
Die Walküre	6.95	6.42	44.31	26.59	0.038	8.62
Siegfried	5.76	5.19	39.23	19.06	0.044	12.97
Götterdämmerung	7.22	5.52	43.4	25.45	-0.47	14.98

Table 1. Mean annotation shapes per page, page performance durations and annotation overhead (both seconds).

5. USER EVALUATION

Post-deployment interviews with the musicologist evaluated the usablity of MuSAK in this small trial according to learnability, efficiency, memorability, and satisfaction [12].

Defined as the degree of ease with which functionality can be learnt and task proficiency gained, *learnability* was evaluated through the experience of acquiring the skills necessary to complete the annotation process. The musicologist found the system non-invasive and in-line with existing annotation pragmatics, minimising training time:

"[Annotation is] very similar to the process that I as a musicologist used to do regularly...I think it worked very well because [it] fit in with actions I was very well adapted to...the tools were very non-invasive."

Efficiency of use was measured in annotation shapes per score page and analysed through mean and standard deviation (Table 1). The average number of annotation shapes per page was between five and seven across all operas, corresponding to an average of 9.7 shapes per minute.

Memorability of the kit – the musicologist's recall of set-up and annotation after a five month period of non-use – was assessed using a think-aloud protocol. The evaluation concluded she remembered both to a very high extent.

A qualitative evaluation assessed whether functionality and performance were *satisfactory*: the musicologist described the experience as follows, believing time needed to make additional freehand annotations and cognitively process observations made page turn annotations inaccurate.

"I was quite well able to keep up with the pace... an important realisation is that making these scenic annotations [...]requires a lot of time to think and... process even if it is only like 10 seconds or 5 seconds.

Page turn analysis (§6) indicates that, on average, the annotator could keep pace with the performance.

The musicologist reported an ability to *capture the idiosyncratic profile of each specific performance*, including deviations from the score or expectations based on the score, as well as staging, lighting, and the behaviour of the actors. The kit was described as *supportive of traditional annotation paradigms*, not necessitating new skills for effective use, and the touchpad screen and stylus were:

"intuitive [...] similar to using pen and paper which everyone [...] analysing music is very used to."

The additional affordances of a digital system were noted, including the automatic *capture of the temporal profile of the performance* and the *benefit of being able to easily create corrections, and undo mistakes*.

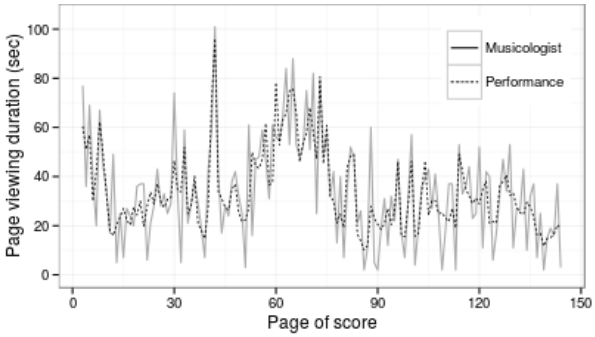

Figure 2. Musicologist's page viewing durations and performance durations for those pages (*Siegfried* Act I).

6. DATA INVESTIGATION

A preliminary analysis considers four research questions to improve our understanding of the data captured. These come with important interpretive contexts: potentially generalisable findings are limited by the scope of data collection to a single performance of the *Ring*; annotations are recorded as continuous shapes from pen touching to leaving the screen, thus symbols comprising several distinct shapes are identified as multiple annotations; and some performance sections are excluded: the start of *Die Walküre* Act III due to kit malfunction, the first page of each opera, and those either side of the intermissions [17] .

6.1 Overhead of annotation task

To determine whether the overhead of annotation interfered with music following, we compared the musicologist's page view durations with the score-page performance events. The corresponding plots reveal strong tracking of the two timelines (Fig.2); Table 1 displays the mean page performance durations, and the time difference compared to the annotator's mean page view durations (the annotation overhead). While performance durations are variable due to changes in tempo and in musical information density on a given page, the magnitude of the mean annotation overhead is below half a second in all four performances. Standard deviations indicate there were periods when annotation acts were delayed, but overall, the musicologist was able to keep up with the music. The value is negative in two performances, indicating a tendency to read ahead.

6.2 Variability of annotation rate

We tested the variability of annotation rates [18] for each night (Table 2; Figure 3). Results demonstrate significant correlations in each performance, accounting for between 18% (*Götterdämmerung*) and 43% (*Walküre*) of the variation in rank between page performance duration and number of annotation shapes per page. The finding of a largely

[17] Pages were left open during the interval so durations are artefactual.

[18] A hypothetical uniform rate would exhibit strong correlation between the duration of a score page performance event and the number of annotation shapes produced for that page.

