

Computational Music Structure Analysis

Edited by

Meinard Müller¹, Elaine Chew², and Juan Pablo Bello³

1 Universität Erlangen-Nürnberg, DE, meinard.mueller@audiolabs-erlangen.de

2 Queen Mary University of London, GB, elaine.chew@qmul.ac.uk

3 New York University, US, jpbello@nyu.edu

Abstract

Music is a ubiquitous and vital part of the lives of billions of people worldwide. Musical creations and performances are among the most complex and intricate of our cultural artifacts, and the emotional power of music can touch us in surprising and profound ways. In view of the rapid and sustained growth of digital music sharing and distribution, the development of computational methods to help users find and organize music information has become an important field of research in both industry and academia.

The Dagstuhl Seminar 16092 was devoted to a research area known as music structure analysis, where the general objective is to uncover patterns and relationships that govern the organization of notes, events, and sounds in music. Gathering researchers from different fields, we critically reviewed the state of the art for computational approaches to music structure analysis in order to identify the main limitations of existing methodologies. This triggered interdisciplinary discussions that leveraged insights from fields as disparate as psychology, music theory, composition, signal processing, machine learning, and information sciences to address the specific challenges of understanding structural information in music. Finally, we explored novel applications of these technologies in music and multimedia retrieval, content creation, musicology, education, and human-computer interaction.

In this report, we give an overview of the various contributions and results of the seminar. We start with an executive summary, which describes the main topics, goals, and group activities. Then, we present a list of abstracts giving a more detailed overview of the participants' contributions as well as of the ideas and results discussed in the group meetings of our seminar.

Seminar February 28–March 4, 2016 – <http://www.dagstuhl.de/16092>

1998 ACM Subject Classification H.5.5 Sound and Music Computing

Keywords and phrases Music Information Retrieval, Music Processing, Music Perception and Cognition, Music Composition and Performance, Knowledge Representation, User Interaction and Interfaces, Audio Signal Processing, Machine Learning

Digital Object Identifier 10.4230/DagRep.6.2.147

Edited in cooperation with Stefan Balke

1 Executive Summary

Meinard Müller

Elaine Chew

Juan Pablo Bello

License  Creative Commons BY 3.0 Unported license
© Meinard Müller, Elaine Chew, and Juan Pablo Bello

In this executive summary, we start with a short introduction to computational music structure analysis and then summarize the main topics and questions raised in this seminar.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Computational Music Structure Analysis, *Dagstuhl Reports*, Vol. 6, Issue 2, pp. 147–190

Editors: Meinard Müller, Elaine Chew, Juan Pablo Bello



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Furthermore, we briefly describe the background of the seminar’s participants, the various activities, and the overall organization. Finally, we reflect on the most important aspects of this seminar and conclude with future implications and acknowledgments.

Introduction

One of the attributes distinguishing music from other types of multimedia data and general sound sources are the rich, intricate, and hierarchical structures inherently organizing notated and performed music. On the lowest level, one may have sound events such as individual notes, which are characterized by the way they sound, i.e., their timbre, pitch and duration. Such events form larger structures such as motives, phrases, and chords, and these elements again form larger constructs that determine the overall layout of the composition. This higher structural level is specified in terms of musical parts and their mutual relations. The general goal of *music structure analysis* is to segment or decompose music into patterns or units that possess some semantic relevance and then to group these units into musically meaningful categories.

While humans often have an intuitive understanding of musical patterns and their relations, it is generally hard to explicitly describe, quantify, and capture musical structures. Because of different organizing principles and the existence of temporal hierarchies, musical structures can be highly complex and ambiguous. First of all, a temporal segmentation of a musical work may be based on various properties such as homogeneity, repetition, and novelty. While the musical structure of one piece of music may be explained by repeating melodies, the structure in other pieces may be characterized by a certain instrumentation or tempo. Then, one has to account for different musical dimensions, such as melody, harmony, rhythm, or timbre. For example, in Beethoven’s Fifth Symphony the “fate motive” is repeated in various ways – sometimes the motive is shifted in pitch, sometimes only the rhythmic pattern is preserved. Furthermore, the segmentation and structure will depend on the musical context to be considered; in particular, the threshold of similarity may change depending on the timescale or hierarchical level of focus. For example, the recapitulation of a sonata may be considered a kind of repetition of the exposition on a coarse temporal level even though there may be significant modifications in melody and harmony. In addition, the complexity of the problem can depend on how the music is represented. For example, while it is often easy to detect certain structures such as repeating melodies in symbolic music data, it is often much harder to automatically identify such structures in audio representations. Finally, certain structures may emerge only in the aural communication of music. For example, grouping structures may be imposed by accent patterns introduced in performance. Hence, such structures are the result of a creative or cognitive process of the performer or listener rather than being an objective, measurable property of the underlying notes of the music.

Main Topics and Questions

In this seminar, we brought together experts from diverse fields including psychology, music theory, composition, computer science, music technology, and engineering. Through the resulting interdisciplinary discussions, we aimed to better understand the structures that emerge in composition, performance, and listening, and how these structures interrelate. For example, while there are certain structures inherent in the note content of music, the perception and communication of structure are themselves also creative acts subject to interpretation. There may be some structures intended by the composer or improviser, which are not fully communicated by symbolic descriptions such as musical score notation. The

performer, if different from the composer, then must interpret structures from the score, and decide on the prosodic means by which to convey them. When a listener then tries to make sense of the performed piece, that act of sense-making, of constructing structure and meaning from an auditory stream is also a creative one. As a result, different people along this communication chain may come up with different solutions, depending on their experiences, their musical backgrounds, and their current thinking or mood.

Based on our discussions of various principles and aspects that are relevant for defining musical patterns and structures, the following questions were raised.

- How can ambiguity in notions such as repetition, similarity, grouping, and segmentation be handled and modeled?
- In which way do these notions depend on the music style and tradition?
- How can one account for the relations within and across different hierarchical levels of structural patterns?
- How can long-term structures be built up from short-term patterns, and, vice versa, how can the knowledge of global structural information support the analysis of local events?
- How can information on rhythm, melody, harmony, timbre, or dynamics be fused within unifying structural models?
- How can the relevance of these aspects be measured?
- How do computational models need to be changed to account for human listeners?

By addressing such fundamental questions, we aimed for a better understanding of the principles and model assumptions on which current computational procedures are based, as well as the identification of the main challenges ahead.

Another important goal of this seminar was to discuss how computational structure analysis methods may open up novel ways for users to find and access music information in large, unstructured, and distributed multimedia collections. Computational music structure analysis is not just an end in itself; it forms the foundation for many music processing and retrieval applications. Computational methods for structuring and decomposing digitized artifacts into semantically meaningful units are of fundamental importance not only for music content but also for general multimedia content including speech, image, video, and geometric data. Decomposing a complex object into smaller units often constitutes the first step for simplifying subsequent processing and analysis tasks, for deriving compact object descriptions that can be efficiently stored and transmitted, and for opening up novel ways for users to access, search, navigate, and interact with the content. In the music context, many of the current commercially available services for music recommendation and playlist generation employ *context-based* methods, where textual information (e. g., tags, structured metadata, user access patterns) surrounding the music object are exploited. However, there are numerous data mining problems for which context-based analysis is insufficient, as it tends to be low on specifics and unevenly distributed across artists and styles. In such cases, one requires *content-based* methods, where the information is obtained directly from the analysis of audio signals, scores and other representations of the music. In this context, the following questions were raised.

- How can one represent partial and complex similarity relations within and across music documents?
- What are suitable interfaces that allow users to browse, interact, adapt, and understand musical structures?
- How can musical structures be visualized?
- How can structural information help improve the organizing and indexing of music collections?

Participants, Interaction, Activities

In our seminar, we had 31 participants, who came from various locations around the world including North America (8 participants from the U.S.), Asia (2 participants from Japan), and Europe (21 participants from Austria, France, Germany, Netherlands, Portugal, Spain, United Kingdom). Many of the participants came to Dagstuhl for the first time and expressed enthusiasm about the open and retreat-like atmosphere. Besides its international character, the seminar was also highly interdisciplinary. While most of the participating researchers are working in the fields of music information retrieval, we have had participants with a background in musicology, cognition, psychology, signal processing, and other fields. This led to the seminar having many cross-disciplinary intersections and provoking discussions as well as numerous social activities including playing music together. One particular highlight of such social activities was a concert on Thursday evening, where various participant-based ensembles performed a wide variety of music including popular music, jazz, and classical music. Some of the performed pieces were original compositions by the seminar's participants.

Overall Organization and Schedule

Dagstuhl seminars are known for having a high degree of flexibility and interactivity, which allows participants to discuss ideas and to raise questions rather than to present research results. Following this tradition, we fixed the schedule during the seminar asking for spontaneous contributions with future-oriented content, thus avoiding a conference-like atmosphere, where the focus tends to be on past research achievements. After the organizers have given an overview of the Dagstuhl concept and the seminar's overall topic, we started the first day with self-introductions, where all participants introduced themselves and expressed their expectations and wishes for the seminar. We then continued with a small number of ten-minute stimulus talks, where specific participants were asked to address some critical questions on music structure analysis in a nontechnical fashion. Each of these talks seamlessly moved towards an open discussion among all participants, where the respective presenters took over the role of a moderator. These discussions were well received and often lasted for more than half an hour. The first day closed with a brainstorming session on central topics covering the participants' interests while shaping the overall schedule and format of our seminar. During the next days, we split into small groups, each group discussing a more specific topic in greater depth. The results and conclusions of these parallel group sessions, which lasted between 60 to 90 minutes, were then presented to, and discussed with, the plenum. Furthermore, group discussions were interleaved with additional stimulus talks spontaneously given by participants. This mixture of presentation elements gave all participants the opportunity for presenting their ideas to the plenum while avoiding a monotonous conference-like presentation format. Finally, on the last day, the seminar concluded with a session we called "self-outroductions" where each participant presented his or her personal view of the main research challenges and the seminar.

Conclusions and Acknowledgment

Having the Dagstuhl seminar, our aim was to gather researchers from different fields including information retrieval, signal processing, musicology and psychology. This allowed us to approach the problem of music structure analysis by looking at a broad spectrum of data analysis techniques (including signal processing, machine learning, probabilistic models, user studies), by considering different domains (including text, symbolic, image, audio representations), and by drawing inspiration from creative perspectives of the agents

(composer, performer, listener) involved. As a key result of this seminar, we achieved some significant progress towards understanding, modeling, representing, extracting, and exploiting musical structures. In particular, our seminar contributed to further closing the gap between music theory, cognition, and the computational sciences.

The Dagstuhl seminar gave us the opportunity for having interdisciplinary discussions in an inspiring and retreat-like atmosphere. The generation of novel, technically oriented scientific contributions was not the focus of the seminar. Naturally, many of the contributions and discussions were on a rather abstract level, laying the foundations for future projects and collaborations. Thus, the main impact of the seminar is likely to take place in the medium to long term. Some more immediate results, such as plans to share research data and software, also arose from the discussions. As measurable outputs from the seminar, we expect to see several joint papers and applications for funding.

Beside the scientific aspect, the social aspect of our seminar was just as important. We had an interdisciplinary, international, and very interactive group of researchers, consisting of leaders and future leaders in our field. Many of our participants were visiting Dagstuhl for the first time and enthusiastically praised the open and inspiring setting. The group dynamics were excellent with many personal exchanges and common activities. Some scientists expressed their appreciation for having the opportunity for prolonged discussions with researchers from neighboring research fields – some thing that which is often impossible during conference-like events.

In conclusion, our expectations of the seminar were not only met but exceeded, in particular with respect to networking and community building. We would like to express our gratitude to the Dagstuhl board for giving us the opportunity to organize this seminar, the Dagstuhl office for their exceptional support in the organization process, and the entire Dagstuhl staff for their excellent service during the seminar. In particular, we want to thank Susanne Bach-Bernhard, Roswitha Bardohl, Marc Herbstritt, and Sascha Daeges for their assistance during the preparation and organizing of the seminar.

