PHENICX: Performances as Highly Enriched aNd Interactive Concert Experiences

Emilia Gómez, Maarten Grachten, Alan Hanjalic, Jordi Janer, Sergi Jordà, Carles F. Julià, Cynthia Liem, Agustin Martorell, Markus Schedl, Gerhard Widmer (Authors in alphabetical order)
PHENICX consortium
`phenicx@upf.edu`

ABSTRACT

Modern digital multimedia and internet technology have radically changed the ways people find entertainment and discover new interests online, seemingly without any physical or social barriers. Such new access paradigms are in sharp contrast with the traditional means of entertainment. An illustrative example of this is live music concert performances that are largely being attended by dedicated audiences only.

This papers introduces the PHENICX project, which aims at enriching traditional concert experiences by using state-of-the-art multimedia and internet technologies. The project focuses on classical music and its main goal is twofold: (a) to make live concerts appealing to potential new audience and (b) to maximize the quality of concert experience for everyone. Concerts will then become multimodal, multi-perspective and multilayer digital artifacts that can be easily explored, customized, personalized, (re)enjoyed and shared among the users. The paper presents the main scientific objectives on the project, provides a state of the art review on related research and presents the main challenges to be addressed.

1. INTRODUCTION

In the current digital age, access to recorded music is readily available. This makes it very easy to serendipitously get confronted with unknown music genres on (social) streaming services. However, barriers can be experienced to really go out and experience a live performance of such an unknown music genre: the walls of an unknown concert venue put up a physical barrier, and at the local etiquette of the social community that identifies most strongly with the performed music puts up a social barrier. If people who would be interested in exploring live performances of unfamiliar music will be faced with an isolated, imposed and standardised concert situation they do not naturally identify with, they thus will remain 'outsiders' to the music and its entourage.

Present-day technologies can change the way we access and enjoy musical concerts today. A wealth of musical information is available on the web, ranging from artist information to scores and lead sheets or other related information about musical pieces. Employing automated analysis techniques, it is possible to find a way through all this supporting information, tailored to our backgrounds and interests. Linking this to live concert performance data, an enriched and deepened experience of the performed music can be created in a personalised way. This can trigger our curiosity to see more of such performances, and share the experience over social media to our friends who then can pick up interest in this as well.

Following these considerations, the PHENICX project was conceived. It focuses on researching how to improve the accessibility of live music concert performances by addressing two main objectives:

Transforming live music concert performances into enriched multimodal, multi-perspective and multilayer digital artefacts

With *multimodal*, we mean different musical modalities, such as audio, video, and symbolic scores. With *multi-perspective*, we mean that a concert performance can be considered from different viewpoints: physical viewpoints in a concert hall and different user perspectives dependent on their backgrounds and intents. With *multilayer*, we mean that multiple music concert performance descriptors can be relevant at the same time, working at differing levels of specificity (e.g. requiring general or sophisticated musical knowledge) and considering different time scale resolutions. Automated and multimodal music description techniques are relevant to this objective, such that they will yield meaningful descriptors from the considered musical pieces. In our approach, performance information is characterised along two dimensions: that of the *musical piece* (*objective* descriptors, valid for any rendition of the piece), as well as *its actual performance* (descriptors on individual expressive and interpretative aspects that make one performance different from another).

Presenting digital music artefacts as engaging digital experiences that can be explored, (re)enjoyed and shared in many customisable and personalised ways

For this, we need advanced user profiling and community characterisation techniques, which will pave the way for sophisticated personalisation techniques. Next to that, techniques for dedicated and adaptive information selection and

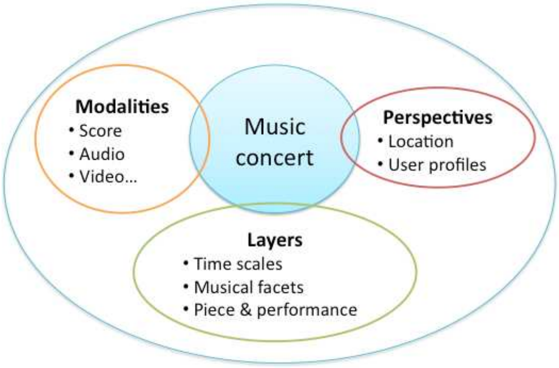

Figure 1. PHENICX view of a musical concert.

presentation has to be investigated, as well as interactive opportunities for audiences to engage even more with the performance.

By working towards these objectives as outlined above, we can transform the current audience experience of concert performances, break the physical and social barriers and decrease the perceived distance between musical performers and their concert audiences.

This transformation entirely takes place in the digital domain. Performing musicians themselves will therefore not have to change their way of performing, maintaining their original performance traditions.

