ROADMAP FOR

# MUSIC INFORMATION RESEARCH

MIRES.CC

ISBN: 978-2-9540351-1-6

This book was designed by Geoffroy Peeters and the MIReS Consortium.
It was typeset using the LaTeX3 typesetting system on distributed GNU/Linux platforms.
The editing of this book is based on the templates created for the Roadmap for Sound and Music Computing of the S2S$^2$ Consortium.



Please cite this document as:

Xavier Serra, Michela Magas, Emmanouil Benetos, Magdalena Chudy, Simon Dixon, Arthur Flexer, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Sergi Jorda, Oscar Paytuvi, Geoffroy Peeters, Jan Schlüter, Hugues Vinet, Gerhard Widmer, "Roadmap for Music Information ReSearch", Geoffroy Peeters (editor), 2013, Creative Commons BY-NC-ND 3.0 license, ISBN: 978-2-9540351-1-6

The most recent version of this document may be downloaded from
http://mires.eecs.qmul.ac.uk/

Roadmap for

# Music Information ReSearch

Version 1.0.0

The MIReS Consortium



The most recent version of this document may be downloaded from
http://mires.eecs.qmul.ac.uk/

# The MIReS Consortium

Music Technology Group (MTG), Universitat Pompeu Fabra, Barcelona, Spain

Stromatolite Design Lab, London, UK

Austrian Research Institute for Artificial Intelligence (OFAI) of the Austrian Society for Cybernetic Studies in Vienna, Austria

Instituto de Engenharia de Sistemas e Computadores (INESC TEC), Porto, Portugal

Institut de Recherche et Coordination Acoustique et Musique (IRCAM), Paris, France

Centre for Digital Music (C4DM), Queen Mary, University of London, UK

Barcelona Music and Audio Technology (BMAT), Barcelona, Spain

| | |
|---|---|
| Coordinators | Michela Magas, Xavier Serra |
| Main Contributors | Emilia Gomez (MTG)<br>Perfecto Herrera (MTG)<br>Sergi Jorda (MTG)<br>Xavier Serra (MTG)<br><br>Michela Magas (Stromatolite)<br><br>Arthur Flexer (OFAI)<br>Jan Schlüter (OFAI)<br>Gerhard Widmer (OFAI)<br><br>Fabien Gouyon (INESC TEC)<br><br>Geoffroy Peeters (IRCAM)<br>Hugues Vinet (IRCAM)<br><br>Emmanouil Benetos (C4DM)<br>Magdalena Chudy (C4DM)<br>Simon Dixon (C4DM)<br><br>Oscar Paytuvi (BMAT) |
| Additional Contributors | Markus Schedl (Johannes Kepler University, Linz)<br>Jaume Vintró (BMAT) |
| Editor | Geoffroy Peeters |

# Executive Summary

This document proposes a Roadmap for Music Information Research with the aim to expand the context of this research field from the perspectives of technological advances, user behaviour, social and cultural aspects, and exploitation methods. The Roadmap embraces the themes of multimodality, multidisciplinarity and multiculturalism, and promotes ideas of personalisation, interpretation, embodiment, findability and community.

From the perspective of technological advances, the Roadmap defines Music Information Research as a research field which focuses on the processing of digital data related to music, including gathering and organisation of machine-readable musical data, development of data representations, and methodologies to process and understand that data. More specifically, this section of the Roadmap examines (i) musically relevant data; (ii) music representations; (iii) data processing methodologies; (iv) knowledge-driven methodologies; (v) estimation of elements related to musical concepts; and (vi) evaluation methodologies. A series of challenges are identified, related to each of these research subjects, including: (i) identifying all relevant types of data sources describing music, ensuring quality of data, and addressing legal and ethical issues concerning data; (ii) investigating more meaningful features and representations, unifying formats and extending the scope of ontologies; (iii) enabling cross-disciplinary transfer of methodologies, integrating multiple modalities of data, and adopting recent machine learning techniques; (iv) integrating insights from relevant disciplines, incorporating musicological knowledge and strengthening links to music psychology and neurology; (v) separating the various sources of an audio signal, developing style-specific musical representations and considering non-Western notation systems; (vi) promoting best practice evaluation methodology, defining meaningful evaluation methodologies and targeting long-term sustainability of MIR. Further challenges can be found by referring to the "Specific Challenges" section under each subject in the Roadmap.