2 Table of Contents

Executive Summary

Meinard Müller, Elaine Chew, and Juan Pablo Bello 147

Stimulus Talks

Computational Music Structure Analysis (or How to Represent the Group’s Interests and Opinions in a Few Slides) <i>Juan Pablo Bello</i>	155
What You Hear and What You Must Make Others Hear <i>Elaine Chew</i>	156
Exposing Hierarchy Through Graph Analysis <i>Brian McFee</i>	156
Music Structure: Seeking Segmentations or Scenes? <i>Cynthia C. S. Liem</i>	157
Looking Beneath the Musical Surface <i>Christopher Raphael</i>	158
Computational Music Structure Analysis: A Computational Enterprise into Time in Music <i>Anja Volk</i>	159
Defining the Emcee’s Flow <i>Mitchell Ohriner</i>	160
Richard Wagner’s Concept of ‘Poetico-Musical Period’ as a Hypothesis for Computer-Based Harmonic Analysis <i>Rainer Kleinertz</i>	161
Large-Scale Structures in Computer-Generated Music <i>Mary Farbood</i>	162
A Composer’s Perspective on MIR <i>Carmine-Emanuele Cella</i>	163
Evolution and Salience <i>Geraint A. Wiggins</i>	163
Beat Tracking with Music Structure <i>Roger B. Dannenberg</i>	164
Can We Reach a Consensus on the Minimum Amount of Originality to Regard a Piece of Music as Original? <i>Masataka Goto</i>	165
Let’s Untie Our Hands! Use All the Data You Have and Stop Making Life Difficult <i>Mark Sandler</i>	166
Towards an Information-Theoretic Framework for Music Structure <i>Frédéric Bimbot</i>	167
Morpheus: Constraining Structure in Music Generation <i>Dorien Herremans, Elaine Chew</i>	168

Using Prior Expectations to Improve Structural Analysis: A Cautionary Tale <i>Jordan Smith</i>	169
Music Segmentation: of what, for what, for who <i>Xavier Serra</i>	170
Flexible Frameworks for the Analysis of Rhythm and Meter in Music <i>Andre Holzapfel</i>	171
Music Structure: Scale, Homegeneity/Repetition, Musical Knowledge <i>Geoffroy Peeters</i>	172

Further Topics and Open Problems

Musical Structure Analysis for Jazz Recordings <i>Stefan Balke, Meinard Müller</i>	173
On the Role of Long-Term Structure for the Detection of Short-Term Music Events <i>Juan Pablo Bello</i>	173
Mid-level Representations for Rhythmic Patterns <i>Christian Dittmar, Meinard Müller</i>	174
Representation of Musical Structure for a Computationally Feasible Integration with Audio-Based Methods <i>Sebastian Ewert</i>	175
Robust Features for Representing Structured Signal Components <i>Frank Kurth</i>	176
Reversing the Music Structure Analysis Problem <i>Meinard Müller</i>	176
Approaching the Ambiguity Problem of Computational Structure Segmentation <i>Oriol Nieto</i>	177
Multi-Level Temporal Structure in Music <i>Hélène Papadopoulou</i>	178
The Ceres System for Optical Music Recognition <i>Christopher Raphael</i>	180
Musical Structure Between Music Theory, Cognition and Computational Modeling <i>Martin Rohrmeier</i>	180
Accessing Temporal Information in Classical Music Audio Recordings <i>Christof Weiß, Meinard Müller</i>	181

Working Groups

Human-in-the-Loop for Music Structure Analysis <i>Participants of Dagstuhl Seminar 16092</i>	182
Computational Methods <i>Participants of Dagstuhl Seminar 16092</i>	183
Applications of Music Structure Analysis <i>Participants of Dagstuhl Seminar 16092</i>	184
Rhythm in Music Structure Analysis <i>Participants of Dagstuhl Seminar 16092</i>	185

Similarity in Music Structure Analysis	
<i>Participants of Dagstuhl Seminar 16092</i>	186
Structure in Music Composition	
<i>Participants of Dagstuhl Seminar 16092</i>	186
Structure Analysis and Music Cognition	
<i>Participants of Dagstuhl Seminar 16092</i>	187
Computation and Musicology	
<i>Participants of Dagstuhl Seminar 16092</i>	188
Music Structure Annotation	
<i>Participants of Dagstuhl Seminar 16092</i>	188
Participants	190

3 Stimulus Talks

3.1 Computational Music Structure Analysis (or How to Represent the Group's Interests and Opinions in a Few Slides)

Juan Pablo Bello (New York University, US)

License  Creative Commons BY 3.0 Unported license
© Juan Pablo Bello

In this talk, I have tried to identify areas of commonality and divergence in the abstracts submitted before the seminar, with the goal of stimulating and seeding the discussions during the Dagstuhl seminar.

First, I highlighted the wide range of applications of interest to the participants, including those driven by musicological concerns like the analysis of Jazz improvisations, performances of Wagner operas, metrical structures in Carnatic music, and notions of flow in Rap music; music information retrieval applications such as automatic rhythm transcription from recorded music, optical music recognition and understanding the structure of large music collections to characterize patterns of originality; and more creative applications in augmenting live music performances, algorithmic composition, the automatic creation of mashups and remixes, and improving the workflow of music production. Paraphrasing Serra's point, it is critical to understand the differences between the signal's properties, the application requirements, and the user context, and design solutions accordingly.

And yet, the abstracts highlighted many common issues that cut across applications. For example, the complex relational structure of music including multiple, hierarchical levels of information with strong interdependencies, and the multi-faceted nature of music information. Also the difficulty of defining notions of similarity and dissimilarity governing those relationships, and of creating representations of music information that are either invariant or sensitive to these patterns of similarity. It is generally agreed upon that current computational methods fail to capture this complexity.

A number of methods and ideas were proposed for addressing this: Markov logic networks, graph analysis, uncertainty functions, combinatorial optimization, statistical regularities in annotations, the use of multiple and rich streams of information, the use of domain-knowledge or the formulation of structural analysis as an information theoretical problem. More fundamentally, the abstracts emphasized the need to revise core assumptions in past literature, notably the shortcomings of the widespread ground-truth paradigm in the context of a task that is ambiguous and thus lends itself to multiple interpretations and disagreements. In this scenario, what should be the goal of computational approaches? To return all or multiple interpretations? Or at least one interpretation deemed to be coherent (or reasonable, plausible, interesting)? How do we define or benchmark coherence? Is this connected to narrative, flow, grammatical consistency? These are open questions that the community needs to address to move the field forward.

Finally, there is the question of the role of humans in the different stages of this process: data collection, computational modeling and benchmarking, and whether the computational task should be redefined as a way to gain insight on the human processing of musical structures, and even the modeling of individual responses.

3.2 What You Hear and What You Must Make Others Hear

Elaine Chew (Queen Mary University of London, GB)

License © Creative Commons BY 3.0 Unported license
© Elaine Chew

URL <http://elainechew-research.blogspot.com/2016/03/new-thoughts-on-piano-performance-sound.html>

*Drawing is not form, but a way of seeing form.
Drawing is not what you see, but what you must make others see.
Drawing is not form, it is your understanding of form.
– Edgar Degas (Gammell 1961, p.22)*

Music, and the performer or composer, works in much the same way. Alternative interpretations or hearings of musical form and structure almost always exist. These differences can be attributed to the listeners' state of knowledge, prior expectations, attention, and ontological commitment [1]. Given a particular hearing of musical structure, the performer can project this structure in performance through musical prosody [2]. The prosody of a musical communication strongly influences the listeners' parsing of its structure and meaning [3, 4]. In this talk, I gave numerous examples to show that structure emerges from performances, and that performances serve as a rich and largely untapped source of information about musical structure.

References

- 1 Smith, J. B. L., I. Schankler, E. Chew (2014). Listening as a Creative Act: Meaningful Differences in Structural Annotations in Improvised Performances. *Music Theory Online*, 20(3). http://www.mtosmt.org/issues/mto.14.20.3/mto.14.20.3.smith_schankler_chew.html
- 2 Chew, E. (2012). About Time: Strategies of Performance Revealed in Graphs. *Visions of Research in Music Education* 20(1). <http://www-usr.rider.edu/~vrme/v20n1/visions/Chew%20Bamberger%20.pdf>
- 3 Chew, E. (2016). Playing with the Edge: Tipping Points and the Role of Tonality. *Music Perception*, 33(3): 344-366. <http://mp.ucpress.edu/content/33/3/344>
- 4 Chew, E. (2016). From Sound to Structure: Synchronizing Prosodic and Structural Information to Reveal the Thinking Behind Performance Decisions. In C. Mackie (ed.): *New Thoughts on Piano Performance*. <http://elainechew-research.blogspot.com/2016/03/new-thoughts-on-piano-performance-sound.html>

3.3 Exposing Hierarchy Through Graph Analysis

Brian McFee (New York University, US)

License © Creative Commons BY 3.0 Unported license
© Brian McFee

Joint work of Juan Bello, Oriol Nieto, Dan Ellis, Brian McFee

In this talk, I presented a graphical perspective on musical structure analysis. While it is common to treat the boundary detection and segment labeling problems independently (or at least sequentially), viewing structure analysis as a graph partitioning problem can provide a unified formalization of both tasks, and motivate new algorithmic techniques. By varying the number of elements in the desired partition, it is possible to reveal multi-level or hierarchical structure in music. I concluded with a discussion of current challenges in both algorithm design and evaluation for hierarchical structure analysis.

References

- 1 Brian McFee, Oriol Nieto, and Juan Pablo Bello: Hierarchical Evaluation of Segment Boundary Detection. In Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR), Málaga, Spain, 406–412, 2015.
- 2 Brian McFee and Dan Ellis: Analyzing Song Structure with Spectral Clustering. In Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, 405–410, 2014.
- 3 Brian McFee and Daniel P. W. Ellis: Learning to segment songs with ordinal linear discriminant analysis, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 5197–5201, 2014.

3.4 Music Structure: Seeking Segmentations or Scenes?

Cynthia C. S. Liem (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Cynthia C. S. Liem

Joint work of Mark S. Melenhorst, Martha Larson, Alan Hanjalic, Cynthia C. S. Liem

When I was asked to give a Stimulus Talk at this Dagstuhl seminar, I reflected on my own past involvement with music structure analysis. As for me, this mostly fell into two categories: first of all, homework assignments for music theory courses back in conservatoire. Secondly, annotations I made for outcome visualizations of past research, in which developments over the course of a music performance (e.g. timing [5] and movement [2]) would partially relate to structure.

However, while I have always read work on automated music structure analysis in the community with interest, somehow I never felt the urge to work on the problem myself. Was it something I just took for granted, causing an interest mismatch similar to those collected within the community after a past Dagstuhl seminar [4]? I came to realize my own interests were not so much in localizing exact segment boundaries, but rather in the events happening in between such boundaries – and their contextualization with respect to various interpretations, at the performer and audience side.

In my talk, I reflected a bit more on this, discussing how notions of structure (even if ambiguous) can be a means to tackling higher-level, bigger questions on how music is realized and interpreted. I did this by discussing three topics:

- the role of structure in musical interpretation [5];
- the importance of narrative and linguistic event structure in music, when moving towards the connection of music to other media [3];
- the role of structure in assisting concert experiences of music audiences through digital interfaces [1].

In my stimulus talk, I particularly emphasized the second of these topics. Besides, following up on the third topic, in a separate demo session, I demonstrated the integrated prototype (see <http://www.phenicx.com>) of our recently concluded European PHENICX project.

References

- 1 Mark S. Melenhorst and Cynthia C. S. Liem. Put the Concert Attendee in the Spotlight. A User-Centered Design and Development approach for Classical Concert Applications. In *Proceedings of the 16th Conference of the International Society for Music Information Retrieval (ISMIR)*, pages 800–806, Málaga, Spain, 2015.

- 2 Cynthia C. S. Liem, Alessio Bazzica, and Alan Hanjalic. Looking Beyond Sound: Unsupervised Analysis of Musician Videos. In *Proceedings of the 14th International Workshop on Image and Audio Analysis for Multimedia Interactive services (WIAMIS 2013)*, Paris, France, 2013.
- 3 Cynthia C. S. Liem, Martha A. Larson, and Alan Hanjalic. When Music Makes a Scene – Characterizing Music in Multimedia Contexts via User Scene Descriptions.. *International Journal of Multimedia Information Retrieval*, 2(1): 15–30, March 2013.
- 4 Cynthia C. S. Liem, Andreas Rauber, Thomas Lidy, Richard Lewis, Christopher Raphael, Joshua D. Reiss, Tim Crawford, and Alan Hanjalic. Music Information Technology and Professional Stakeholder Audiences: Mind the Adoption Gap. In Meinard Müller, Masataka Goto, and Markus Schedl, editors, *Multimodal Music Processing*, Dagstuhl Follow-Ups, volume 3, pages 227–246. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2012. <http://www.dagstuhl.de/dagpub/978-3-939897-37-8>
- 5 Cynthia C. S. Liem, Alan Hanjalic, and Craig Stuart Sapp. Expressivity in musical timing in relation to musical structure and interpretation: A cross-performance, audio-based approach. In *Proceedings of the 42nd International AES Conference on Semantic Audio*, pages 255–264, Ilmenau, Germany, July 2011.

3.5 Looking Beneath the Musical Surface

Christopher Raphael (Indiana University – Bloomington, US)

License  Creative Commons BY 3.0 Unported license
© Christopher Raphael

In my talk, I discussed two interrelated music interpretation problems that are both essentially about musical structure. The first deals with the problem of rhythm recognition. While people effortlessly understand complex polyphonic rhythm from audio, in many cases it is nearly impossible for a person or algorithm to correctly perform rhythmic transcription on a sequence of uninflected taps or clicks. The slightest amount of inaccuracy or tempo change can render the sequence ambiguous and “lose” the listener. The reason is that the sequence of times leaves out a great deal of important information that is essential for organizing the music. An interesting formulation of the problem tries to transcribe rhythm given a sequence of onset times now labeled with their pitches, much like what one gets from MIDI stream. The essential human strategy seems to employ a fundamental assumption: when we hear the “same” thing it usually falls in the same position in the measure, beat or other unit in the rhythmic hierarchy. This “same” thing could be a long note, a particular pitch configuration, a short pattern such as a dotted rhythm, etc. This is, of course, the basis for structural analysis by autocorrelation. While this idea is a powerful heuristic it is hard to formulate generally. People are good at recognizing many kinds of variations on the “same” thing though they can involve transformations on many different musical axes, such as leaving out or adding decorative notes, shifting by an interval or chord inversion, contour inversion, reharmonization, etc. The musical understanding seems to require that we identify these variations as a kind of repetition in order to mentally organize the material at an equivalent rhythmic location in the rhythmic structure (e.g. measure). Perhaps the recognition requires that we learn these building blocks or motives in addition to the rhythm we seek to identify. Thus we do not come to each new piece of music we try to understand with a fixed model for rhythm and pitch, but rather recognize by adapting a flexible model to the data and recognizing simultaneously. For instance, a possible recognition strategy could model the

music as a sequence of measures with one of several possible types. Standard HMM parameter estimation strategies could be employed to learn the pitch/rhythm tendencies of each of these possible types while parsing the data according to these types. As we seek parsimonious explanations we penalize our models to favor few types of measures. Aside from any specific formulation, the essential idea is to recognize and model the music simultaneously.