2. PRACTICAL SETTING

The PHENICX project will mainly focus on Western classical music in large ensemble settings. Classical music is a very strong example of European cultural heritage, which suffers from very strong audience stereotypes, and the general image of being a complex and possibly boring genre. As such, it does not straightforwardly attract new audiences, and its live performance tradition may be endangered as outlined in our Motivation. As will be outlined in the remainder of this paper, Western classical music also poses several interesting research challenges that are not encountered for other musical genres, and thus can help in pushing scientific advances in music and multimedia information retrieval forward. The project will be structured in four different research and development areas.

Multimodal musical piece analysis: the project will research on suitable analysis techniques for automatic description and enrichment of the considered musical pieces.

Multimodal musical performance analysis: the project will also address performance aspects by extracting expression-related features from audio signals and other modalities (e.g. score, video), synchronizing performances with their score and with alternative performances of the same piece, and characterizing performer's or conductor's gestures.

Profiling and personalization, in order to adapt concert experiences to different user profiles.

Exploration and interaction, as a way to enhance the concert experience through music visualization, personalised musical information and interactive systems for conductor/performer impersonation.

In the following sections, we discuss relevant scientific state-of-the-art in these areas and corresponding challenges as foreseen for PHENICX. A discussion on how these challenges can be validated in real world end-user settings will be available in [1].

3. MULTIMODAL MUSICAL PIECE ANALYSIS

The main goal of the research related to musical piece analysis in the project is to provide the audience with meaningful information about the music material played in the concert, including musical descriptors (e.g. theme, melodic line, key, structure), semantic labels (e.g. mood), similar pieces, or links to existing online information about performers, composers or instruments. Moreover, the project will research on audio processing technologies to separate the different sections of the orchestra from mixed recordings in order to allow multi-perspective listening experiences.

3.1 Content-based feature extraction and similarity

Current techniques are capable of automatically obtaining features from music recordings related to different musical facets such as melody, harmony, rhythm and instrumentation. These descriptors are exploited by music retrieval and recommendation systems to compute similarity distances and to classify musical pieces according to e.g. artist, genre or mood [2].

However, there is a glass ceiling in current feature extractors. The accuracy of state-of-the-art methods for audio feature extraction does not go beyond 80% (results slightly vary for different tasks, e.g. onset detection, genre classification, chord detection, predominant melody extraction), even though they are not always evaluated on realistic situations (limited, e.g. to simple music material). In addition, there is a semantic gap between existing descriptors and expert musicological analyses. For instance, similarity algorithms have been traditionally based on low-level timbre descriptors, beat tracking is not accurate for expressive music with varying tempo, and melodic/harmonic descriptors are often limited to global key, which has shown to be poor to represent the tonal content of a musical piece. In the foreseen project, we will research on the best strategies for our particular repertoire, classical music in large ensemble settings. We will address the limitations of state-of-the-art methods for predominant melody estimation [3], rhythm description [4] and tonal analysis [5] to deal with our particular music material.

In the project, we should finally investigate to what extent differing application contexts may have different notions of similarity, both from a systems and user perspective, and we will consider hybrid approaches (integrating different descriptors and temporal resolutions) as proposed in [6]. For example, a scholar studying a particular piece may wish to gather many recordings of the piece and will consider these recordings to be dissimilar in comparison to each other, while to a novice unfamiliar with the piece, all these recordings will sound very similar to each other, and any differences between them are not as relevant. While this has not frequently been addressed in literature yet, it is an important topic to investigate since it will influence the ultimate success of a music information system.

3.2 Music auto-tagging

The process of automatically assigning semantically meaningful labels to representations of music is commonly known as auto-tagging. So far, auto-tagging has mostly been performed on the level of artists (e.g. [7, 8]) or songs (e.g. [9, 10]); only few works [11, 12] have addressed tagging of segments within a song.

To the PHENICX project, it is useful to automatically be able to obtain label descriptions for recorded performances. However, this means that research beyond the current state-of-the-art is needed. First of all, current methods strongly focused on pop music. In classical music, a 'song' will typically be much longer than in pop music, which means that more work is needed into obtaining segment-level descriptors. Furthermore, the multimodal and social setting of the project allows for the consideration additional data sources such as social tags, which can be gathered from collaborative tagging systems, textual features extracted from web pages or microblogs, or even simple visual features mined from images (e.g., album covers or photographs). However, once again, if these additional data sources were considered in previous work (which is uncommon, since the predominant focus has been on audio information only), this was in the pop music domain, and it should still be investigated to what extent they will be equally informative for the classical music domain.

3.3 Linking web sources of music

In PHENICX, we should extract information about performers and instruments, aim for a multimodal approach which enriches the presentation with videos, images and other supporting material, including possible alternative performances of the same piece. This means that different sources of music information need to be linked together.