Opera	Rheingold	Walküre	Siegfried	Götterdämmerung
r_s	.46	.65	.57	.42
r_s^2	.21	.43	.32	.18

Table 2. Spearman correlation ($p<.001$) between page performance duration and annotation shapes per page.

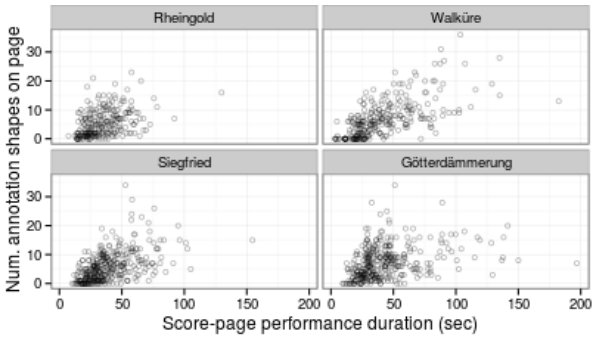

Figure 3. Correlation between score page performance duration and annotation shapes per page.

consistent annotation rate reflects the suitability of the annotation key for the task, suggesting the musicologist could adopt symbols with different granularities of meaning according to the time available. It suggests that even seemingly small 'events' such as gestures were not overlooked.

6.3 Annotation density; Performance correspondence

Finally, we investigated whether periods of high annotation density reflect consistent types of stage events, or corresponded to scenes of high activity or intensity. For this analysis, periods in *Siegfried* with a number of annotations per minute exceeding a threshold of the mean plus two standard deviations, as well as three major peaks in annotation activity for the final act of *Götterdämmerung*, were mapped to page sequences in the annotated score. The musicologist reinspected the corresponding pages, determining that the symbols occurring at these periods predominantly indicate changes in performers' posture or position. These symbols largely consist of four or more shapes; thus, the high rate of annotation during these relative to other periods may be partly artefactual. These peak periods may refer to key dramatic moments, e.g. when Siegfried kisses and awakens Brünnhilde in Act III, or parts of larger dialogic scenes, where every utterance is interpolated by a postural change. Certain passages with greater staged activity (for instance when Siegfried kills Fafner) were not observed within this subset of the data. A possible explanation is that these scenes were staged with a high reliance on lighting effects, annotated with simple symbols; they were also generally drawn out for longer periods, and thus potentially overlooked by our per-minute-metric. The three peaks in the third act of *Götterdämmerung* each reflected essential moments: the Rhinemaidens telling Siegfried about the curse; Hagen killing Gunther; and Hagen struggling with the Rhinemaidens for the Ring. One other expected scene, Siegfried's death, took up over two minutes, and was thus represented by two rate observations that both came close to the threshold without quite meeting it. A refined measure of annotation density accounting for variations in granularity by considering the immediate temporal context could accommodate this issue.

7. CONCLUSIONS AND FUTURE WORK

We have described the software developed to, in combination with off-the-shelf hardware, form a kit used to capture data informing performance studies and the MIR analyses that may be applied to them. We have reported its use to annotate a complete production of Wagner's *Ring* and evaluation of the kit's performance after the deployment. An initial data driven investigation of the annotations has shown it can support and enrich analysis of the performance, and that the corpus could be developed as a 'ground truth' for MIR research.

Investigations to date have focused on temporal analysis of acts of annotation, whereas our next step will examine semantics within the symbols, realizing further benefits for indexing and searching within performance data. We will trial computer vision techniques to categorise pictograms in the annotation layers, and revisit options for encoding stronger symbol semantics during the annotation. While the desirable affordances of the current interface preclude full taxonomic symbol selection, our data analysis suggests even a very coarse grained categorisation (e.g. complex vs. simple events) would yield a much improved musicological understanding of the data. Our work informs future design of symbols used within the kit: ensuring greater uniformity of semantic complexity which would simplify analysis, as would the ability to more clearly delimit writing events, either by the reduction of all symbols to single (rather than compound) drawing, or through a metric combining of temporal and geometric distance. Future deployments of the kit will also record instances of 'undo'.

Our data indicates events with complex layering of type and meaning throughout the performances, cautioning against formulation of naively phrased MIR tasks such as identifying "musicologically interesting parts in this annotated score". Reflecting how tools can be utilised for musicology, our preliminary study makes clear there is unlikely to be a 'perfect' feature to automatically complete a study; instead the method is iterative, with computational analysis informed by musicology research questions and vice versa – through this iteration a fuller understanding of the question, investigation, and its limitations can be found.

8. ACKNOWLEDGEMENTS

This work was supported by the UK Arts and Humanities Research Council *Transforming Musicology* project (AH/L006820/1), part of *Digital Transformations*. The authors gratefully acknowledge assistance during the *Ring* and *Hearing Wagner* events: the Birmingham Hippodrome team, especially Zara Harris and Paul Kaynes; Mariinsky Opera Company; and the greater project team. We thank Jeff Fuller for invaluable Union Platform help and advice.

