

INTERACTIVE SCORES IN CLASSICAL MUSIC PRODUCTION

Simon Waloschek, Axel Berndt, Benjamin W. Bohl, Aristotelis Hadjakos

Center of Music And Film Informatics (CeMFI)

University of Music Detmold, Germany

{waloschek,berndt,hadjakos}@hfm-detmold.de, bohl@edirom.de

ABSTRACT

The recording of classical music is mostly centered around the score of a composition. During editing of these recordings, however, further technical visualizations are used. Introducing digital interactive scores to the recording and editing process can enhance the workflow significantly and speed up the production process. This paper gives a short introduction to the recording process and outlines possibilities that arise with interactive scores. Current related music information retrieval research is discussed, showing a potential path to score-based editing.

1. INTRODUCTION

Classical music generally revolves around the musical score. It is used as fundamental interpretation directions by musicians and conductors. During recording of classical music, scores are used as a means of communication as well as direct working material of record producers. Successive working steps towards a finished music production however utilize additional views upon the recorded audio material while still frequently referring to the score. This media disruption can take a great deal of time since these different views are not synchronized in any way. Although most technologies that are needed to overcome this disadvantage are already present, they have not been used in this specific field of application. This paper therefore summarizes the ongoing efforts of introducing interactive scores to the classical music production process and discusses open issues.

2. CLASSICAL MUSIC PRODUCTION

In order to understand the score-related needs and issues arising during classical music production, a brief overview of the production process will be given. It is neither complete in terms of performed steps taken nor does it claim to comprehensively address every aspect of the production



Figure 1. Annotations made in the score during a recording session.

process. Tasks that have implications for further considerations will be outlined in more detail.

The production of classical music recordings is usually divided into three major phases: *pre-production*, *production* and *post-production* [5]. During *pre-production* an essential goal is to set up a production plan in concordance with the artistic director or conductor. From the record producer's perspective this of course includes analyzing the piece(s) of music to be recorded. This includes several aspects, such as identifying challenging passages. Generally speaking, the record producer's main goal is to familiarize himself with the piece in a way that will allow him to perform his later tasks in a best possible manner, e.g. by listening to existing recordings and studying the score. During this process, the record producer might annotate and mark passages in the score for later consideration. As the score will later be a major means of communication with the conductor or musicians, it should be identical to the conducting score with respect to its appearance, e.g. page-layout or reference points.

Capturing the raw audio material from the musicians' performance in the *Digital Audio Workstation* (DAW) is the main goal of the *production phase*. This might be done in several recording sessions, depending on the scope and nature of the music. Moreover it is common practice to repeat musically or technically unsatisfying passages multiple times but yet keep all of the recorded *takes*.



© Simon Waloschek, Axel Berndt, Benjamin W. Bohl, Aristotelis Hadjakos. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Simon Waloschek, Axel Berndt, Benjamin W. Bohl, Aristotelis Hadjakos. "Interactive Scores in Classical Music Production", 17th International Society for Music Information Retrieval Conference, 2016.

The responsible record producer has to carefully listen to the played music as well as pay attention to the musical score at the same time. Deviations from the score and other score-related comments—positive and negative—are mostly annotated directly in the score as shown in Figure 1. It is to be noted that there is no standardized set of symbols that is commonly used by record producers, e.g. beginnings or endings of the takes or annotating the quality of a certain passage during a take. Every producer develops his own set of symbols based on his personal experience and specific needs.

Oftentimes, an additional *take list* is manually maintained that reflects the beginning and ending measures of the individual takes in relation to their take number, see Table 1.

On the basis of their observations during individual takes, the record producer’s tasks include keeping an overview of whether all passages have been recorded in a satisfying manner. If they have not, they will communicate this to the musicians and ask for additional takes. Communication with the musicians is done orally using dedicated audio lines aka talkback channels. For this purpose the score and its consistency with the conductor’s score and the musicians’ parts is an essential basis for communication. Page numbers, measure numbers and other reference marks of the score will be used to communicate the precise location to be considered.

The following editing process is dominated by selecting and splicing parts from the takes that best conform to the musical aesthetics and technical demands of the production. This often requires reviewing huge amounts of audio data with identical sections spread across the various takes. With the takes usually being rendered consecutively in the editing environment as *waveforms* (see Figure 2), navigation for auditory comparisons becomes a time-consuming task, see Section 3.4.