Classical music is ontologically more complex than pop music: in many cases, we are not just dealing with songs performed by artists, but with a piece consisting of multiple movements, written by a composer, and interpreted by varying groups of performing artists. In terms of Semantic Web technology facilities, the Music Ontology [13] is a rare example of an ontology which has been expanded to deal with classical music cases, and as such will be of active interest for PHENICX. However, in the imperfect real world, (metadata) information on classical music may not always be cleanly and consistently labeled following this ontology. Therefore, effort will be investigated in techniques to still match this imperfect data.

As an example of a multimodal music information system involving web-scale information, [14] should be mentioned, presenting a system offering information about similarities between music artists or bands, prototypicality of an artist or a band for a genre, descriptive properties of an artist or a band, band members and instrumentation, and images of album cover artwork is performed. Once again, this system was aimed at popular music, and it should be verified to what extent the approach will translate to the classical domain.

3.4 Multi-perspective audio description: source localisation and separation

One characteristic of orchestral music concerts compared to other amplified musical live performances is how sound is propagated from the performers to the audience. Sound sources are spread over a large stage area creating an acoustic image in front of the audience, which is affected then by the acoustics of the concert hall. A recording setup might consist principally of a stereo pair microphones placed near the conductor. In a typical setup, however, this stereo track can be complemented with a number of zenithal microphones covering specific instrumental sections. These zenithal tracks are used to find the right balance in the final mastering mix.

One of the objectives of the project consists in obtaining the localisation of the active instruments on stage from a set of recorded tracks. This process shall include means of providing a source signal separation. In our scenario, we might take advantage of additional data such as the score or source positioning informations (e.g. instrument sections).

State of the art methods of source localisation include beamforming techniques, which take input signals from sensor arrays. Other specific techniques address the case of stereo signals [15]. Regarding source separation, state of the art techniques involve Non-negative Matrix Factorisation (NMF) and PLCA [16], but more recent techniques are also based on signal-models that exploit musical knowledge [17]. Score-informed techniques such as [18] are specially relevant in the context of the project.

4. MULTIMODAL MUSICAL PERFORMANCE ANALYSIS

The central purpose of research related to musical performance analysis in the PHENICX project is to give the audience or music consumer deeper insights into the subtle art of expressive performance, which is so central to classical music. This requires methods for computing expressive aspects (e.g., tempo and timing) from recorded or live performance – which in turn requires methods for aligning performances to scores, or to each other –, models for explaining, predicting, and visualising expressive aspects, and methods for recognising and characterising expressive actions by the musicians that are not readily apparent from the audio signal (for instance, gestures by the conductor). The latter will also be used to devise ways of directly interacting with performances via gestures.

4.1 Score-performance alignment, performance-to-performance matching, and real-time score following

Computing a one-to-one alignment between a performance and another representation of the same piece is important for several purposes in the project. We distinguish three cases: (1) aligning a recorded performance (audio recording) to the musical score ("score-performance alignment"),

(2) aligning two or more performances (audio recordings) to each other ("performance-to- performance matching"), and (3) aligning an ongoing performance (coming in as an audio stream) to the score in real time ("performance tracking" or "real-time score following").

In the case of *score-performance alignment*, the score is usually either rendered to audio, or acoustic features are computed directly from the score. Most alignment algorithms then use some kind of Dynamic Time Warping (DTW) to find an optimal global alignment [23–25], or model the musical processes via statistical graphical models [26, 27]. PHENICX will focus on the DTW approach, starting from and improving the methods proposed in [25], which rely on the percussiveness of the considered instruments sounds. Although recent efforts towards timbre-invariant audio features are promising [28], generalising the above methods to the wide variety of orchestral instruments will require the design of new audio features, as well as fundamental modifications to the general top-down alignment strategy. A second class of challenges concerns the possibility of structural differences between score and performance, or between performances [29, 30]. We believe these problems can more easily solved in DTW-based methods.

With respect to *real-time score following*, there are also two competing approaches, again based on either (online) DTW (OLDTW) or graphical models and probabilistic inference (e.g., [31,32]. Recent research on DTW-based performance tracking [29] looks extremely promising – not only with respect to computational efficiency and low latency, but also w.r.t. robustness against playing errors, omissions and insertions.

The biggest challenge in real-time tracking of classical music is to design more effective predictive tempo models, for the system to be able to anticipate abrupt changes in local tempo, or the return of the soloist or orchestra after a long rest. Here, the above predictive performance models will play an important role.