An interesting and related problem is that of performing structural and motivic analysis on a simple folk song, carol, anthem, etc, given in score form (rather than performance data). While there could be many motivations for such analysis, one would be to create algorithms that compose simple music that “makes sense” – this might avoid the repetitive music in many computer games, instead providing an inexhaustible supply. As always, there is a close relationship between analysis and synthesis, so it isn’t too far fetched to suppose that an analysis engine could form the heart of an algorithmic composition system. Analysis of even the simplest familiar folk song shows a tight reuse of figures and motives at various time scales, while allowing for many possible variations. An algorithmic analysis might try to represent the music in terms of simple building blocks, encoding the music in terms of these basic figures as well as the necessary transformations that lead to the actual notes and rhythm. One could view this approach as analysis by compression, where we seek the minimal number of bits necessary to represent the music. We doubt the power of a strict Lempel-Ziv compression seeking to represent music in terms of literal repetition, since repetition in music is often not exact repetition. Rather one must learn the basic musical objects as well as the transformations applied to them that lead to the musical surface. In formulating the problem scientifically, it may begin with the familiar parsing through a probabilistic context free grammar, though, as before, we do not believe in a generic grammar for music. Rather, as with the previous problem, each piece has its own grammar which must also be learned as part of the recognition strategy. How could one formulate this problem in a way that is both musically plausible and tractable?

3.6 Computational Music Structure Analysis: A Computational Enterprise into Time in Music

Anja Volk (Utrecht University, NL)

License © Creative Commons BY 3.0 Unported license
© Anja Volk

In this talk I addressed the relation between computational music structure analysis and the study of time in music. Computational music structure analysis requires the modeling of time processes in music. Structures such as segments, salient patterns, rhythmic-metric structures, and the like are inferred from either symbolic or audio musical content, taking temporal information into account. Often it is crucial for the success of a computational model for solving a specific task within music structure analysis *what* temporal information is taken into account, such as whether large scale or local temporal information is considered. However, even from the perspective of the human information processing in music, we often know only very little about what temporal information we need to consider in a certain context. While music has been argued to be the “art of time,” most theories of music have been concerned predominantly with pitch and not with time, such that we are far from understanding the different functions of time and temporal information in music [5].

Time is not only a challenging concept in music analysis, but in many disciplines: what is time, how do we perceive time and how do we successfully employ time in our interactions

with the world? While we do not have a sensory organ for perceiving time (as we have, for instance, for colors), studies in cognition demonstrate that the auditory domain is superior to other domains in processing temporal information [4], backing up that it is worthwhile to investigate the different relations between music and time. What is the relation between time as employed in the musical structure and our experience of time when we listen to and make sense of music?

Understanding this relation would help us to employ temporal information in computational music structure analysis in such a way as to find meaningful elements of musical structure, such as temporal building blocks. Current challenges regarding temporal information and music structure analysis discussed during the seminar link to the question of what is the interrelations between different temporal scales in music, such as between temporal microstructure [3] (as studied in expressive timing), small-to-medium-scale temporal phenomena (as studied in the area of rhythm and meter [6]), and medium-to-large scale temporal phenomena (as studied in theories of musical form [2])? For instance, as discussed during the seminar, automatically generating large scale music structures provides an unsolved issue in automatic composition. In the area of rhythm and meter, we know little about the interaction between rhythmic structures and other parts of musical structures, such as melodic patterns. I discussed examples on the role of time for improving music structure analysis for tasks such as repetition-based finding of segments [7], and the discovery of shared patterns in a collection of Dutch folk songs [1]. Hence, computational music structure analysis can help to elucidate the role of time in music for recognizing structure, as well as it will benefit from a better understanding of the human processing of time in music.

References

- 1 Boot, P., Volk, A., and de Haas, B. (2016). Evaluating the Role of Repeated Patterns in Folk Song Classification and Compression. Submitted for publication.
- 2 Caplin, W. E., Hepokoski, J., and Webster, J. (2009). *Musical Form, Forms and Formenlehre: Three Methodological Reflections*, P. Bergé (ed.), Cornell University Press, New York.
- 3 Clarke, E. (1999). Rhythm and timing in music. In Deutsch, D.: *The Psychology of Music*, second edition, University of California, San Diego, 473–500.
- 4 Grahn, J. A., Molly J. Henry, J. H., and McAuley, J. D. (2011). MRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. In *NeuroImage*, 54:1231–1243.
- 5 Kramer, J. D. (1988). *The Time of Music*. New York: Schirmer Books.
- 6 London, J. (2004). *Hearing in Time*, Second Edition. Oxford University Press.
- 7 Rodriguez-Lopez, M. E. and Volk, A. (2015). Location Constraints for Repetition-Based Segmentation of Melodies. In *Proceedings of the 5th International Conference in Mathematics and Computation in Music (MCM)*, 73–84.

3.7 Defining the Emcee’s Flow

Mitchell Ohriner (Shenandoah University – Winchester, US)

License  Creative Commons BY 3.0 Unported license
© Mitchell Ohriner

In discourse on rap music, the word “flow” takes on several meanings. On the one hand, the flow of a verse is understood as all the musical features of the emcees rapping that verse—rhythms, accents, rhyme, articulation, etc. Flow might also refer to less musical

features such as word choice, topic, or reputation. On the other hand, emcees and critics alike attest that individual emcees have a distinctive flow that transcends individual verses or tracks. At the Dagstuhl seminar, using the emcee Black Thought as an example, I presented a method for locating an emcee in a feature space through a large corpus of his verses in comparison to another corpus representative of rap music generally.

References

- 1 Adams, K. (2009). On the metrical techniques of flow in rap music. *Music Theory Online* 15(5).
- 2 Adams, K. (2015). The musical analysis of hip-hop. In Justin Williams (ed.), *The Cambridge Companion to Hip-Hop*, pp. 118–135. Cambridge University Press.
- 3 Condit-Schultz, N. (2016). MCFLOW: A digital corpus of rap transcriptions. *Empirical Musicology Review* 11(1).
- 4 Kautny, O. (2015). Lyrics and flow in rap music. In Justin Williams (ed.), *The Cambridge Companion to Hip-Hop*, pp. 101–117. Cambridge University Press.
- 5 Krims, A. (2004). The musical poetics of a ‘revolutionary’ identity. In *Rap Music and the Poetics of Identity*. Cambridge University Press.
- 6 Ohriner, M. (2013). Groove, variety, and disjuncture in the rap of Eminem, André, and Big Boi. Paper present at the Annual Meeting of the Society for Music Theory, November 2, 2013, Charlotte, NC USA.
- 7 Ohriner, M. (2016). Metric ambiguity and flow in rap music: A corpus-assisted study of Outkast’s ‘Mainstream’ (1996). *Empirical Musicology Review* 11(1).

3.8 Richard Wagner’s Concept of ‘Poetico-Musical Period’ as a Hypothesis for Computer-Based Harmonic Analysis

Rainer Kleinertz (*Universität des Saarlandes, DE*)

License © Creative Commons BY 3.0 Unported license
© Rainer Kleinertz

Joint work of Meinard Müller, Christof Weiß, Rainer Kleinertz

In the third part of his large theoretical work *Oper und Drama* (Leipzig 1852), Wagner developed the idea of a ‘poetico-musical period.’ Based on the *drama* (i.e. text and action), he tries to motivate modulations: The *musician* (composer) would receive an incitement to step outside the once selected key only when an opposite emotion occurs (e.g., “Die Liebe bringt Lust und Leid.”). When this new, opposite emotion returns to the original emotion (e.g., “Doch in ihr Weh auch webt sie Wonnen.”), then harmony would return in the original key. In Wagner’s eyes the most perfect artwork would be that, in which many such ‘poetico-musical periods’ – as he calls them – present themselves “in utmost fulness.”

These pages on his *Drama of the Future* were applied to Wagner’s musical dramas by Alfred Lorenz in his highly influential study *Das Geheimnis der Form bei Richard Wagner* (4 vols., Berlin 1924–1933). Among others, Lorenz analyzed the entire Ring as a series of such periods. In the 1960s, Carl Dahlhaus rejected Lorenz’ analyses as being completely erroneous and against Wagner’s musical ideas. In 2002 Werner Breig postulated that the concept should be ignored, as Wagner – when he coined it – had not yet composed a single note of the Ring, and even the texts of *Rheingold*, *Die Walküre*, and *Siegfried* did not yet exist. My own hypothesis – as published in [2] – is that Wagner had indeed something in mind which he realized at least partly in his subsequent Ring composition: The concept of the ‘poetico-musical period’ serves to describe more or less ‘closed’ parts of the Ring in which

a strong emotional change motivates modulations leaving the original key and returning to it. As a paradigmatical example for such a ‘poetico-musical period’ may serve Sieglinde’s narration in the first act of *Die Walküre* with its two interior modulations out of and back to the framing tonality of E minor. Consequently, the ‘poetico-musical period’ should not be regarded as a mere ‘way’ of music between a certain tonality and its return, but as a harmonic construct around a central modulation.

This musical-philological assumption of what Wagner may have had in mind when he wrote *Oper und Drama* may serve as a meaningful hypothesis for computer-based harmonical analysis. In a current research project of the Deutsche Forschungsgemeinschaft (DFG) – a cooperation of Meinard Müller’s group in Erlangen and Rainer Kleinertz’ group in Saarbrücken – harmonic analysis of the entire Ring based on audio data may allow a verification or falsification of this hypothesis. This approach may become a paradigm for a cooperation between historical musicology and computer science where the fundamentally different methods of both disciplines are applied in favour of new *objectified* results. Hermeneutical-musicological understanding and computer-based proceedings would allow new insights in complex musical works.

References

- 1 Verena Konz, Meinard Müller and Rainer Kleinertz (2013). A Cross-Version Chord Labelling Approach for Exploring Harmonic Structures – A Case Study on Beethoven’s *Appassionata*, *Journal of New Music Research*, DOI: 10.1080/09298215.2012.750369
- 2 Rainer Kleinertz (2014). Richard Wagners Begriff der ‘dichterisch-musikalischen Periode’, *Die Musikforschung* 67, pp. 26–47.

3.9 Large-Scale Structures in Computer-Generated Music


Mary Farbood (*New York University, US*)

License  Creative Commons BY 3.0 Unported license
© Mary Farbood

The generation of musical structure is closely related to (and in most cases dependent on) the analysis of structure. Short-term musical structure has been effectively modeled in many algorithmic composition systems ranging from those that compose music off-line such as David Cope’s EMI to improvisatory systems such as François Pachet’s Continuator. These systems often use Markov models or some type of generative grammar to create longer sections of music from shorter segments or individual notes. The success of these systems in composing convincingly human-sounding music is dependent on how well lower-level generation and assembly of smaller segments make stylistic sense. Implementing a system that produces convincing computer-generated music formalizes (at least to some extent) how listeners intuitively perceive style. However, computer-generated music that is coherent and interesting on a large-scale structural level is very difficult to achieve. No system currently exists that produces aesthetically compelling music (at least from a structural perspective) over long time spans. At Dagstuhl, we discussed this problem in the context of cognitive constraints and aesthetic considerations.

3.10 A Composer’s Perspective on MIR

Carmine-Emanuele Cella (École normale supérieure, F)

License  Creative Commons BY 3.0 Unported license
© Carmine-Emanuele Cella

A sound transformation is, in a general sense, any process that changes or alters a sound in a significant way. Transformations are closely related to representations: each action is, indeed, performed on a specific representation level. For example, a time stretch performed in the time domain by means of granular synthesis gives inferior results (perceptually) to the same transformation performed in the frequency domain, where one has access to phase information. In the same way, a pitch shift operated in the frequency domain gives inferior results to the same operation performed using a spectral envelope representation, where one has access to dominant regions in the signal.

In the two cases discussed above, we passed from a low-level representation (waveform) to a middle-level representation (spectral envelope). We could, ideally, iterate this process by increasing the level of abstraction in a representation, thus giving access to specific properties of sound that are perceptually relevant; by means of a powerful representation it could therefore be possible to access a semantic level for transformations.

An example will clarify the ideas outlined: suppose we want to elongate the minor chords present in a sound by a certain factor, but only if they are located in non-transient regions. At the same time, we want to pitch shift them by some semitones, but only if they are played by a piano. Obviously, this kind of textual description is very easy to understand by humans, but extremely difficult to code in an algorithm. We envision, therefore, the possibility in the future to have such kind of semantic transformations.