A well-organized take list helps to decrease the time needed to identify the takes containing a specific section of the recorded composition. Nevertheless, deciding which takes to splice might be a process of consecutive comparisons of several takes that often cannot be broken down to mere technical aspects (in the sense of the musicians’ playing technique, as well as the recording quality) but has to account for aesthetic aspects too, e.g. the quality of a passage’s musical performance. When a decision has been made about which takes to splice, it comes to rendering the splice imperceptible. Besides some technical aspects like selecting an adequate zero crossing in the waveforms, ad-

Take	Pos. / Measures	Comment
1	$\alpha - \Omega$	
2	$\alpha - 17$	Beautiful start
3	13 - 31	
4	22 - 31	Quarters not in time
5	29 - 52	
⋮	⋮	

Table 1. Exemplary list of takes from a recording session.

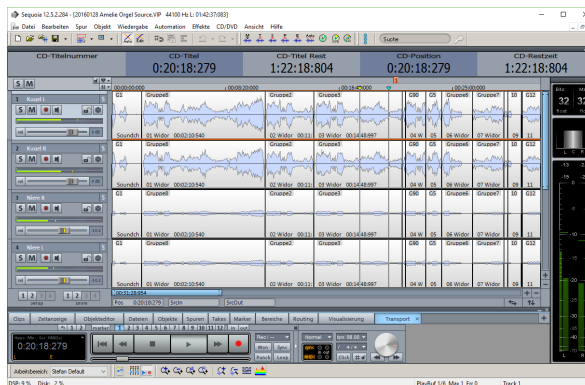


Figure 2. DAW with multiple takes lined up in a row.

justing the loudness of the takes, and optimizing the cross-fade between the two takes, it is the editor’s ears that will allow them to assess the quality of the splice. All in all, this often means specifying the precise edit location by ear and on the basis of a sound music-aesthetic sensitivity of the editor.

The mixing phase ensures that the sound of the recording has the right balance in terms of each instrument’s volume levels. Dynamics and panoramic position are manipulated and filters and effects such as reverb may be added to produce a mix that is more appealing to the listener. These tasks as well as the final mastering are not further considered, as the annotated score and take list do not play major roles in them.

3. INTERACTIVE SCORES AS MUSIC PRODUCTION INTERFACES

Replacing conventional paper scores with their interactive counterparts opens up a variety of workflow optimizations for music production and establishes fruitful connections to further research and application areas such as digital music edition. The following sections describe the parts of the workflow that could benefit from a more ubiquitous use of digital scores. The corresponding research topics are mainly located in the field of human-computer interaction (pen and touch interaction, gesture development and recognition, user interface design), audio processing and music information retrieval (audio-to-audio and audio-to-score alignment).

3.1 Sheet Music Interaction and Annotations

Sheet music marks a central work object in the pre-production phase, recording session (here in particular) and in the editing phase as the preceding introduction to classical music production shows. Handwritten annotations in the score are an effective, easily accessible and versatile means to document the recording process and communicate with the musicians. However, the number of remarks increases drastically during a recording session and tends to hamper readability and complicate the assignment of annotations to specific takes.

Figure 3. Mock-up of a Pen and Touch-based User Interface for Take Selection and Editing

A transformation of the analog writing on the music sheet into the digital domain can be achieved via digital pen and paper technology such as Anoto. An overview of respective technologies can be found in [20]. The advantage of digital paper lies in its possibility to link the annotations with the recorded audio data and process them accordingly. Limitations become apparent when musical sections have to be recorded repeatedly, each take introducing new annotations until the paper sheet is overfull and hardly readable. Furthermore, the step from the recording to the editing phase introduces a media disruption. In the latter, the focus lies on navigating, selecting and editing takes which cannot be done with the sheet music. Here, a continuous staff layout, aligned with the audio data, is desirable.

Even though printed sheet music has some practical limitations, it features the aforementioned clear advantages. These motivate a fundamental design decision that will underlay our further concept development: *The interactive score should become the central interface widget during the recording and editing phase.* Up to now, the DAW (see Figure 2) marks the central productive interface and the printed score serves as a secondary medium to hold further information. To preserve the advantages of (digital) pen and paper we regard pen and touch displays as the most promising technology. The score layout can easily be switched from a traditional page-wise style during the recording sessions to a continuous staff layout in the editing phase. Annotations can likewise be shifted as they are linked to score positions (e.g., measures and single notes). Figure 3 demonstrates the continuous score layout, aligned with the recorded audio material and supplemented by handwritten annotations.

We conceive pen and touch interaction with interactive scores for music production according to the following scheme. Touch input is mainly used for navigation (turning score pages, panning and zooming) and to toggle buttons and input mode switches. Productive input—primarily the

creation of annotations and the editing of takes—is done with the pen as these require a higher precision. At this, we follow the same precedent as Yee [19], Hinckley et al [10] and Frisch [7].