4.2 Explanatory and predictive computational models of expressive performance

Despite considerable research over several decades, our knowledge of the factors that shape musical expression is still far from complete. Valuable explanatory models do exist, but they tend to focus on highly specific aspects of performance, such as the form of a final ritard [19] and the effect of phrase structure on tempo [20]. With advances in both sensor technology and automatic transcription of musical audio, much more substantial empirical data is now becoming available [21], and these now allow for a paradigm-shift from the classical music-theory driven approach to a data mining approach, inspiring new computational models of expressive performance.

In this context, Grachten and Widmer [22] recently proposed a framework for modeling expressive performance. It allows to estimate the contribution of arbitrary features of the musical score (including, but not limited to expressive markings annotated in the score) in shaping expressive characteristics of the performance, such as tempo, loudness, and articulation. Musical features are represented as basis functions, which are linearly combined over one or more performances, to approximate their expressive characteristics. This framework can be used for explanatory modeling, and thereby provide the users with precise characterisations and explanations (e.g. in what ways do different ensembles perform the same piece differently?). Moreover, as a computational model, the framework also allows for predictive modeling. Accurate hypotheses about the shape of musical expression in a performance can improve score-performance alignment and real-time score following [29].

4.3 Gesture recognition

The purpose of gesture recognition in the project is to provide additional insight into how expressive performances are realized, and to facilitate interactive music-making scenarios. In the literature on the recognition of body gestures, we can distinguish two main approaches: Machine learning (usually supervised – e.g., [33]) and analytical techniques. The analytical description of gestures in order to recognise them is the most used technique right now in commercial applications and devices that require gesture recognition. Other frameworks allow describing the gestures rather than program them directly, as in [34] that allows this description in a form of regular expressions.

In PHENICX we will consider an analytical approach to recognise the principal components of specific symbolic gestures for different instruments using a composition technique and an agent-based framework [35], as well as general features of the whole body movement, and try to recognise concurrent performances of these gestures at the same time for multi-user interaction. More precisely, in the field of studying body movement of music performers we can find several approaches, like recording precise movement of a violin bow to synthesise its sound [36] or (more related to our approach) studies about "Air playing" [37]. Our research will try to link body movements and gestures to high level properties of music, such as loudness, tonality, tempo and note density.

5. PROFILING AND PERSONALISATION

PHENICX strives to offer personalised music experiences. This means that adequate user and recommendation models need to be set up.

An important direction to consider here is that of profiling and personalisation through social media mining, in which we build forth on techniques proposed in existing work including [38–41]. Of these references, only [41] explicitly deals with music recommendation, showing that users prefer social recommendations (taking into account friends) over non-social ones, and that social recommendations are particularly well-suited to discover relevant and novel music. However, the proposed user model is relatively coarse. Furthermore, in general it is important to realize that apart from general taste, a person's preference for a certain item will also be influenced by ad hoc context and search intent.

In existing music-related work, the concept of 'context' has been defined, gathered and incorporated in varying ways. In [42], a study is presented investigating if and how various context factors relate to music taste (e.g., human movement, emotional status, and external factors such as temperature and lighting conditions). Other work involving context e.g. includes temporal context ([43]), listening history and weather conditions ([44]), walking pace or heartbeat rate([45, 46]), geographical location [47] and driving circumstances [48]. As for the latter work, while eight different contextual driving factors are considered, the application scenario is quite restricted and the system relies on explicit human feedback. In PHENICX, upon establishing relevant context factors to the practical application scenarios of interest, we will rather aim to rely on implicit user feedback to adhere to the requirement of unintrusiveness, which is a prerequisite for wide user acceptance.

The concept of ‘intent’ deals with the ‘why’ behind an action. In terms of information search, moving beyond textual search, search intent is now increasingly being studied for image and video domains (e.g. [49, 50]. In PHENICX, we strive to make another step forward in this field, by explicitly studying and considering search intent and particular information needs in the music domain as well.

Finally, there are two recommendation aspects which have not been studied extensively yet, but are well-known and deserve closer examination within our project. First of all, especially if different performances of the same piece are considered to be different entities in a recommender system (e.g. because the metadata does not fully match), ‘long tail’ issues [51] will occur, in which many musical items will have relatively low consumption counts. Furthermore, we wish to advance towards serendipitous findings, building forth on a model for serendipitous music retrieval and recommendation proposed by Schedl et al. [52], and establishing proper evaluation methodologies for this, as e.g. presented in [53].

6. EXPLORATION AND INTERACTION

In the area of exploration and interaction, the project will research on two different areas. The first one is to provide meaningful visualization of musical pieces and performances from different layers as extracted by multimodal piece and performance analysis (Section 3-4). The second one is to allow the audience to interact with the concert from different perspectives according to source (section 3.4) and user profile (section 5).