3.11 Evolution and Saliency

Geraint A. Wiggins (Goldsmiths, University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Geraint A. Wiggins

My work is currently focused on a cognitive architecture that is intended to explain the structuring of sequential and semantic information in the mind/brain. Because it is fundamentally sequential, it is directly applicable to music and language. One aspect of the approach is that it ought to explain what we mean by “saliency” in our musicological discussions, in line with the paradigmatic analytical approach of Ruwet, and the related psychological theory of cue abstraction, due to Deliège [1]. There are several ways to look at music, and they can tell us different things. One way is via evolution. When I consider music this way, in the context of the information-processing view of mind, I am led to the notion of saliency, which is something we don’t often discuss in MIR.

References

- 1 G. A. Wiggins. Cue abstraction, paradigmatic analysis and information dynamics: Towards music analysis by cognitive model. *Musicae Scientiae*, Special Issue: Understanding musical structure and form: papers in honour of Irène Deliège, pages 307–322, 2010.

3.12 Beat Tracking with Music Structure

Roger B. Dannenberg (Carnegie Mellon University, US)

License  Creative Commons BY 3.0 Unported license
© Roger B. Dannenberg

I want to encourage more thinking about music structure as the “missing link” in music understanding. The MIR community has made great progress in using larger datasets, machine learning, and careful search and optimization to improve the performance of many music understanding tasks, including beat tracking. However, our algorithms still make obvious and silly mistakes, at least when viewed from a human perspective. My sense is that most current algorithms for music understanding tasks are very selective about the information they use. The information is highly effective for most problems, which is why systems work at all, but when much of the available information is ignored, algorithms can never be robust.

I believe that music is intentionally confusing and created in a way that demands attention to many different aspects of rhythm, melody, and harmony. It is not clear why music works this way, but it seems natural that music would exercise human intelligence, and the brain does seem to enjoy making sense out of things, especially when they are not completely obvious. If this is the way music works, then we must think about integrating many sources of information in order to get our machines to understand music. Music understanding is largely a problem of finding patterns and structure in music. For example, if we can identify the structure of conventional music in terms of phrases and measures, most of the work of beat tracking is done.

With this premise in mind, let us consider how we might use non-beat information to help with the beat-tracking problem. This approach is mainly a review of an earlier ISMIR paper [1], considered here in the context of this seminar on music structure. Nearly all beat trackers optimize two basic things:

- Beats and “beat features” are highly correlated. In other words, something in the signal (amplitude change, increase in high frequencies, spectral difference) indicates where beats are likely to occur.
- Tempo is fairly steady. In other words, the spacing between nearby beats is about the same.

These constraints are expressed differently in different beat tracking algorithms, but seem to be the core principles.

I propose a third principle based on music structure that can provide additional information: Where repetitions occur in music, the beats in the two repetitions should correspond. For example, suppose a beat tracker labels a verse correctly at the beginning of a song and the verse repeats later. Rather than relabeling the same music (and possibly getting it wrong), this constraint tells us that the beats in the two sections should correspond; thus, we have essentially labeled the beats just by identifying the structural repetition. Alternatively, we can process these two repetitions as one. Then, we have twice as much information to help us find the beats. Of course, the information may be conflicting, but generally it should be easier to find one solution than two.

An implementation of this approach uses a self-similarity matrix based on chroma features to find repetition in music. The result of this step is a set of continuous mappings from one interval of time to another. There is one mapping for each repetition that is discovered. Then, each time a beat is proposed at time t , we simply map t to all the places this beat should be repeated and propose there must be beats at those locations as well. In practice,

chroma-based alignments do not have high time resolution, so the mappings are not very useful for placing single beats. However, we can modify the constraint to say that tempo is consistent in repeated sections. Thus, when we place several measures of beats in one location, we can use the mappings to assert the tempo in other locations. The tempo may not be identical due to slight tempo changes during the performance, but the derivatives of the mappings can be used to estimate the tempo change.

In practice, these constraints have been imposed through a gradient descent algorithm. Essentially, a few beats are placed according to initial guesses based on the autocorrelation of beat features. These guesses are replicated according to repetitions in the music. Then, beat locations are adjusted to optimize the combination of all three constraints: beats correspond to beat features, tempo is steady, and tempo is consistent across repetitions. The algorithm continues by alternately proposing a few more beats (based on the steady-tempo principle) and then optimizing using gradient descent, until the entire piece is covered by beats.

This is only one approach, presented to motivate thinking about ways we can use music structure in music processing tasks. One of the problems with this approach in general is that while structure can offer helpful information, the information can also be wrong and misleading. Building more holistic algorithms does not guarantee improvement over simpler approaches that benefit from being more tractable and more amenable to training with large datasets. Another challenge is to jointly estimate structure along with information such as tempo, beats, rhythm, and harmony.

References

- 1 Roger B. Dannenberg. *Toward Automated Holistic Beat Tracking, Music Analysis, and Understanding* in Proceedings of the 6th International Conference on Music Information Retrieval Proceedings (ISMIR), London, pages 366–373, 2005.

3.13 Can We Reach a Consensus on the Minimum Amount of Originality to Regard a Piece of Music as Original?

Masataka Goto (AIST – Tsukuba, JP)

License  Creative Commons BY 3.0 Unported license
© Masataka Goto

In the age of digital music, future musicians may find it more difficult to be truly original in the face of ever-expanding archives of all past music. The amount of digital musical pieces that can be accessed by people has been increasing and will continue to do so in the future. Since the amount of similar musical pieces is monotonically increasing, musicians will be more concerned that their pieces might invite unwarranted suspicion of plagiarism. All kinds of musical pieces are influenced by existing pieces, and it is difficult to avoid the unconscious creation of music partly similar in some way to prior music. The monotonic increase in musical pieces thus means that there is a growing risk that one’s piece will be denounced as being similar to someone else’s.

To address this issue, we started a research project called OngaCREST [1] to build an information environment in which people can know the answers to the questions “What is similar here?” and “How often does this occur?” Although human ability to detect musical similarity and commonness (typicality) [2] is limited for a large-scale music collection, future advanced technologies would enable people to compute musical similarity between any pairs of musical pieces and musical commonness of a musical piece to a set of pieces.

Once such technologies could be available to compute musical similarity and commonness in detail, people could naturally understand that any musical piece has similar pieces with regard to some aspects. The concept of originality would then be discussed in a more quantitative way and might be revised. If some (or most) aspects are always similar, how can we measure the amount of originality? To be regarded as an original musical piece, how many different aspects or elements should it have? Can we, as a global society, reach a consensus on the minimum amount of originality to regard a piece of music as original?

References

- 1 Masataka Goto. *Frontiers of Music Information Research Based on Signal Processing*, Proceedings of the 12th IEEE International Conference on Signal Processing (ICSP), pages 7–14, October 2014.
- 2 Tomoyasu Nakano, Kazuyoshi Yoshii, and Masataka Goto. *Musical Similarity and Commonness Estimation Based on Probabilistic Generative Models*, Proceedings of the IEEE International Symposium on Multimedia (ISM), pages 197–204, December 2015.

3.14 Let’s Untie Our Hands! Use All the Data You Have and Stop Making Life Difficult

Mark Sandler (*Queen Mary University of London, GB*)

License © Creative Commons BY 3.0 Unported license
 © Mark Sandler
URL <http://www.semanticaudio.ac.uk/>

At the Dagstuhl seminar, I said a few words on music structure analysis in the context of a large project called “Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption” (FAST-IMPACT) I am coordinating. Thinking about musical structure, one needs to ask, who is this for? Is it for the professional in the studio (a strong focus of FAST-IMPACT) or the consumer, or even for some intermediary needing to make money somehow from the content – hopefully on behalf of the artists and creators? The needs of different categories of user are very different. For the producer it is probably connected with navigation around a particular piece of music in an ongoing project. For the consumer/listener this could be true, but there is potentially the added need to navigate within collections. (As I write I realise that the latter is also true for professionals, though the collections are different!) Taking a step back, we can say that for many musics, what we need to do to help these participants is analyse the audio signal and extract meaningful, and above all, useful information from the audio. We need to do this to the best of the capabilities of the available technologies and algorithms. Everyone would find this hard to dispute, I think. Yet, why do we all, to my knowledge with zero (or close to) exception, make use of anything but a monophonic down mix? I would therefore propose that we start to investigate ways that use the maximum amount of data and information available to us, and to stop making our investigations overly and unnecessarily difficult. I would start with stereo signals – which I see some researchers describe as ‘legacy’!

3.15 Towards an Information-Theoretic Framework for Music Structure

Frédéric Bimbot (CNRS/IRISA, Rennes, FR)

License © Creative Commons BY 3.0 Unported license
© Frédéric Bimbot

Music is a communication signal and the estimation of music structure is essentially an information-theoretic problem. The structure S of a music content M can be understood as the “proper type” and “right quantity” of latent side information which provides an economical explanation of M by minimizing $Q(M, S)$, i.e., the quantity of information needed to jointly describe M and S . Two philosophies can support the definition of Q (see for instance [4]):

- Shannon’s Information (SI), also called lossy source-coding scheme, which relates information Q to the distortion of M with respect to a prototypical structure itself derived from a probabilistic model of all possible structures, and
- Kolmogorov’s Complexity (KC), sometimes referred to as algorithmic compressibility, which considers M as the output of a short, standalone program (within a class of valid structure generating programs), whose size is related to Q .

Shannon’s approach is fundamentally a knowledge-based (inter-opus) approach, where statistically typical forms provide templates that guide the recognition of music content organization (stylistic structure). Kolmogorov’s framework is rather based on a data-driven (intra-opus) viewpoint and focuses on internal redundancy as a primary criterion for grouping musical material into consistent structural patterns (“semiotic” structure [1, 2]). Both conceptions of information are meaningful, but understanding and exploiting their interaction remains a fundamental scientific bottleneck – in MIR, in Computational Musicology, and also in many other scientific domains. The duality between SI and KC in music is for instance illustrated by Schenker’s [10] versus Narmour’s [7, 8] conceptions of music structure, and KC approaches are becoming increasingly popular in MIR (see for instance [5]).

However, current approaches in Music Structure Analysis [6] fail in explicitly accounting for both aspects simultaneously, even though they are presumably present with a different balance across musical genres (this could be one of the causes of ambiguities in human perception of structure [11]). Note that, even though, neither SI nor KC can actually be calculated exactly, they can be estimated using models, i.e. family of distributions for SI such as Hidden Markov Models (see for instance [9]) and classes of programs for KC (as prefigured by the System & Contrast Model [3]). Approaching the diverse views of music structure within the common framework of Information Theory appears as a relevant move towards a better understanding of what music structure is, but also as a key for a more efficient use of music structure in computational contexts. Though music is not a uniquely decodable code, it can be assumed that the number of reasonable structural hypothesis is sufficiently limited so as to be tackled in a relatively unified framework, encompassing the two main facets of data compression (SI and KC). By understanding the interaction between the “two sides of a same coin,” music could become a case study that would help bridging a fundamental gap in Information Sciences.

References

- 1 Bimbot, F., Deruty, E., Sargent, G., Vincent E. (2012). Semiotic Structure Labeling of Music Pieces: Concepts, Methods and Annotation Conventions. Proc. International Society on Music Information Retrieval Conference (ISMIR), pages 235–240, Porto.

- 2 Bimbot, F., Sargent, G., Deruty, E., Guichaoua, C., Vincent, E. (2014). Semiotic Description of Music Structure: An Introduction to the Quaero/Metiss Structural Annotations. Proc. 53rd AES Conference on Semantic Audio, 12 pages, London.
- 3 Bimbot, F., Deruty, E., Sargent, G., Vincent, E. (2016). System & Contrast: A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces. *Music Perception*. 41 pages, to appear.
- 4 Grünwald, P., Vitanyi, P. (2004). Shannon Information and Kolmogorov Complexity. arXiv preprint [cs/0410002](https://arxiv.org/abs/cs/0410002). 2004, updated 2010.
- 5 Meredith, D. (2012). Music Analysis and Kolmogorov Complexity. Proc. XIX CIM.
- 6 Müller, M. (2015). Music structure analysis. In *Fundamentals of Music Processing*, chapter 4, pages 167–236, Springer Verlag.
- 7 Narmour, E. (1977). *Beyond Schenkerism*. University of Chicago Press.
- 8 Narmour, E. (2000). Music expectation by cognitive rule-mapping. *Music Perception*, XVII/3, pages 329–398.
- 9 Pauwels, J., Peeters, G. (2013). Segmenting music through the joint estimation of keys, chords and structural boundaries. Proc. 21st ACM International Conference on Multimedia, New York.
- 10 Schenker, H. (1935). *Der freier Satz*, Universal, Vienna.
- 11 Smith, J. (2014). *Explaining Listener Differences in the Perception of Musical Structure*. PhD thesis, Queen Mary University of London, UK.

3.16 MorpheuS: Constraining Structure in Music Generation

Dorien Herremans, Elaine Chew (Queen Mary University of London, GB)

License © Creative Commons BY 3.0 Unported license
© Dorien Herremans, Elaine Chew

A major problem with much of the automatically generated music is that it lacks a structure and long-term coherence. We have defined the music generation problem as a combinatorial optimization problem [2, 5]. The advantage of this approach is that it gives us the freedom to impose both hard and soft constraints. These constraints can be used to define different types of structure.