Annotations are layered for each take, i.e., each recording starts with a non-annotated score. However, previous annotations can be switched on, if required. This overcomes the problem of overfull music sheets during the recording session. Hence, the record producer can make annotations at their exact place in the score without having to deal with previous annotations. Annotations can be structured, moved, hidden, or deleted. This allows, for example, to show only those annotations that have been written throughout the last three takes, helping to keep an overview. Mostly, annotations are also rated as positive, neutral or negative which helps the record producer to select the best takes in the editing phase. Such ratings may be indicated by symbols such as “+”, “-” and “~”. All annotations have to be made very quickly during the music performance and each additional mark costs time. Instead of such additional symbols, the side-switch of a digital pen and its eraser can be used as mode switches and the annotations may be color coded accordingly. These may come in handy during the editing phase to quickly find the right takes (see Section 3.4).

Moreover, annotations can even serve as control gestures. Record producers typically note the start and stop position of a take in the score by “-” and “-” and a serial number. Instead of controlling the recording functionality at a different place, it can be triggered immediately when the input symbol is recognized as a control gesture. The take’s serial number and naming can be generated. The symbols and their positions in the score further help to align the recorded audio material with the score.

3.2 Protocol Automation

Centering the recording and editing workflow around digital scores is advantageous also during the pre-production



Figure 4. A pen and touch display used for score annotation.

phase. Digital scores can be generated from commonly used notation formats such as MusicXML and MEI (Section 4.1 addresses music formats with regard to technological requirements). In the latter case, i.e. MEI, elements of the editors' critical report can be included and help to clarify the musical idea.

The information that are provided by MusicXML, MEI and other symbolic music representation formats, that may underlay the digital score, often include information about the instrumentation, the number of voices and voice groupings. From these information a basic recording project can be automatically initialized, i.e. the creation of audio tracks and their naming. Even a rough panning setup can be generated from the typical orchestra setting and from the miking plan that is made during the preparatory meeting.

During the recording sessions, the record producer might maintain a take list, see Table 1. This is done in addition to the interactions in the DAW and the annotations in the score. The information in the take list is actually redundant with the score annotations and can be generated from them which relieves the record producer. Take list comments can be generated from the qualitative connotations of the score annotations (positive, negative, neutral) and from textual annotations and symbols that may be recognized by the system.

3.3 Communication

Interactive scores can help communicating with the musicians and serve as communication channel. In the recording session the music notation is displayed in a traditional page-wise score layout that matches with the layout that the musicians have. This eases the communication. Referring to a position in the music is mostly done in the format (page, staff line, measure). If the musicians use digital music stands, the digital score can even become a means for visual communication. Here, the record producer may make annotations in their score that are synchronously displayed on the musicians' music stands. This audiovisual mode can help making the communication more effective, less ambiguous and faster.

3.4 Take Selection & Editing

The editing phase and the recording phase utilize different facets of the interface. During the recording phase, the record producer needs to be able to orient himself quickly in the digital score that must be consistent with the musicians' score¹ to facilitate communication. In the editing phase the score and all its additional information (annotations, take list etc.) should facilitate a quick selection of suitable takes which are then spliced via cross-fades.

Since the music editing process changed from analog to digital, the average splice count and frequency increased drastically [18]. We conducted a preliminary survey with 15 record producers to determine the approximate durations of specific tasks in the editing phase. While recording a 10 minute piece takes approx. 2:26h, the pure navigation and splicing process takes 1.62 times as much. The actual selection of the takes in terms of aesthetics was not considered. Navigation between suitable takes marks the most time-consuming part of the editing phase, 54.9% of the time.

Cues that help to identify promising takes are spread over the score and the take list. Takes are arranged consecutively and not aligned in accordance with their actual musical content. However, it is possible to tackle these flaws and approach a solution similar to Figure 3, i.e., a continuous score, each take aligned with it and color coded for qualitative indication.

From the control gestures ("←" and "→", see Section 3.1) we know each take's corresponding score position. This helps aligning them with each other and with the score. Annotations made in the recording phase are linked to positions and even regions in the score. They can also be linked to the actual audio recordings via an audio-to-score alignment. Thereby, problematic passages can be indicated directly in the takes since annotations have a qualitative connotation. Selecting a take reveals its annotations in the score. Based on these qualitative connotations, takes can be recommended and automatically combined into a "raw edit version". This does not make the detailed work of the editor obsolete but accelerates the time-consuming search for good candidates.