6.1 Visualisation of music pieces and performances

We can distinguish two qualitatively distinct sources of information to be exploited in visualisation: the score itself, from which users can be informed about melodic lines, harmony, motifs or structure; and a specific performance, the specific way the written music was actually realized. Performance differences such as timing, phrasing and dynamics are quite notable for the symphonic repertoire, being one of the main sources of engagement and enjoyment for the audience. Moreover, both dimensions – score and performance – can inform and enrich each other. For users with different musical backgrounds (e.g. naive listeners, basic musical training, professional musicians), the most relevant musical descriptions and their corresponding visualisations will differ in terms of modalities, types and specificity levels. As outlined in section 3.1, we will employ different types of automatic music descriptors, which take different temporal scales into account, ranging from short-time melodic description to global key properties.

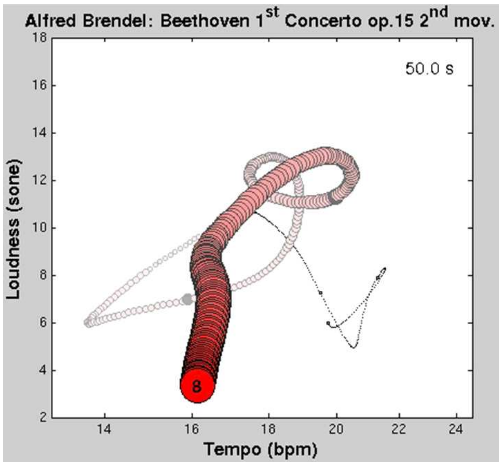

Figure 3. ‘Performance Worm’ visualisation of expressive timing and dynamics in Beethoven’s First Piano Concerto.

Existing real-time music visualisation tools for tonality include dynamic tracking in both audio and symbolic domains [54], [55], but most of them are mostly intended to inform musicians (in music theory terms). We propose an extension of temporal multi-scale techniques for the analysis and representation of a variety of audio and/or symbolic features, through time-scale summarisation and mapping into feature spaces and geometrical colourspaces. This has been proposed for temporal multiscale tonality representations and interactive navigation of music pieces [5], illustrated in figure 2. For tonality, some of these models have been validated as perceptually relevant by cognitive psychology methodologies [56], and they have been used to inform real-time music performances, such as jazz improvisations. This approach is being currently extended beyond usual tonal simplifications, covering other musical (non-tonal) representation domains, and as interactive controllers for music creation.

In addition to properties of the music (the composition) itself, we also want to visualise interesting aspects to the specific *performance*. Examples of performance visualization are the Performance Worm [57] and more general phase plane representations [58]. While these uncover rather local timing and dynamics patterns, multi-level visualisations such as the Timescapes used in [59] can visualise how expressive timing shapes a piece at many levels simultaneously, making explicit also long-term developments and large-scale structure in a performance.

For live, real-time visualisation on stage, visualisation methods must be integrated with a real-time performance tracker, which is less trivial than it may seem. For instance, all of the above-mentioned methods rely on some kind of

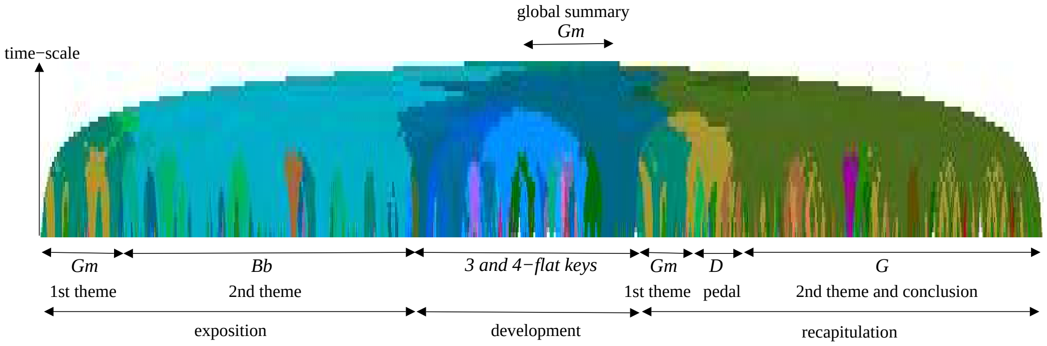

Figure 2. Multiscale tonality visualization of Finale of Haydn's "Rider" String Quartet, op.74 n.3 (in Gm).

smoothing over time, and in doing so, effectively need to look 'into the future' of a given point in time. Predictive performance models may alleviate this problem. Moreover, timing and (global) dynamics are only two aspects of a much more complex and multi-faceted phenomenon. We will also investigate new ways of visualising dimensions such as articulation, balance of the voices/instrument sections in the orchestra, etc.