One example of a structure that can be imposed by hard constraints is based on repeated and transposed patterns. The *cosiatec* pattern detection algorithm [3] was used to find maximum translatable patterns. These patterns were then used to constrain the output of a music generation algorithm called *MorpheuS* (<http://dorienherremans.com/software>).

A second form of structure, which is soft constrained, is a tension profile. This type of tension could be relevant to, for instance, automatic generation of game or video music. We have developed a model [4] that captures aspects of tonal tension based on the spiral array [1], a three dimensional model for tonality. Our approach first segments a musical excerpt into equal length subdivisions and maps the notes to clouds of points in the spiral array. Using vector-based methods, four aspects of tonal tension are identified from these clouds. First, the cloud diameter measures the dispersion of clusters of notes in tonal space. Second, the cloud momentum measures the movement of pitch sets in the spiral array. Third, the tensile strain measures the distance between the local and global tonal context. Finally, the cosine similarity measures the directional change for movements in tonal space.

The results of generating polyphonic piano music with constrained patterns and fit to a tension profile are very promising and sound musically interesting. The reader is invited to listen to full pieces generated by the algorithm at <http://dorienherremans.com/Morpheus>.

References

- 1 Chew, Elaine (2014). *The Spiral Array*. Mathematical and Computational Modeling of Tonality. Springer, pages 41–60.
- 2 Herremans, Dorien, and Sørensen, Kenneth (2013). *Composing fifth species counterpoint music with a variable neighborhood search algorithm*. Expert Systems with Applications 40(16):6427–6437.
- 3 Meredith, David (2013). *COSIATEC and SIATECCompress: Pattern discovery by geometric compression*. in Music Information Retrieval Evaluation Exchange (Competition on “Discovery of Repeated Themes & Sections”) of the International Society for Music Information Retrieval Conference.
- 4 Herremans, D., and Chew, E. (2016). *Tension ribbons: Quantifying and visualising tonal tension*. Second International Conference on Technologies for Music Notation and Representation (TENOR). Cambridge, UK.
- 5 Herremans, D., Weisser, S., Sørensen, K., and Conklin, D. (2015). *Generating structured music for bagana using quality metrics based on Markov models*. Expert Systems with Applications 42(21):7424–7435.

3.17 Using Prior Expectations to Improve Structural Analysis: A Cautionary Tale

Jordan Smith (AIST – Tsukuba, JP)

License © Creative Commons BY 3.0 Unported license
© Jordan Smith


Joint work of Masataka Goto, Jordan Smith

Annotations of musical structure tend to have strong regularities: average segment size is roughly 20 seconds, the number of segments per annotation is roughly 12, and segments usually have a uniform size, meaning that the average ratio of segment length to the median segment length is very close to 1. These regularities are consistent even between different collections of annotations, but are not often used by algorithms to refine estimates. By treating these regularities as prior expectations, we can use a committee-based approach to structural analysis: first, make several estimates using a variety of algorithms; second, choose the estimate that is likeliest given the prior distributions. Although the method may seem like ‘cheating’ (even when appropriate leave-one-out training regimes are followed), the approach is guaranteed to give at least a modest gain in f-measure.

Except, we tried it, and it didn’t work. Why not? We are still trying to decide.

3.18 Music Segmentation: of what, for what, for who

Xavier Serra (UPF – Barcelona, ES)

License  Creative Commons BY 3.0 Unported license
© Xavier Serra

Within the field of Computational Musicology, we study music through its digital traces, or digital artifacts, that we can process computationally. Ideally we want to start from data that is as much structured as possible and the goal is to extract musical knowledge from it. However most current computational research is still focused on trying to increase the structuring of the existing data by computational means. Music segmentation is a key structuring element and thus an important task is to do this as automatically as possible.

Most, if not all, music data processing problems, and music segmentation is no exception, should be approached by taking into consideration the following issues: What signal and music are we processing? What application are we aiming at? Who is the user and context being targeted?

Each type of music signal (including audio, symbolic scores, lyrics, etc.) requires different segmentation methodologies and implies different segmentation concepts. Each musical facet (including timbre, melody, rhythm, harmony, etc.) also requires a different music segmentation strategy and can be used for different tasks.

The targeted application of a music segmentation process is also critical. It is very different wanting to perform music analysis for music understanding or wanting to solve some engineering task-driven problem. It is nice when a task-driven problem can be based on a musically grounded concept, but it is not always possible, nor even adequate.

Personal subjectivity, cultural bias, and other contextual issues greatly affect the concept of music segmentation. The analysis approach has to take that into account and assume that the results obtained should be different depending on the context being targeted.

In general, the concept of music segmentation means many things, even within the MIR community. Maybe the most common meaning relates to music structure, which is a musically grounded concept. But strictly speaking, practically all music processing tasks have an implicit or explicit segmentation, and this segmentation has a big impact on the results obtained. We process music signals by first segmenting them into discrete events, such as audio samples, audio frames, symbolic notes, phrases, songs, and so on. We use a multiplicity of segments with more or less agreed definitions and standard approaches to be computed. Clearly any particular piece of music has many possible multilevel segmentations that might be of use for different applications and different contexts.

As a summary I want to emphasize that we cannot talk about music segmentation without taking into account

- the type of signals we start from,
- the targeted application, and
- the particular context in which the signal and application is part of.

In the project CompMusic (see <http://compmusic.upf.edu>), we have worried about these three issues. Let me go through what has been our approach.

With respect to the issue “of what,” we have focused on the musical repertoires of five music cultures: Hindustani, Carnatic, Turkish-makam, Beijing Opera, and Arab-Andalusian. The data we have been processing has been mainly audio recordings, scores, and editorial metadata. We also have extracted audio features from the audio recordings that are then used as inputs to other processing tasks. Segmenting the audio and the scores in a unified way has been an important task in some of the music repertoires.

With respect to the issue “for what,” our main goal has been to develop tools and systems with which to explore the selected musical repertoires. *Dunya* (see <http://dunya.compmusic.upf.edu>) is one of the prototypes we have developed in order to evaluate the developed technologies, a prototype that can be used to explore the music collections we have compiled and with it you can listen to music pieces while visualizing information that can help the understanding and enjoyment of the music.

With respect to the issue “for who,” we have aimed at developing tools that can be of interest to the music lovers of each of the music traditions we have been studying. Thus we target people with some knowledge of the music they are exploring and listening to.

3.19 Flexible Frameworks for the Analysis of Rhythm and Meter in Music

Andre Holzapfel (OFAI-Wien, AT)

License  Creative Commons BY 3.0 Unported license
© Andre Holzapfel

In a recent study [1] on metered Turkish makam music performance, we illustrated differences in the ways notes are distributed in compositions from different cultures. Based on these findings, we are now able to track meters that go beyond simple meters in 4/4 and 3/4. However, our evaluation measures are tailored towards simple/symmetric meters, our evaluation data is limited in style, annotations are mostly from one annotator, and in most cases only the beat is annotated, while other metrical layers are ignored. Recent developments both in Bayesian Networks and Deep Neural Networks push the state of the art in meter tracking to a new level. However, how far can we go given the limited amounts of annotated data, and possibly more importantly, the limited amount of understanding of musics that we engineers have about the diverse structures we aim to subject to a correct analysis? We developed a Bayesian framework [2] for meter tracking in music that is able to track meter, given a small amount of annotated representative samples. A Bayesian framework allows to adapt the model to new features, and to different types of tempo and meter properties. In a discussion of important steps to take in future, I would like to emphasize that including a complete set of observables into account is highly timely; Learning meter in Indian music without looking at performers seems odd. Furthermore, music performances are shaped by humans who move and breathe together, and the aspects in which their various biosignals correlate remain widely unknown. In short, we need methodologies for the sophisticated observation of performance events. But still, they will not reveal the meanings and mental representations that these structures are evoke in various contexts. I believe that in this aspect interdisciplinary collaborations between music psychology, engineering, and ethnomusicology can indicate promising directions that go beyond observationalism towards a more complete understanding of music, driven by sophisticated computational analysis.

References

- 1 Holzapfel, A. (2015). Relation between surface rhythm and rhythmic modes in Turkish makam music. *Journal for New Music Research* 44(1):25–38.
- 2 Krebs, F., Holzapfel, A., Cemgil, A. T., and Widmer, G (2015). Inferring metrical structure in music using particle filters. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 23(5):817–827.

3.20 Music Structure: Scale, Homogeneity/Repetition, Musical Knowledge

Geoffroy Peeters (UMR STMS – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Geoffroy Peeters

Joint work of Florian Kaiser, Johan Pauwels, Geoffroy Peeters

In my talk at the Dagstuhl seminar, I discussed three aspects of automatic music structure estimation.

The first aspect relates to the temporal scale considered when trying to estimate the structure (i.e., the duration of the segments). This scale is usually a priori unknown. To solve this issue, we proposed in [1] to use a multi-scale approach in which a set of checkerboard-kernels of increasing size is used to segment a given self-similarity matrix.

The second aspect relates to the underlying process that creates the structure. Currently two major assumptions are used: homogeneity/novelty and repetition [3] leading to the so-called “state” and “sequence” approaches [5]. Also, this underlying process is typically a priori unknown. To solve this issue, we proposed in [2] a joint estimation based on the two assumptions leading to a large increase in the estimation results.

Finally, I discussed how musical structure can be estimated exploiting musical knowledge. As an example, I reviewed our work [4] on the joint estimation of chord, key, and structure, where the structure arises from the variation of chord perplexity at the end of each segment.

References

- 1 F. Kaiser and G. Peeters. Multiple hypotheses at multiple scales for audio novelty computation within music. In Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, British Columbia, Canada, pages 231–235, 2013.
- 2 F. Kaiser and G. Peeters. A simple fusion method of state and sequence segmentation for music structure discovery. In Proc. of the International Society for Music Information Retrieval Conference (ISMIR), Curitiba, PR, Brazil, pages 257–262, 2013.
- 3 J. Paulus, M. Müller, and A. Klapuri. Audio-based Music Structure Analysis. In Proc. of the International Conference on Music Information Retrieval (ISMIR), Utrecht, The Netherlands, pages 625–636, 2010.
- 4 J. Pauwels, F. Kaiser, and G. Peeters. Combining harmony-based and novelty-based approaches for structural segmentation. In Proc. of the International Society for Music Information Retrieval (ISMIR), Curitiba, PR, Brazil, pages 601–606, 2013.
- 5 G. Peeters. Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation: Sequence and State Approach. In Proc. Computer Music Modeling and Retrieval (CMMR), Lecture Notes in Computer Science 2771, Springer, pages 142–165, 2004.

4 Further Topics and Open Problems

4.1 Musical Structure Analysis for Jazz Recordings

Stefan Balke, Meinard Müller (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Stefan Balke, Meinard Müller

Analyzing jazz recordings by famous artists is the basis for many tasks in the field of jazz education and musicology. Although jazz music mainly consists of improvised parts, it follows common structures and conventions which allow musicians to play and interact with each other. For example, at jam sessions and in traditional jazz recordings, musicians introduce a song by playing its main melody based on a characteristic harmonic progression. This part is also called the *head-in*. Afterwards, this progression is repeated while the melody is replaced by improvised solos by the various musicians. After all solos have been played, the song is concluded with another rendition of the main melody, a part also referred to as *head-out*. Based on this musical knowledge, we investigated automated methods for detecting (approximate) repetitions of the harmonic progression, certain melodic elements, and transitions between soloists as cues to derive a coarse structure of the jazz recording. The discussions at the Dagstuhl seminar showed that the integration of specific domain knowledge is essential for dealing with the possible musical and acoustic variations one encounters in jazz music.

References

- 1 Klaus Frieler, Wolf-Georg Zaddach, Jakob Abeßer, and Martin Pfeiderer. Introducing the Jazzomat Project and the melody Library. In *Third International Workshop on Folk Music Analysis*, 2013.
- 2 The Jazzomat Research Project. <http://jazzomat.hfm-weimar.de>.

4.2 On the Role of Long-Term Structure for the Detection of Short-Term Music Events

Juan Pablo Bello (New York University, US)

License © Creative Commons BY 3.0 Unported license
© Juan Pablo Bello

In MIR it is often assumed that there is a universal “ground truth” to music events such as beats, downbeats, chords, and melodic lines. This conveniently ignores the fact that music analysis is an interpretative task, with multiple outcomes possible. But, if there is no single answer, what should we expect from computational approaches?

One possible objective is to produce at least one valid answer, one that could have plausibly been produced by a human. I would argue that plausibility is partly a function of the long-term structural coherence of the system’s output, an aspect that is largely ignored during the design, training and evaluation of current approaches. As a result, music event detection is typically performed as a series of short-term, (semi-)independent tasks, with outputs that are often incoherent and thus implausible.

Take for example chord estimation, where methods are trained on maximum likelihood objectives derived from windows of information rarely spanning more than a single chord; dynamic models, whenever used, are almost certain to have short memories; and evaluation

is based on an aggregation of the accuracy of short-term detections. Earlier work tried to leverage long-term repetitions to enhance the robustness of feature representations with promising results [1, 2], but those strategies have not been widely adopted, having next to no impact on the feature extraction, model training and evaluation methodologies currently in use.

During the Dagstuhl seminar we have discussed the multiple ways in which the long-term, hierarchical structure of musical pieces can be used to improve the validity, and thus usability, of computational music analyses. However, some of these issues were only discussed briefly and tentatively, and much remains open for future discussion and development within the community.