9. REFERENCES

[1] D. Bainbridge, X. Hu, and J. S. Downie. A musical progression with Greenstone: How music content analysis and linked data is helping redefine the boundaries to a music digital library. In *Proceedings of the 1st International Workshop on Digital Libraries for Musicology in conjunction with the ACM/IEEE Digital Libraries conference 2014*, pages 1–8, 2014.

[2] S. Bechhofer, K. R. Page, and D. De Roure. Hello Cleveland! Linked data publication of live music archives. In *14th International Workshop on Image and Audio Analysis for Multimedia Interactive Services*, pages 1–4, 2013.

[3] C. Cannam, C. Landone, M. Sandler, and J. Bello. The Sonic Visualiser: A visualisation platform for semantic descriptors from musical signals. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 324–327, 2006.

[4] H. Christenson. HathiTrust. *Library Resources & Technical Services*, 55(2):93–102, 2011.

[5] N. Cook. Methods for analysing recordings. In *The Cambridge Companion to Recorded Music*, pages 221–245, 2009.

[6] N. Cook. *Beyond the Score: Music as Performance*. Oxford University Press, Oxford, 2014.

[7] T. Crawford, B. Fields, D. Lewis, and K. R. Page. Explorations in linked data practice for early music corpora. In *Proceedings of the IEEE/ACM Joint Conference on Digital Libraries*, pages 309–312, 2014.

[8] M. Doğantan Dack. The art of research in live music performance. *Music Performance Research*, 5:34–48, 2012.

[9] H. Danuser. Urteil und Vorurteil im Interpretationsvergleich. *Zeitschrift für Musiktheorie*, 6(2):76–88, 1975.

[10] M. Doerr, C. Bekiari, P. LeBoeuf, and Bibliotheque Nationale de France. FRBRoo, a conceptual model for performing arts. In *2008 Annual Conference of CIDOC, Athens*, pages 06–18, 2008.

[11] D. Dolan, J. Sloboda, H. J. Jensen, B. Crüts, and E. Feygelson. The improvisatory approach to classical music performance: An empirical investigation into its characteristics and impact. *Music Performance Research*, 6, 2013.

[12] X. Ferré, N. Juristo, H.Windl, and L. Constantine. Usability basics for software developers. *IEEE Software*, 18(1):22–29, 2001.

[13] X. Hu and Y. Yang. Cross-cultural mood regression for music digital libraries. In *Proceedings of the IEEE/ACM Joint Conference on Digital Libraries*, pages 471–472, 2014.

[14] B. Kershaw. Performance as research: Live events and documents. In *The Cambridge Companion to Performance Studies*, pages 23–45, 2008.

[15] A. Marsden. 'What was the question?': Music analysis and the computer. In *Modern Methods for Musicology*, pages 137–147, 2009.

[16] C. A. Mullin. International music score library project/Petrucci music library (review). *Notes*, 67(2):376–381, 2010.

[17] T. Nurmikko-Fuller, D. M. Weigl, and K. R. Page. On organising multimedia performance corpora for musicological study using linked data. In *Proceedings of the 2nd International Workshop on Digital Libraries for Musicology*, pages 25–28, 2015.

[18] K. Okumura, S. Sako, and T. Kitamura. Stochastic modeling of a musical performance with expressive representations from the musical score. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 531–536, 2011.

[19] K. R. Page, B. Fields, B. J. Nagel, G. O'Neill, D. De Roure, and T. Crawford. Semantics for music analysis through linked data: How country is my country? In *e-Science (e-Science), 2010 IEEE Sixth International Conference on*, pages 41–48, 2010.

[20] A. Porter, M. Sordo, and X. Serra. Dunya: a system to browse audio music collections exploiting cultural context. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 101–106, 2013.

[21] R. C. Repetto and X. Serra. Creating a corpus of Jingju (Beijing opera) music and possibilities for melodic analysis. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 313–318, 2014.

[22] V. Sebastien, D. Sebastien, and N. Conruyt. Annotating works for music education: Propositions for a musical forms and structures ontology and a musical performance ontology. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 451–456, 2013.

[23] J. B. Smith, J. A. Burgoyne, I. Fujinaga, D. De Roure, and J. S. Downie. Design and creation of a large-scale database of structural annotations. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 555–560, 2011.

[24] S. Tresize, N. Cook, E. Clarke, D. Leech-Wilkinson, and J. Rink. The recorded document: Interpretation and discography. In *The Cambridge Companion to Recorded Music*, pages 186–209, 2009.

[25] M. Winget. 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):1878–1897, 2008.