6.2 Multi-perspective audio processing: source auralisation

With the separated signals and information about the instruments location as presented in Section 3.4, we can attend a meaningful process of auralisation. Recent approaches have addressed the concept of upmixing (i.e. providing a spatial multi-channel output from a mono or stereo audio signal) by means of source separation techniques [60]. The challenge here is to provide a meaningful auralisation of the orchestral content by exploring different options from an acoustic zoom for a given instrument section, to virtually place the listener in a specific position on stage.

6.3 User-generated and multi-perspective concert video

Finally, it is of relevance to mention recent approaches regarding multi-perspective and user-generated concert video content. This topic has been emerging in several recent works, and since such content reflects collective strategies taking into account a particular person's view on a concert, it can be of interest for PHENICX too.

As for existing work,in [61] audio fingerprints are used to synchronise multiple user-generated concert video recordings, and key moments within a concert are detected based on the amount of overlap between multiple user clips. In [62], an automated video mashup system is presented, synchronising different user videos through camera flashes, and generating an aesthetic mashup result based on formalised requirements as elicited from video camera users. Finally, in [63] a concert video browser is demonstrated based on segment-level visual concept detectors, in which crowdsourcing mechanisms are used to improve the indexing results. It is striking that none of these existing methods actually base their analyses or evaluations on musical audio content, nor do they try to relate obtained results to musical content. In contrast, in PHENICX, since multi-perspective video and social information are to be used to get a better insight into the live musical performance, musical aspects will need to be taken into account explicitly.

Acknowledgments

The PHENICX consortium includes the Music Technology Group, Universitat Pompeu Fabra (project coordinator), Delft Multimedia Information Retrieval Lab - Technische Universiteit Delft (scientific coordinator), Department of Computational Perception, Johannes Kepler Universitt Linz, Royal Concertgebouw Orchestra, Video Dock BV, Austrian Research Institute for Artificial Intelligence and Escola Superior de Música de Catalunya. The project is funded by the European Union Seventh Framework Programme FP7 / 2007- 2013 under grant agreement n 601166.

References

[1] C. C. S. Liem, R. van der Sterren, M. van Tilburg, Álvaro Sarasúa, J. Bosch, J. Janer, M. Melenhorst, E. Gómez, and A. H. and, "The phenicx project: Towards interactive and social classical music performance consumption," in *Workshop on Interactive Content Consumption at EuroITV*, submitted.

[2] D. Bogdanov, J. Serrà, N. Wack, P. Herrera, and X. Serra, "Unifying low-level and high-level music similarity measures," *IEEE Transactions on Multimedia*, vol. 13, pp. 687–701, 08/2011 2011.

[3] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, pp. 1759–1770, 08/2012 2012.

[4] A. Holzapfel, M. Davies, J. Zapata, J. Oliveira, and F. Gouyon, "Selective sampling for beat tracking evaluation," *IEEE Transactions on Audio Speech and Language Processing*, 2012.

[5] A. Martorell and E. Gómez, "Two-dimensional visual inspection of pitch-space, many time-scales and tonal uncertainty over time," in *3rd International Conference on Mathematics and Computation in Music*, Paris, June 2011.

[6] K. Seyerlehner, G. Widmer, and T. Pohle, "Fusing Block-Level Features for Music Similarity Estimation," in *International Conference on Digital Audio Effects*, Graz, Austria, September 2010.

[7] M. Schedl and T. Pohle, "Enlightening the Sun: A User Interface to Explore Music Artists via Multimedia Content," *Multimedia Tools and Applications: Special Issue on Semantic and Digital Media Technologies*, vol. 49, no. 1, pp. 101–118, August 2010.

[8] J. H. Kim, B. Tomasik, and D. Turnbull, "Using Artist Similarity to Propagate Semantic Information," in *International Society for Music Information Retrieval Conference)*, Kobe, Japan, October 2009.

[9] M. Sordo, "Semantic Annotation of Music Collections: A Computational Approach," Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, Spain, 2012.

[10] K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees, "Automatic Music Tag Classification based on Block-Level Features," in *Sound and Music Computing Conference*, Barcelona, Spain, July 2010.

[11] M. I. Mandel, R. Pascanu, D. Eck, Y. Bengio, L. M. Aiello, R. Schifanella, and F. Menczer, "Contextual Tag Inference," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 7S, no. 1, pp. 32:1–32:18, 2011.

[12] M. I. Mandel and D. P. W. Ellis, "A Web-Based Game for Collecting Music Metadata," *Journal of New Music Research*, vol. 37, no. 2, pp. 151–165, 2008.

[13] Y. Raimond, C. Sutton, and M. Sandler, "Interlinking Music-Related Data on the Web," *IEEE MultiMedia*, vol. 16, no. 2, pp. 52–63, 2009.