4. TECHNOLOGICAL STATE OF THE ART

Many of the previously outlined aspects and issues have already been addressed by current research. This section provides an overview of the relevant developments and links together fundamental techniques that can be used to implement the aforementioned features.

4.1 Digital Score Format

The most basic requirement for score interactivity is the availability of scores in a digital format. Unfortunately, most publishers do still publish their scores solely in a printed form. In order to produce digital counterparts, two different ways can be employed: *Optical Music Recognition (OMR)* [15], which aims at digitizing printed scores

¹ at least with the conductor's score.

via image recognition algorithms, and direct encoding of notes in an appropriate data format. While OMR techniques have been used for linking scores with audio [6, 11], they present the same layout related issues discussed in Section 3.1. Therefore, further considerations concentrate mainly on symbolic music encoding.

In order to adequately represent symbolic music data, the *Music Encoding Initiative* (MEI) [16] was founded and elaborated a data format (also called MEI), that is able to hold both musical and layout data of a score.

Although encoding music directly in such formats is a rather time-consuming task, it offers the best flexibility in terms of post processing capabilities and manual layout restructuring. The latter oftentimes provides a tremendous advantage in readability over automatically generated score layouts.

Ongoing digitization efforts of major music publishers will most likely help to overcome the lack of digital scores in the near future.

4.2 Digital Score Rendering

Various approaches have been used in rendering digital scores paired with interactivity, mainly for musical practice. MOODS [2] and muse [8] are two early examples.

MOODS is a collaborative music editor and viewer, which lets musicians, the conductor and the archivist view and annotate the score based on their user privileges. A similar approach would be useful to support the communication between the record producer and the musicians during music production (see Section 3.3). The muse lets the user view and annotate the music score. It turns pages automatically using audio to score alignment [8]. Page-turning in the MOODS system is based on a horizontal separator that splits the page into two parts: the part below the separator is currently being played, the part above is a preview of the next page, which is good for musicians who oftentimes read ahead.

An alternative representation is an infinite scroll of music score without page breaks as shown in Figure 3. In the score editing software *Sibelius* this approach is called “Panorama” and its “Magic Margins” provide information about clef, key and measure number on the left side.²

With the advent of *Verovio* [14], a more modern approach to score rendering is available. Providing an easy interface to render MusicXML as well as MEI files with custom layout properties, it is gaining usage amongst web-based score applications.

4.3 Linking Scores and Audio

In order to visualize the recorded audio takes time-synchronously to the score (see Figure 3), both representations have to be aligned algorithmically; Each position in the individual takes should be linked to the equivalent position in the score and vice versa.

² <http://www.sibelius.com/products/sibeliusedu/5/panorama.html> (last accessed March 2016)

Symbolic music representations offer the possibility to be transformed into audio files, reducing the *score-to-audio-alignment* task to the (commonly considered) simpler problem of *audio-to-audio-alignment*. This way, the generated audio can be annotated automatically with cue points from the score. Tools such as *music21* [9] and *meico*³ are able to transform MEI, MusicXML etc. into MIDI, which in return can be used to output audio data.

In the next step, these audio data are to be aligned with the recorded takes. Various approaches have been applied and are thoroughly discussed by Thomas et al. [17] and Müller [13]. Annotations in the score during the recording phase as described in Section 3.1 can be used to bypass false alignments in situations, where identical musical passages occur multiple times throughout a composition.

Scenarios with partial usage of alignment techniques in recording situations can be found in [4, 12], though practical implementations are still lacking. This renders comprehensive evaluations of new score-based editing methods impossible.

4.4 Interaction

Changing from printed to digital scores allows for a wide range of interaction possibilities as presented in Section 3.4. Rendering engines like *Verovio* output the score as *Scalable Vector Graphics* (SVGs), which can be viewed in every modern web browser. SVG however developed from a pure visualization format into a major interactivity framework. The content can be changed programmatically with low expenditure of time and allows for pixel-precise reaction to pen and touch input. Using a web browser as front-end enables the usage of JavaScript as the underlying programming language.

Ready-to-use gesture frameworks, e.g. the \$N Multistroke Recognizer [1], can easily be incorporated and adapted to make use of established Pen & Touch interaction modalities [3] and the aforementioned gestures to start and stop the recording etc. due to its JavaScript nature.

Hardware-wise, large pen-enabled touch screen displays as shown in Figure 4 are already widely used by media designers and can be adapted to the discussed scenario without further modifications.

5. CONCLUSIONS

Although many technical aspects of interactive scores in recording scenarios have already been addressed by research, the main issue remains in bringing their results together in a usable manner. This usability is generally driven by in-depth insights into the workflow of record producers. Classical music recording is in its roots a rather conservative task that remains skeptical about new developments and, therefore, requires comprehensive analysis in advance. Unfortunately, this field of work seems to be not very accessible in terms of open exchange of working strategies.