References

- 1 Matthias Mauch, Katy Noland, and Simon Dixon. *Using Musical Structure to Enhance Automatic Chord Transcription*. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Kobe, Japan, pages 231–236, 2009.
- 2 Taemin Cho and Juan P. Bello. *A Feature Smoothing Method for Chord Recognition using Recurrence Plots*. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Miami, USA, pages 651–656, 2011.

4.3 Mid-level Representations for Rhythmic Patterns

Christian Dittmar, Meinard Müller (Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 3.0 Unported license
© Christian Dittmar, Meinard Müller

For music retrieval and similarity search, one important step is to convert the music data into suitable mid-level features. Ideally, these representations should capture relevant characteristics of the music while being invariant to aspects irrelevant for the given task. For example, rhythmic patterns that are played in different tempi may be perceived as similar by human listeners, while being numerically quite different. In this context, one requires mid-level representations that capture rhythmic characteristics while being invariant to tempo changes. During the Dagstuhl seminar we revisited different mid-level features that have been proposed in earlier works to capture rhythmic information. An established technique for analyzing rhythmic patterns is based on computing a local version of the autocorrelation function (ACF) of some onset-related function [1]. Together with Andre Holzapfel, we discussed open issues related to applying the scale transform [3] to rhythmic patterns for improving tempo invariance. In a follow-up discussion with Brian McFee, we highlighted the relation between the scale transform and the Log-lag ACF [2]. Together with Frank Kurth, we investigated the suitability of shift-ACF [4] for characterizing rhythmic structures with multiple repetitions.

References

- 1 Peter Grosche, Meinard Müller, and Frank Kurth. Cyclic tempogram – a mid-level tempo representation for music signals. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5522–5525, Dallas, Texas, USA, 2010.
- 2 Matthias Gruhne and Christian Dittmar. Improving rhythmic pattern features based on logarithmic preprocessing. In *Proceedings of the Audio Engineering Society (AES) Convention*, Munich, Germany, 2009.

- 3 Andre Holzapfel and Yannis Stylianou. Scale transform in rhythmic similarity of music. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1): 176–185, 2011.
- 4 Frank Kurth. The Shift-ACF: Detecting multiply repeated signal components. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, New Paltz, NY, USA, 2013.

4.4 Representation of Musical Structure for a Computationally Feasible Integration with Audio-Based Methods

Sebastian Ewert (Queen Mary University of London, GB)

License © Creative Commons BY 3.0 Unported license
© Sebastian Ewert

In terms of terminology, “musical structure” has been used in several, different contexts. In one interpretation, musical structure is essentially equivalent to musical form, which can be considered as a genre or rather style specific definition of the expectation of how a piece is composed on a rather global level. Another interpretation of structure is closer to the corresponding mathematical notion, where structure yields properties and regularity.

Both interpretations lead to various interesting questions. In the context of the first interpretation, a popular task is to determine the temporal boundaries of elements used to describe the form, such as the chorus in pop music or the recapitulation in a sonata form. One group of methods focuses on the detection of boundaries by detecting novelty or sudden changes in terms of a feature representation. This is essentially equivalent to a local, focused expectation violation of some sort. In this context, we discussed what this means for various music and composition styles, and how this could be expressed computationally using audio recordings as input.


This is directly connected to a question we raised in the context of the second interpretation of structure. Here, structure can refer to various regularities or expectations about the harmony, the rhythm, the melody or any other musical concept. In this context, music signal processing as a field has been criticized for not making enough use of these properties to obtain better results in specific tasks. While this criticism is valid it often leads to simplifying conclusions about the underlying reasons for why structure is neglected. In particular, a major obstacle is that the detection of musical low-level events is still an unsolved problem (e.g. note transcription). Therefore, good signal processing methods typically avoid making hard decisions (e.g. “This is a C-major chord”) but preserve uncertainty in a model as long as possible. This, however, leads to an exponential explosion of the underlying state space for longer time ranges and, therefore, we simply often cannot represent or integrate complex expectation models that require long time ranges – at least not using classical, symbolic Bayesian modelling techniques.

Recently, neural networks, in contrast to attempts modelling expectations explicitly, express expectations implicitly and thus can be used to build complex language models (i.e. expectation models) for polyphonic music, even down to a note level. This led to measurable but small improvements in tasks such as music transcription, and thus can be considered as a first step. A more recently developed mathematical tool are uncertainty functions, which avoid an explosion of the state space similar to neural network language models but at the same time enable the integration of explicit knowledge (to some degree). In this context, we discussed approaches and best practices to representing musical structure for specific (!) cases, where the Bayesian network philosophy fails – in particular with respect to usable,

practical ways for integrating such representations into audio signal processing methods while preserving computational feasibility.

4.5 Robust Features for Representing Structured Signal Components

Frank Kurth (*Fraunhofer FKIE – Wachtberg, DE*)

License  Creative Commons BY 3.0 Unported license
© Frank Kurth

In the last years we have developed several features for robustly representing repeating signal components [1]. A main focus in this was the robust detection of such signal components in real-world audio recordings. Applications included bioacoustical monitoring [2] (e.g., detection of repeating bird calls or sounds of marine mammals) and speech detection [3]. In the latter, the harmonic structure of voiced signal parts constitute the repeating components. One of my interests during the Dagstuhl seminar was to discuss possible applications of such features in music structure analysis.

Generally speaking, repetitions can be seen as building blocks for more complex structure elements of audio signals, which is particularly obvious for music. Thus, another interesting thing discussed at the seminar was that of possible generalizations of the repetition-based features proposed in [2] to represent such complex structures. As a third possible application, the usability of such features to the extraction and separation of mixtures of repeating components (e.g, for multipitch extraction [3] or the detection of overlapping rhythmic components) was discussed.

References

- 1 Frank Kurth. *The shift-ACF: Detecting multiply repeated signal components*. In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2013.
- 2 Paul M. Baggenstoss and Frank Kurth. *Comparing Shift-ACF with Cepstrum for Detection of Burst Pulses in Impulsive Noise*, Journal of the Acoustical Society of America 136(4):1574–1582, 2014.
- 3 Alessia Cornaggia-Urrigshardt and Frank Kurth. *Using enhanced F0-trajectories for Multiple Speaker Detection in Audio Monitoring Scenarios*. In Proc. of the European Signal Processing Conference (EUSIPCO), Nice, France, pages 1093–1097, 2015.

4.6 Reversing the Music Structure Analysis Problem

Meinard Müller (*Universität Erlangen-Nürnberg, DE*)

License  Creative Commons BY 3.0 Unported license
© Meinard Müller

The general goal of music structure analysis is to divide a given music representation into temporal segments that correspond to musical parts and to group these segments into musically meaningful categories [1, 2]. In general, there are many different criteria for segmenting and structuring music. For example, a musical structure may be related to recurring patterns such as repeating sections. Or a certain musical sections may be characterized by some homogeneity property such as a consistent timbre, the presence of a

specific instrument, or the usage of certain harmonies. Furthermore, segment boundaries may go along with sudden changes in musical properties such as tempo, dynamics, or the musical key [1]. When recognizing and deriving structural information, humans seem to combine different segmentation cues in an adaptive and subjective fashion [3]. The listener-dependent and context-sensitive relevance of different segmentation principles make structure analysis an extremely challenging task when approached with computer-based systems. During the Dagstuhl seminar, we discussed a task that may be regarded as a kind of reversal of the structure analysis problem: Given a structure annotation made by a human listener, find out possible segmentation cues that support the annotation. A similar task was suggested by Smith and Chew [4], where a given structure annotation was used to estimate the relevance of features at certain points in the music recording. During the Dagstuhl seminar, we extended this discussion by not only considering the relevance of certain feature types (e.g. representing instrumentation, harmony, rhythm, or tempo), but also the relevance of different segmentation principles based on repetition, homogeneity, and novelty. What are the musical cues used for deriving a specific segmentation boundary? Is there an unexpected musical event or a sudden change in tempo or harmony? Did the listener recognize a repeating section or a familiar phrase? Finding answers to such questions may help better understand what one may expect from automated methods and how to make computer-based approaches adaptive to account for a wide range of different segmentation cues.

References

- 1 Meinard Müller: *Fundamentals of Music Processing*. Springer Verlag, 2015.
- 2 Jouni Paulus, Meinard Müller, Anssi Klapuri: Audio-based Music Structure Analysis. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2010, pp. 625–636.
- 3 Jordan B. L. Smith, J. Ashley Burgoyne, Ichiro Fujinaga, David De Roure, J. Stephen Downie: Design and creation of a large-scale database of structural annotations. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2011, pp. 555–560.
- 4 Jordan Smith, Elaine Chew: Using quadratic programming to estimate feature relevance in structural analyses of music. In *Proceedings of the ACM International Conference on Multimedia*, 2013, pp. 113–122.

4.7 Approaching the Ambiguity Problem of Computational Structure Segmentation

Oriol Nieto (Pandora, US)

License © Creative Commons BY 3.0 Unported license
© Oriol Nieto

The identification of music segment boundaries has shown to be ambiguous; two subjects might disagree when annotating the same piece [1, 7]. This exposes a significant problem when developing computational approaches, which tend to be evaluated against references composed of a single annotation per track. These inadequately called “ground-truth” annotations will likely yield spurious results as long as they do not capture the inherent ambiguity of the given task.

In this seminar we discussed various ideas to approach this ambiguity problem:

- To make use of as many human annotations as possible when evaluating an algorithm's estimation for a specific music track. The SALAMI [7] and SPAM [5] datasets already contain multiple annotations for each piece.
- To weight each boundary based on a confidence value. It has been shown that humans generally agree when stating the per-boundary confidence of their annotations [1].
- To produce more than a single estimation. Let the user decide which estimation fits best her needs.
- To design and re-think current evaluation metrics using cognitive studies. Computational methods should produce estimations that are better aligned to perceptual cues. An example of this has already been published in [6].
- To annotate and estimate hierarchical boundaries. An algorithm can then be tuned to specific layers in the reference annotations. The depths of these hierarchies might differ based on the annotator, but some layers might contain a smaller amount of variations, thus reducing the ambiguity when focusing on them. Recent work towards computational hierarchical approaches can be found in [2, 3, 4].

References

- 1 Bruderer, M. J. (2008). Perception and Modeling of Segment Boundaries in Popular Music. PhD Thesis, Technische Universiteit Eindhoven. Retrieved from <http://alexandria.tue.nl/extra2/200810559.pdf>
- 2 McFee, B., and Ellis, D. P. W. (2014). Analyzing Song Structure with Spectral Clustering. In Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, pages 405–410.
- 3 McFee, B., and Ellis, D. P. W. (2014). Learning to Segment Songs With Ordinal Linear Discriminant Analysis. In Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Florence, Italy, pages 5197–5201.
- 4 McFee, B., Nieto, O., and Bello, J. P. (2015). Hierarchical Evaluation of Music Segment Boundary Detection. In Proc. of the 16th International Society of Music Information Retrieval (ISMIR), Taipei, Taiwan, pages 406–412.
- 5 Nieto, O. (2015). Discovering Structure in Music: Automatic Approaches and Perceptual Evaluations. PhD Thesis, New York University.
- 6 Nieto, O., Farbood, M. M., Jehan, T., and Bello, J. P. (2014). Perceptual Analysis of the F-measure for Evaluating Section Boundaries in Music. In Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, pages 265–270.
- 7 Smith, J. B., Burgoyne, J. A., Fujinaga, I., De Roure, D., and Downie, J. S. (2011). Design and Creation of a Large-Scale Database of Structural Annotations. In Proc. of the 12th International Society of Music Information Retrieval (ISMIR), Miami, FL, USA, pages 555–560.

4.8 Multi-Level Temporal Structure in Music

Hélène Papadopoulos

License © Creative Commons BY 3.0 Unported license
© Hélène Papadopoulos

Joint work of George Tzanetakis, Hélène Papadopoulos

Human beings process the global musical context in a holistic fashion. Music signals exhibit a complex relational structure at multiple representation levels. They convey multi-faceted

and strongly interrelated information (e.g. harmony, melody, metric, semantic structure), which are structured in a hierarchical way. For instance the highest-level expression of the structure (segmentation into verse/chorus, ‘ABA’ form etc) is dependent on musically lower-level organization such as beats and bars. Another example is that there is often a strong similarity between the chord progression of two semantically same segments.

Current computational models in MIR are limited in their capacities of capturing this complex relational structure. They usually have a relatively simple probabilistic structure and are constrained by limiting hypotheses that do not reflect the underlying complexity of music. In particular, during the Dagstuhl seminar, the problem that music analysis is typically performed only at a short time scale has been discussed. A stimulus talk on the use of semantic structure to constrain a beat tracking program has highlighted the benefit of combining longer-term analysis with shorter-term event detection (see also the abstract of Section 4.2 titled “On the Role of Long-Term Structure for the Detection of Short-Term Music Events”). How the hierarchical temporal structure of music can be described is a question that has been briefly evoked, but it remains an open discussion.