[14] M. Schedl, G. Widmer, P. Knees, and T. Pohle, "A music information system automatically generated via web content mining techniques," *Information Processing & Management*, vol. 47, 2011.

[15] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures," in *International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, 2000, pp. 2985–2988 vol.5.

[16] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, vol. 3, no. 3, pp. 177–180, 2003.

[17] J. Durrieu, G. Richard, B. David, and C. Févotte, "Source/Filter Model for Unsupervised Main Melody Extraction From Polyphonic Audio Signals," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 18, no. 3, pp. 564–575, Mar. 2010.

[18] S. Ewert and M. Müller, "Using Score-Informed Constraints for NMF-based Source Separation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Kyoto, Japan: IEEE, 2012.

[19] A. Friberg and J. Sundberg, "Does Music Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners," *Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1469–1484, 1999.

[20] N. P. Todd, "A Computational Model of Rubato," *Contemporary Music Review*, vol. 3, no. 1, pp. 69–88, 1989.

[21] S. Flossmann, W. Goebl, M. Grachten, B. Niedermayer, and G. Widmer, "The Magaloff Project: An Interim Report," *Journal of New Music Research*, vol. 31, no. 4, pp. 363–377, 2010.

[22] M. Grachten and G. Widmer, "Linear Basis Models for Prediction and Analysis of Musical Expression," *Journal of New Music Research*, vol. 41, no. 4, pp. 311–322, 2012.

[23] S. Dixon and G. Widmer, "MATCH: A Music Alignment Tool Chest," in *International Society for Music Information Retrieval Conference*, London, UK, 2005.

[24] M. Müller, H. Mattes, and F. Kurtz, "An Efficient Multiscale Approach to Audio Synchronization," in *International Society for Music Information Retrieval Conference*, Victoria, Canada, 2006.

[25] B. Niedermayer, "Accurate Audio-to-Score Alignment: Data Acquisition in the Context of Computational Musicology," Ph.D. dissertation, Johannes Kepler University Linz, Austria, 2012.

[26] A. Cont, "A Coupled Duration-focused Architecture for Real-time Music-to-Score Alignment," *IEEE TPAMI*, vol. 32, no. 6, pp. 974–987, 2010.

[27] C. Raphael, "Aligning Music Audio with Symbolic Scores Using a Hybrid Graphical Model," *Machine Learning*, vol. 65, no. 2-3, pp. 389–409, 2006.

[28] M. Müller and S. Ewert, "Towards Timbre-invariant Audio Features for Harmony-based Music," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 649–662, 2010.

[29] A. Arzt and G. Widmer, "Towards Effective 'Anytime' Music Tracking," in *STAIRS Conference*, Lisbon, Portugal, 2010.

[30] C. Fremerey, M. Müller, and M. Clausen, "Handling Repeats and Jumps in Score-performance Synchronization," in *International Society for Music Information Retrieval Conference*, Utrecht, Netherlands, 2010.

[31] A. Cont, "ANTESCOFO: Anticipatory Synchronisation and Control of Interactive Parameters in Computer Music," in *International Computer Music Conference*, 2008.

[32] C. Raphael, "Music Plus One and Machine Learning," in *International Conference on Machine Learning*, 2010.

[33] T. Schlömer, B. Poppinga, H. Henze, and S. Boll, "Gesture Recognition with a Wii Controller," in *International Conference on Tangible and Embedded Interaction*, New York, 2008.

[34] K. Kin, B. Hartmann, and T. DeRose, "Proton: Multitouch Gestures as Regular Expressions," in *CHI 2012*, 2012.

[35] C. F. Julià, S. Jordà, and N. Earnshaw, "Gestureagents: An agent-based framework for concurrent multi-task multi-user interaction," in *TEI 2013*.

Barcelona, Spain: ACM, February 2013.

[36] E. Maestre, A. Perez, and R. Ramirez, "Gesture Sampling for Instrumental Sound Synthesis," in *International Computer Music Conference*, 2010.

[37] R. Godoy, E. Haga, and A. Jensenius, "Playing"Air Instruments": Mimicry of Sound-producing Gestures by Novices and Experts," in *Gesture in Human-Computer Intraction and Simulation*, 2006.

[38] Z. Cheng, J. Caverlee, and K. Lee, "You Are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users," in *ACM International Conference on Information and Knowledge Management*, October 26-30 2010, pp. 759–768.

[39] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short Text Classification in Twitter to Improve Information Filtering," in *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Geneva, Switzerland, July 19–23 2010.

[40] M. Clements, A. P. de Vries, and M. J. Reinders, "The Task Dependent Effect of Tags and Ratings on Social Media Access," *ACM Transactions on Information Systems*, vol. 2, no. 3, November 2008.