³ <http://www.zemfi.de/resources/meico-mei-converter/> (last accessed March 2016)

Once the (classical) recording process is well documented, new interface and usability concepts as exemplary outlined in Section 3 can be developed and evaluated. To have an impact on the actual work processes, the conception and development should be as close as possible to potential users.

The software implementation of such an interface can gain advantage of present MIR and usability engineering research and brings together several topics in a new way. Thus, it provides an interesting and unique use case for future research on music information retrieval methods combined with user interaction analysis.

6. REFERENCES

- [1] L. Anthony and J. O. Wobbrock. \$N\$-Protractor: A Fast and Accurate Multistroke Recognizer. In *Proc. of Graphics Interface*, pages 117–120. Canadian Information Processing Soc., 2012.
- [2] Pierfrancesco Bellini, Paolo Nesi, and Marius B Spinu. Cooperative visual manipulation of music notation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(3):194–237, 2002.
- [3] P. Brandl, C. Forlines, D. Wigdor, M. Haller, and Ch. Shen. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proc. of the Working Conf. on Advanced Visual Interfaces*, pages 154–161. ACM, 2008.
- [4] R. B. Dannenberg and N. Hu. Polyphonic Audio Matching for Score Following and Intelligent Audio Editors. In *Proc. of the Int. Computer Music Conf.*, San Francisco, 2003.
- [5] J. Eargle. *Handbook of Recording Engineering*. Springer US, 2006.
- [6] Chr. Fremerey, M. Müller, F. Kurth, and M. Clausen. Automatic Mapping of Scanned Sheet Music to Audio Recordings. In *Proc. of the Int. Soc. for Music Information Retrieval Conf.*, pages 413–418, 2008.
- [7] M. Frisch. *Interaction and Visualization Techniques for Node-Link Diagram Editing and Exploration*. Verlag Dr. Hut, Munich, Germany, 2012.
- [8] Christopher Graefe, Derek Wahila, Justin Maguire, and Orya Dasna. Designing the muse: A digital music stand for the symphony musician. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 436–ff. ACM, 1996.
- [9] A. Hankinson, P. Roland, and I. Fujinaga. The Music Encoding Initiative as a Document-Encoding Framework. In *Proc. of the Int. Soc. for Music Information Retrieval Conf.*, pages 293–298, 2011.
- [10] K. Hinckley, K. Yatani, M. Pahud, N. Coddington, J. Rodenhouse, A. Wilson, H. Benko, and B. Buxton. Pen + Touch = New Tools. In *Proc. of the 23rd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 27–36, New York, NY, USA, 2010. ACM.
- [11] Ö. İzmirlı and G. Sharma. Bridging printed music and audio through alignment using a mid-level score representation. In *Proc. of the Int. Soc. for Music Information Retrieval Conf.*, pages 61–66, 2012.
- [12] N. Montecchio and A. Cont. Accelerating the mixing phase in studio recording productions by automatic audio alignment. In *Proc. of the Int. Soc. for Music Information Retrieval Conf.*, 2011.
- [13] M. Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer International Publishing, 2015.
- [14] L. Pugin, R. Zitellini, and P. Roland. Verovio: A Library For Engraving MEI Music Notation Into SVG. In *Proc. of the Int. Soc. for Music Information Retrieval Conf.*, Taipei, Taiwan, 2014.
- [15] Ana Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. S. Marcal, C. Guedes, and J. S. Cardoso. Optical Music Recognition: State-of-the-Art and Open Issues. *International Journal of Multimedia Information Retrieval*, 1(3):173–190, 2012.
- [16] Perry Roland. The music encoding initiative (mei). In *Proceedings of the First International Conference on Musical Applications Using XML*, pages 55–59, 2002.
- [17] V. Thomas, Chr. Fremerey, M. Müller, and M. Clausen. Linking Sheet Music and Audio – Challenges and New Approaches. *Dagstuhl Follow-Ups*, 3, 2012.
- [18] S. Weinzierl and C. Franke. 'Lotte, ein Schwindel!' – History and practice of editing recordings of Beethoven's symphony No. 9. In 22. *Tonmeistertagung – VDT International Convention*, 2003.
- [19] K. Yee. Two-handed Interaction on a Tablet Display. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1493–1496, Vienna, Austria, 2004. ACM.
- [20] R. B. Yeh. *Designing Interactions That Combine Pen, Paper, and Computer*. PhD thesis, Stanford, CA, USA, 2008.