The work carried out in the emerging research area of *Statistical Relation Learning* offers very interesting ideas for modeling multi-relational and heterogeneous data with complex dependencies. In particular the framework of Markov logic networks (MLNs), which combines probability and logic, opens compelling perspectives for music processing. They seem suitable to design a multi-level description of music structure at various time scales (beat, measures, phrase, etc.) in which information specific to the various strata interact. In addition to encompassing most traditional probabilistic models (e.g. HMM), this framework allows much more flexibility for representing complex relational structure. For instance, earlier work have used structural repetitions to enhance chord estimation [1]. In a piece of music, repeated segments are often transformed up to a certain extent and present variations from one occurrence to another. Although usually strongly related, chord progressions in such repeated segments may not be exactly the same. Such variations can be accommodated by MLNs [2].

Also, among other appealing features, MLNs allow building probabilistic models that incorporate expert knowledge (constraints) in a simple and intuitive way, using logical rules. Using a language that is intuitive may be a way to make easier collaborations between musicologists and computer science people (see also the abstract of the working group “Computation and Musicology” in Section 5.8).

References

- 1 M. Mauch, K. Noland, and S. Dixon. *Using Musical Structure to Enhance Automatic Chord Transcription*. In Proc. of the International Society for Music Information Retrieval Conference (ISMIR), pages 231–236, 2009.
- 2 H. Papadopoulos and G. Tzanetakis. *Modeling Chord and Key Structure with Markov Logic*. In Proc. of the International Society for Music Information Retrieval Conference (ISMIR), pages 127–132, 2012.

4.9 The Ceres System for Optical Music Recognition

Christopher Raphael

License  Creative Commons BY 3.0 Unported license
© Christopher Raphael

We presented our current state of the art in optical music recognition (OMR) with a demonstration at the Dagstuhl conference.

The core recognition performed by our Ceres system understands and represents the grammatical relationships between music notation primitives (note heads, beams, accidentals etc.) necessary for useful results. Within this context, the various recognition problems are cast as dynamic programming searches that seek the grammatically-consistent representation of an object best explaining the pixel data.

In addition to the core recognition technology, Ceres is a human interactive system in which the user guides the computer in understanding a music document. Within Ceres the user can choose candidates for symbol recognition (chords, beamed groups, slurs, etc.). After recognition the user can correct errors in two ways. In the first, the user labels individual pixels with a symbol or primitive type (beam, open note head, accidental, etc), while the system then re-recognizes subject to the user-imposed constraint. In the second, the user can change basic parameters of the recognition models, while the system re-recognizes according to the new model. For instance, we may allow or disallow augmentation dots, beams that span grand staves, two-way stems, or other possible model variations. The combination gives a flexible tool for resolving recognition problems that doesn't require any understanding of the inner workings of the system.

Our goal is to create a tool that serves as the foundation for a global effort to create large, open, symbolic music libraries. Such data are needed for digital music stands, computational musicology and many other uses. For this goal to be achieved, we must create a system where high-quality symbolic data can be created efficiently. During the Dagstuhl workshop we identified a useful partnership with Meinard Müller's group, who hope to use OMR, and the resulting symbolic music representations, to relate music scores to audio and allow search and retrieval. We look forward to pursuing this collaboration.

4.10 Musical Structure Between Music Theory, Cognition and Computational Modeling

Martin Rohrmeier (TU Dresden, DE)

License  Creative Commons BY 3.0 Unported license
© Martin Rohrmeier

The experience of music relies on a rich body of structural knowledge and the interaction of complex cognitive mechanisms of learning and processing. Even seemingly simple everyday listening experiences, such as the build up of musical tension, instantaneous recognition of a “sour” note, recognition of a variation of a familiar tune or the surprise caused by an unexpected turn of phrase, rely on a body of (largely implicit) structural knowledge. The understanding of such varieties of musical experiences lies in the interdisciplinary intersection between music theory, music cognition and computational modeling [5]. Various music-theoretical approaches proposed formalizations of high-level syntactic relations in music (e.g. [3, 4, 1]). Some of the theoretical proposals have been evaluated in experimental research (see,

e.g. [2, 8]), yet the scientific potential in exploring music perception in the overlap of theory, psychology and computation remains large. During the Dagstuhl seminar, we discussed in which ways progress in complex high-level computational models of human music listening as well as in future MIR applications can be achieved by taking into account interdisciplinary insights developed in the intersection of music theory and music cognition.

References

- 1 Granroth-Wilding, M., Steedman, M. (2014). A robust parser-interpreter for jazz chord sequences. *Journal of New Music Research*, 43(4):355–374.
- 2 Koelsch, S., Rohrmeier, M., Torrecuso, R., Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*, 110(38):15443–15448.
- 3 Lerdahl, F., Jackendoff, R. (1984). *A generative theory of tonal music*. Cambridge, MA, MIT Press.
- 4 Lerdahl, F. (2001). *Tonal pitch space*. Oxford University Press.
- 5 Pearce, M., Rohrmeier, M. (2012). Music cognition and the cognitive sciences. *Topics in cognitive science*, 4(4):468–484.
- 6 Pearce, M. T., Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in cognitive science*, 4(4): 625–652.
- 7 Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1):35–53.
- 8 Rohrmeier, M., Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in cognitive science*, 4(4), pp. 525–553.
- 9 Steedman, M. J. (1996). The Blues and the Abstract Truth: Music and Mental Models, in: A. Garnham, J. Oakhill (eds.), *Mental Models in Cognitive Science* (Erlbaum, Mahwah, NJ, 1996).

4.11 Accessing Temporal Information in Classical Music Audio Recordings

Christof Weiß, Meinard Müller (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Christof Weiß, Meinard Müller

Joint work of Vlora Arifi-Müller, Thomas Prätzlich, Rainer Kleinertz, Christof Weiß, Meinard Müller

Music collections often comprise documents of various types and formats including text, symbolic data, audio, image, and video. For example, in one of our projects, we are dealing with operas by Richard Wagner, where one has different versions of musical scores, libretti, and audio recordings. When exploring and analyzing the various kinds of information sources, the identification and establishment of semantic relationships across the different music representations becomes an important issue. For example, when listening to a performance given as CD recording, time positions are typically indicated in terms of physical units such as seconds. On the other hand, when reading a musical score, positions are typically specified in terms of musical units such as measures. Knowing the measure positions in a given music recording not only simplifies access and navigation, but also allows for transferring annotations from the sheet music to the audio domain (and vice versa). In our Wagner project, we have started to annotate measure positions within various performances for the opera cycle “Der Ring des Nibelungen” either supplied by human annotators or generated by automated music synchronization techniques. Surprisingly, even the manually generated

annotations (not to speak of the annotations obtained by automated methods) often deviate significantly.

At the Dagstuhl seminar, we presented this scenario and reported typical problems. In particular, we discussed why and to which extent the task of identifying measure positions in performed music is ambiguous and what the (musical) reasons for highly deviating measure annotations are. Furthermore, we raised the question how one should evaluate automated procedures in the case where human annotators disagree. Over the course of the seminar, we found similar problems in the work of other participants. In particular, the issue of ambiguity when determining structural relevant time positions was an ongoing topic throughout the seminar.

References

- 1 Vlora Arifi, Michael Clausen, Frank Kurth, Meinard Müller. *Synchronization of Music Data in Score-, MIDI- and PCM-Format*. In Walter B. Hewlett and Eleanor Selfridge-Fields (ed.), MIT Press, 13:9–33, 2004.
- 2 Meinard Müller. *Fundamentals of Music Processing*. Springer Verlag, 2015.

5 Working Groups

5.1 Human-in-the-Loop for Music Structure Analysis

Participants of Dagstuhl Seminar 16092

License © Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 16092

Most of prior research on music structure analysis can be found in the realm of popular music. Typically, one starts with an audio recording which is then split up into segments by the system based on principles such as repetition, novelty, and homogeneity [5, 6]. Each part may then be labeled with a meaningful description such as intro/verse/chorus or a symbol ‘A’ as shown in the AABA form. Although the perception of structures is highly subjective, a tremendous amount of work has been proposed on developing segmentation algorithms for music structure analysis, and more importantly, on the evaluation metrics for such algorithms [9, 5, 4, 7]. However, tasks such as analyzing formal structure in tonal music require musically trained experts, who often disagree with each other in their annotations because of the ambiguity or complexity in music compositions [2]. In such cases, interactive systems that incorporate humans in the loop [8] or that are based on active learning approaches seem promising.

In this working group discussion, we shared best practices for collecting data from participants [3, 1]. We also discussed the importance of asking suitable questions to obtain meaningful answers relevant for the desired structure analysis task. Obtaining a large amount of high-quality annotated data from trained and motivated participants remains a challenge for all disciplines. However, such data, enhanced by multiple structure annotations, is especially important for MIR due to structural ambiguities (i.e., natural variations) in music. Also, multiple annotations allow for eliminating incorrect or clarifying flawed annotations. In summary, framing a question from the perspective of musicology and music cognition with quantifiable measures is one key for music structure analysis. Also, targeting an application and understanding its users may render the difficult evaluation task more feasible.

References

- 1 Alessio Bazzica, Cynthia C. S. Liem, and Alan Hanjalic. *On detecting the playing/non-playing activity of musicians in symphonic music videos*. Computer Vision and Image Understanding, 144(C), pages 188–204, 2016.
- 2 Ching-Hua Chuan and Elaine Chew. *Creating Ground Truth for Audio Key Finding: When the Title Key May Not Be the Key*. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Porto, Portugal, pages 247–252, 2012.
- 3 Masataka Goto, Jun Ogata, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, and Tomoyasu Nakano. *PodCastle and Songle: Crowdsourcing-Based Web Services for Retrieval and Browsing of Speech and Music Content*. In CrowdSearch, pages 36–41, 2012.
- 4 Brian McFee, Oriol Nieto and Juan P. Bello. *Hierarchical Evaluation of Segment Boundary Detection*. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Málaga, Spain, pages 406–412, 2015.
- 5 Meinard Müller. Music Structure Analysis. In Fundamentals of Music Processing, chapter 4, pages 167–236, Springer Verlag, 2015.
- 6 Jouni Paulus, Meinard Müller, and Anssi Klapuri. Audio-based Music Structure Analysis. In Proc. of the International Conference on Music Information Retrieval (ISMIR), Utrecht, The Netherlands, pages 625–636, 2010.
- 7 Marcelo Rodriguez-López and Anja Volk. *On the Evaluation of Automatic Segment Boundary Detection*. Proceedings of the International Symposium on Computer Music Multidisciplinary Research, Plymouth, UK, June, 2015.
- 8 Markus Schedl, Sebastian Stober, Emilia Gómez, Nicola Orio, and Cynthia C. S. Liem. *User-Aware Music Retrieval*. In Meinard Müller, Masataka Goto, and Markus Schedl, editors, *Multimodal Music Processing, Dagstuhl Follow-Ups*, volume 3, pages 135–156, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2012. <http://www.dagstuhl.de/dagpub/978-3-939897-37-8>
- 9 Jordan Smith and Meinard Müller. *Music Structure Analysis*. Tutorial at the International Society for Music Information Retrieval Conference (ISMIR), 2014, <http://www.terasoft.com.tw/conf/ismir2014/tutorialschedule.html>.

5.2 Computational Methods

Participants of Dagstuhl Seminar 16092

License © Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 16092

This working group discussion began by examining the performance gap between human annotators of musical structure and computational models. One key question was how to close this gap primarily by leveraging what is in the signal while recognizing that some important information used by human annotators may not be there. To this end, two different approaches were discussed in detail – one which already exists and the other which was more abstract.

- The deep neural network approach of Grill and Schlüter at OFAI [1]. The group questioned whether this highly effective system was being optimized according to the best evaluation metric, and posed a secondary hypothetical question: Could such a deep architecture learn wrongly-labeled structural boundaries?
- Theoretical large scale graphical model that extracts “everything.” This theoretical approach was discussed based on evidence that some tasks which are trained across two

simultaneous (but related) tasks can outperform those trained only on one. However, the importance of defining precisely which tasks should be estimated in parallel (e.g., key, structure, chords, beats, downbeats, onsets, and so on) was considered essential.

The working group recognized the extremely high computationally demands required to run such a model, and the unresolved issue of how to train it – in particular whether all the different layers really would interact with each other equally. The group considered how key and beat information are not strongly related, but that keys are linked with chords, and beats can be associated with chord changes. A further issue concerned how to account for the fact that not all dimensions of interest might be present in the signal being analyzed. Hence, such a system would be ineffective without a measure of confidence or salience in relation to the presence or absence of such dimensions.

In summary, the group determined that it was not fruitful to attempt to build a huge model of everything musical (in this sense, it could be considered an anti-grand challenge for computational music structure analysis). Instead, the most promising next step in relation to computational models should be to identify cases where there is a meaningful and well-defined interaction between different sources of information, such as the approaches by Holzapfel et al. [2] (jointly considering beat, tempo, and rhythm), Dannenberg [3] (combining beat tracking and structural alignment), or Papadopoulos et al. [4] (jointly estimating chords and downbeats). Finally, the group revisited the issue of evaluation and how the calculated accuracy scores could be revisited in order to optimize new computational methods to the most perceptually valid measures of performance.

References

- 1 Thomas Grill and Jan Schlüter. *Structural Segmentation with Convolutional Neural Networks MIREX Submission*. MIREX abstract, 2015.
- 2 Andre Holzapfel, Florian Krebs, and Ajay Srinivasamurthy. *Tracking the “Odd”: Meter Inference in a Culturally Diverse Music Corpus*. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, pages 425–430, 2014.
- 3 R. Dannenberg. *Toward Automated Holistic Beat Tracking, Music Analysis and Understanding*. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), London, UK, pages 366–373, 2005.
- 4 Hélène Papadopoulos and Geoffroy Peeters. *Joint Estimation of Chords and Downbeats from an Audio Signal*. IEEE Trans. Audio, Speech and Language Processing 19(1):138–152, 2011.