[41] S. D. Cedric S. Mesnage, Asma Rafiq and R. Brixtel, "Music Discovery with Social Networks," in *Workshop on Music Recommendation and Discovery*, Chicago, IL, USA, October 2011.

[42] S. Cunningham, S. Caulder, and V. Grout, "Saturday Night or Fever? Context-Aware Music Playlists," in *International Audio Mostly Conference of Sound in Motion*, October 2008.

[43] T. Cebrián, M. Planagumà, P. Villegas, and X. Amatriain, "Music Recommendations with Temporal Context Awareness," in *ACM Conference on Recommender Systems*, Barcelona, Spain, 2010.

[44] J. S. Lee and J. C. Lee, "Context Awareness by Case-Based Reasoning in a Music Recommendation System," in *Ubiquitous Computing Systems*, ser. Lecture Notes in Computer Science, H. Ichikawa, W.-D. Cho, I. Satoh, and H. Youn, Eds. Springer Berlin / Heidelberg, 2007, vol. 4836, pp. 45–58.

[45] B. Moens, L. van Noorden, and M. Leman, "D-Jogger: Syncing Music with Walking," in *Sound and Music Computing Conference*, Barcelona, Spain, 2010, pp. 451–456.

[46] H. Liu and J. H. M. Rauterberg, "Music Playlist Recommendation Based on User Heartbeat and Music Preference," in *International Conference on Computer Technology and Development*, Bangkok, Thailand, 2009, pp. 545–549.

[47] M. Kaminskas and F. Ricci, "Location-Adapted Music Recommendation Using Tags," in *User Modeling, Adaption and Personalization*, ser. Lecture Notes in Computer Science, J. Konstan, R. Conejo, J. Marzo, and N. Oliver, Eds. Springer Berlin / Heidelberg, 2011, vol. 6787, pp. 183–194.

[48] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, K.-H. Lüke, and R. Schwaiger, "InCarMusic: Context-Aware Music Recommendations in a Car," in *International Conference on Electronic Commerce and Web Technologies (EC-Web)*, Toulouse, France, Aug–Sep 2011.

[49] M. Lux, C. Kofler, and O. Marques, "A classification scheme for user intentions in image search," in *ACM CHI '10*, 2010.

[50] A. Hanjalic, C. Kofler, and M. Larson, "Intent and its discontents: the user at the wheel of the online video search engine," in *ACM Multimedia*, 2012.

[51] O. Celma, *Music Recommendation and Discovery – The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Berlin, Heidelberg, Germany: Springer, 2010.

[52] M. Schedl, D. Hauger, and D. Schnitzer, "A Model for Serendipitous Music Retrieval," in *16th International Conference on Intelligent User Interfaces: 2nd International Workshop on Context-awareness in Retrieval and Recommendation*, Lisbon, Portugal, February 14 2012.

[53] Yuan Cao Zhang, Diarmuid O Seaghdha, Daniele Quercia, Tamas Jambor, "Auralist: Introducing Serendipity into Music Recommendation," in *ACM Int'l Conference on Web Search and Data Mining*.

[54] E. Chew, "Towards a mathematical model of tonality," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[55] P. Toiviainen, "Visualization of tonal content in the symbolic and audio domains," *Tonal theory for the digital age*, p. 187, 2007.

[56] P. Toiviainen and C. Krumhansl, "Measuring and modeling real-time responses to music: The dynamics of tonality induction," *Perception*, vol. 32, pp. 741–766, 2003.

[57] J. Langner and W. Goebl, "Visualizing Expressive Performance in Tempo-Loudness Space," *Computer Music Journal*, vol. 27, no. 4, pp. 69–83, 2003.

[58] M. Grachten, W. Goebl, S. Flossmann, and G. Widmer, "Phase-plane Representation and Visualisation of Gestural Structure in Expressive Timing," *Journal of New Music Research*, vol. 38, no. 2, pp. 183–195, 2009.

[59] C. Sapp, "Comparative analysis of multiple musical performances," in *International Conference on Music Information Retrieval*, Vienna, Austria, September 23-27 2007, pp. 497–500.

[60] D. Fitzgerald, "Upmixing from mono - a source separation approach," in *International Conference on DSP*, 2011, pp. 1–7.

[61] L. Kennedy and M. Naaman, "Less talk, more rock: automated organization of community-contributed collections of concert videos," in *WWW '09*, 2009.

[62] P. Shrestha, P. H. N. de With, H. Weda, M. Barbieri, and E. H. L. Aarts, "Automatic mashup generation from multiple-camera concert recordings," in *ACM Multimedia*, 2010.

[63] C. G. M. Snoek, B. Freiburg, J. Oomen, and R. Ordelman, "Crowdsourcing rock n' roll multimedia retrieval," in *ACM Multimedia*, 2010.