5.3 Applications of Music Structure Analysis

Participants of Dagstuhl Seminar 16092

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 16092

This working group focused on discussing application scenarios for music structure analysis. We began with an attempt to categorize existing applications into more scientifically and practically oriented ones. The group identified a number of applications of practical relevance in areas such as music appreciation and interaction, automatic music generation, music education and teaching, music repurposing, video production, as well as toys and games. Further examples of commercial applications using music structure analysis are, for example, *Adobe Premiere Clip* and *Jukedeck*. As for the scientific applications, the group concluded

that music structure analysis is essential to reduce the complexity of music, test theories of perception, music generation, and musicology in general. The idea was brought up that, instead of struggling to reveal structural elements derived from existing music theory, music structure analysis could be applied to infer propositions for a music theory in contexts where such a theory has not yet been developed. Especially for oral music traditions this might be a promising direction for future research. Diverse findings from practical experiences were brought up throughout the discussion. For example, in the automatic creation of mash-ups, downbeats are the most important kind of structural information. The problem of recruiting and keeping users interested in the evaluation of music structure analysis results expanded upon for the case of the PHENICX project, see [1].

References

- 1 Cynthia Liem, Emilia Gómez, and Markus Schedl *PHENICX: Innovating the classical music experience*. Proceedings of the International Conference on Multimedia Expo Workshops (ICMEW), 2015

5.4 Rhythm in Music Structure Analysis

Participants of Dagstuhl Seminar 16092

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 16092

This working group discussed issues that relate rhythm analysis to music structure analysis. On the one hand, there is a multitude of MIR tasks that involve rhythm such as beat, downbeat and tempo tracking, microtiming analysis, onset detection, and time signature estimation. On the other hand, conventional approaches to music structure analysis are often based solely on timbre and harmony, but neglect aspects related to rhythm. Recently, significant performance improvements have been achieved in tempo and meter tracking even for difficult music with soft onsets and varying tempo [2]. However, from a musicological point of view, rhythm has too often been studied in isolation from related concepts. The group discussed and identified a number of grand challenges and open issues related to rhythm analysis.

- The importance of microtiming for groove styles and short-term temporal structure [3, 1].
- Interaction between rhythm and other musical properties relevant for musical structures.
- Tracking of structures above the measure level (supra-metrical levels).
- Robust downbeat tracking.
- Exploiting larger scale repetitions for analyzing rhythmic patterns.
- Discrimination of expressive timing from bad timing.
- Visualization of conducting styles and gestures that shape the tempo.
- Identification of metrical levels that go into body movement in dance music.

References

- 1 Martin Pfeleiderer. *Rhythmus: Psychologische, Theoretische und Stilanalytische Aspekte Populärer Musik*. Transcript, 2006.
- 2 Florian Krebs and Sebastian Böck and Gerhard Widmer, *An Efficient State-Space Model for Joint Tempo and Meter Tracking*. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Málaga, Spain, October, pages 72–78, 2015.
- 3 Richard Ashley *Grammars for Funk Drumming: Symbolic and Motor-Spatial Aspects*. Abstract in Proc. Cognitive Science Society, Portland, Oregon, 2010.

5.5 Similarity in Music Structure Analysis

Participants of Dagstuhl Seminar 16092

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 16092

In this working group, we discussed the importance of music similarity. Early on, we agreed that the measurement of music similarity by computational means is the key ingredient for music structure analysis. Unfortunately, there is no general definition of similarity readily available, as the concept depends on the signal's properties, the application context, and the user background. Usually, corpora need to be defined that reflect similarity ratings. Under the assumption that there is a common understanding of similarity among composers, performers and listeners, one then tries to measure similarity in a linear continuum. Besides the fact that measuring similarity of symbolic music representations is still an open issue, even more intricate problems arise from other music traditions that are strongly based on learning music by imitation. This was made clear by some examples from the Dunya Makam corpus [1]. While for this Turkish music tradition, similarity based on discrete pitches is reasonable, in Chinese and Indian music traditions the performance characteristics (such as inflections and vibrato) are much more important. Moreover, even though music segments may be similar from a semantic viewpoint, they often exhibit significant spatial or temporal differences. Therefore, we pointed out the importance of the design of musically meaningful mid-level features so that the subsequent similarity rating can be handled very efficiently using simple metrics (basically inner products and correlation measures).

References

- 1 Holzapfel, A., Şimşekli U., Şentürk S., and Cemgil A. T. *Section-level Modeling of Musical Audio for Linking Performances to Scores in Turkish Makam Music*. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 141–145, Brisbane, Australia, 2015

5.6 Structure in Music Composition

Participants of Dagstuhl Seminar 16092

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 16092

Nobody really listens to generated music. We believe this is due to the fact that computer-generated music typically lacks long-term structure. Systems employing Markov models usually only look at transition probabilities on a very local level. Context-free grammars offer a step-up, but have not been very successful in constraining global structure. These days, many believe in deep learning and speak of the potential of convolutional neural nets for generating music with structure; we have not yet seen convincing results.

Since music generation can be viewed as the flip side of music analysis (as stated by one of the participants), we can use the information retrieved from the structural analysis to facilitate the music generation process. We therefore believe that generation should be a two-step process: generate the structure first (i.e., a narrative that the music follows), then generate the music based on that structure. A possible approach could be to use a tension profile model to capture structure. Such models [1, 2] can capture a lot more than people realize and listeners recognize it. This approach could potentially be combined with structure

captured by information content. These two types of structure are closely related, but do not always have a linear relationship. In the approach followed by Herremans in the MorpheuS project (<http://dorienherremans.com/MorpheuS>), generating music that fits a tonal tension profile is combined with enforcing a higher level, top-down structure defined by a pattern detection algorithm. In conclusion, generating music into a structure allows us to overcome the problem that we have long been facing in the field of automated composition. We propose that tension and information content might be suitable structural profiles, especially when combined with a larger top-down structure.

References

- 1 Farbood, M. (2012). *A parametric, temporal model of musical tension*. *Music Perception* 29(4):387–428.
- 2 Herremans, D., and Chew, E. (2016). *Tension ribbons: Quantifying and visualising tonal tension*. Second International Conference on Technologies for Music Notation and Representation (TENOR). Cambridge, UK.

5.7 Structure Analysis and Music Cognition

Participants of Dagstuhl Seminar 16092

License © Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 16092

For the discussion in this session two different directions were proposed. First, the question was posed as to what aspects of the cognitive sciences and music psychology can be helpful for computational approaches to structure analysis. Second, the question on how cognitive and psychological studies can be advanced with the help of computational analysis of musical structure was raised. David Temperley, in his keynote talk at the CogMIR Conference 2013, outlined some thoughts about the relation between music cognition and music information retrieval (MIR). He stated that, for well-defined problems, MIR can work alone; for many ill-defined problems, however, MIR can profit from collaborations with the cognitive sciences. Examples that were discussed in the working group session were:

- In MIR, one rarely encounters the concept of reduction as it has been proposed in theories such as the Generative Theory for Tonal Music (GTTM) by Lerdahl and Jackendoff [1].
- Learned signal representations derived from deep learning could profit from incorporating perceptually motivated low-level features.
- Mechanisms that help humans to predict music events and shape their expectations about what is going to happen next [2] should be better exploited.

In summary, most participants of our working group agreed that MIR approaches to structure analysis should incorporate findings from the cognitive sciences related to the aspects that help human listeners chunk a piece of music into memorable segments. These processes are mainly those of expectation, reduction, and tension. However, an important problem that such an interdisciplinary exchange will face is the fact that most models for these processes have been derived and documented using notated representations of music. Making these models work with audio recordings is a major challenge that impedes straightforward incorporation of the models into MIR approaches.

References

- 1 Lerdahl, F. and Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press Cambridge.
- 2 Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

5.8 Computation and Musicology

Participants of Dagstuhl Seminar 16092

License  Creative Commons BY 3.0 Unported license
 © Participants of Dagstuhl Seminar 16092

At the Dagstuhl Seminar, participants representing musicology and the computer/engineering sciences identified the hurdles that prevent more widespread collaboration between these two disciplines. In addition to their different philosophical orientations, the two fields have very different cultures. Musicological research privileges sole-authored, narrowly focused, hermeneutical, score-centered, and (sometimes) strongly opinionated writings. Computational research is multi-authored, broadly focused, descriptive, audio-centered, and leaves interpretation to the reader. Still, we also identified connecting points such as our shared interest in normal practices of musicians as well as our shared interest in analytical reduction and classification.

5.9 Music Structure Annotation

Participants of Dagstuhl Seminar 16092

License  Creative Commons BY 3.0 Unported license
 © Participants of Dagstuhl Seminar 16092

In this working group, we discussed various aspects regarding the generation, the usage, and the availability of structure annotations for music recordings. The following list summarizes the main points of our discussion.

- Several systems and methodologies have been proposed to annotate the music structure of an audio soundtrack. This is not a problem in itself but should be kept in mind when analyzing the results of an algorithm on a given test-set, since the evaluation results may significantly depend on the system and methodology used to create the annotations of the test-set.
- For a single system and methodology, the annotation may also vary from one annotator to another. We therefore need to collect the information on why an annotator did his or her annotation.
- Various kinds of (known and unknown) ambiguities as well as subjectivity are sources of annotation variations.
- There seem to be a tendency for annotators to agree on large-scale structures, but not necessarily on small-scale structures.
- Annotations should be considered as a sparse graph (lattice of hypotheses), rather than a sequence of segmentation boundaries.

- The ambiguity as well as the various annotation systems and methodologies should be handled explicitly rather than being “swept under the carpet.” The best solution is to collect several annotations representing the ambiguities.
- The following corpora of annotations have been identified and discussed.
 - Public: SALAMI (1400), IRISA (380), RWC (200), Isophonics (360)
 - Private (audio non-public): IRCAM (300), TUT-Paulus (600), Billboard (1000), CompMusic, ACM-Multimedia
- We should produce a document listing the various corpora and explaining their corresponding annotation systems and guidelines. It would be interesting to compare annotations of the same track resulting from the various systems.
- It is easier to detect the beginning of a segment than its end. We need a format that would easily allow for the representing of fuzzy endings. Maybe the JAMS (JSON-based music annotation) format could be used for storage [3].

References

- 1 G. Peeters and E. Deruty. *Is music structure annotation multi-dimensional? A proposal for robust local music annotation*. In Proc. of the International Workshop on Learning the Semantics of Audio Signals (LSAS), Graz, Austria, 2009.
- 2 F. Bimbot, E. Deruty, S. Gabriel, and E. Vincent. *Methodology and resources for the structural segmentation of music pieces into autonomous and comparable blocks*. In Proc. of the International Society for Music Information Retrieval (ISMIR), Miami, Florida, USA, pages 287–292, 2011.
- 3 E. Humphrey, J. Salamon, O. Nieto, J. Forsyth, R. Bittner, and J. Bello. *JAMS: A JSON Annotated Music Specification for Reproducible MIR Research*. In Proc. of the International Society for Music Information Retrieval (ISMIR), Taipei, Taiwan, pages 591–596, 2014.
- 4 J.B.L. 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 Proc. of the International Society for Music Information Retrieval (ISMIR), Miami, Florida, USA, pages 555–560, 2011.

Participants

- Stefan Balke
Univ. Erlangen-Nürnberg, DE
- Juan Pablo Bello
New York University, US
- Frédéric Bimbot
IRISA – Rennes, FR
- Carmine Emanuele Cella
ENS – Paris, FR
- Elaine Chew
Queen Mary University of
London, GB
- Ching-Hua Chuan
University of North Florida, US
- Roger B. Dannenberg
Carnegie Mellon University, US
- Matthew Davies
INESC TEC – Porto, PT
- Christian Dittmar
Univ. Erlangen-Nürnberg, DE
- Sebastian Ewert
Queen Mary University of
London, GB
- Mary Farbood
New York University, US
- Masataka Goto
AIST – Tsukuba, JP
- Dorien Herremans
Queen Mary University of
London, GB
- Andre Holzapfel
OFAI-Wien, AT
- Rainer Kleinertz
Universität des Saarlandes, DE
- Frank Kurth
Fraunhofer FKIE –
Wachtberg, DE
- Cynthia C. S. Liem
TU Delft, NL
- Brian McFee
New York University, US
- Meinard Müller
Univ. Erlangen-Nürnberg, DE
- Oriol Nieto
Pandora, US
- Mitchell Ohriner
Shenandoah University –
Winchester, US
- Hélène Papadopoulos
L2S – Gif sur Yvette, FR
- Geoffroy Peeters
UMR STMS – Paris, FR
- Christopher Raphael
Indiana University –
Bloomington, US
- Martin Rohrmeier
TU Dresden, DE
- Mark Sandler
Queen Mary University of
London, GB
- Xavier Serra
UPF – Barcelona, ES
- Jordan Smith
AIST – Tsukuba, JP
- Anja Volk
Utrecht University, NL
- Christof Weiß
Univ. Erlangen-Nürnberg, DE
- Geraint A. Wiggins
Queen Mary University of
London, GB