In terms of user behaviour, the Roadmap addresses the user perspective, both in order to understand the user roles within the music communication chain and to develop technologies for the interaction of these users with music data. User behaviour is examined by identifying the types of users related to listening, performing or creating music. User interaction is analysed by addressing established Human Computer Interaction methodologies, and novel methods of Tangible and Tabletop Interaction. Challenges derived from these investigations include analysing user needs and behaviour carefully, identifying new user roles related to music activities; developing tools and open systems which automatically adapt to the user; designing MIR-based systems more holistically; addressing collaborative, co-creative and sharing multi-user applications, and expanding MIR interaction beyond the multi-touch paradigm.

Social and cultural aspects define music as a social phenomenon centering on communication and on the context in which music is created. Within this context, Music Information Research aims at processing musical data that captures the social and cultural context and at developing data processing methodologies with which to model the whole musical phenomenon. The Roadmap analyses specifically music-related collective influences, trends and behaviours, and multiculturalism. Identified challenges include promoting methodologies for modeling music-related social and collective behavior, adapting complex networks and dynamic systems, analysing interaction and activity in social music networks, identifying music cultures that can be studied from a data driven perspective, gathering culturally relevant data for different music cultures, and identifying specific music characteristics for each culture.

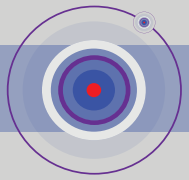
The exploitation perspective considers Music Information Research as relevant for producing exploitable technologies for organising, discovering, retrieving, delivering, and tracking information related to music, in order to enable improved user experience and commercially viable applications and services for digital media stakeholders. This section of the Roadmap focuses specifically on music distribution applications, creative tools, and other exploitation areas such as applications in musicology, digital libraries, education and eHealth. Challenges include demonstrating better exploitation possibilities of MIR technologies, developing systems that go beyond

recommendation and towards discovery, developing music similarity methods for particular applications and contexts, developing methodologies of MIR for artistic applications, developing real-time MIR tools for performance, developing creative tools for commercial environments, producing descriptors based on musicological concepts, facilitating seamless access to distributed data in digital libraries, overcoming barriers to uptake of technology in music pedagogy and expanding the scope of MIR applications in eHealth. For a full list of challenges, please refer to the relevant sections of the Roadmap.

The Music Information Research Roadmap thus identifies current opportunities and challenges and reflects a variety of stakeholder views, in order to inspire novel research directions for the MIR community, and further inform policy makers in establishing key future funding strategies for this expanding research field.

# Contents

# INTRODUCTION



# CHAPTER 1
# Introduction

For the purpose of this Roadmap we consider Music Information Research (MIR) as a field that covers all the research topics involved in the understanding and modelling of music and that use information processing methodologies. We view this research to be very much within the field of Information Technologies thus with the major aim of developing methods and technologies with which to process musically relevant data and develop products and services with which to create, distribute and interact with music information.

This Roadmap aims at identifying research challenges of relevance to the MIR community. The document should also be of relevance to the policy makers that need to understand the field of information technologies, identifying the state of the art in music information technology research and the relevant research problems that should be worked on.

The Roadmap has been elaborated by the researchers involved in the MIReS project. They have been responsible to get input and feedback from the different MIR stakeholders. Many events have been organised to gather information from experts throughout the elaboration of the Roadmap. This document is the result of a very participative process in which many people have been involved. The task of the writers of the document has mainly been an editorial one, trying to capture and summarise what the different stakeholders expressed as being relevant for the future of the field.

There have been some European initiatives with similar aims. It is specially relevant to mention the S2S$^2$ Coordination Action, within which a Roadmap was elaborated and published in 2007 [1]. That Roadmap covered the whole Sound and Music Computing field, thus it had a broader perspective than MIR and it also went beyond research issues. A more recent and also relevant document has been the discussion paper entitled "Musicology (Re-) Mapped" promoted by the European Science Foundation and published in 2012 [2]. This paper covered the field of Musicology from a modern perspective and discussed some of the current research trends, some of which overlap with MIR. It is also relevant to mention the report on the 3rd CHORUS+ Think-Tank that took place at MIDEM 2011 and addressed the future of music search, access and consumption from an industrial perspective [3].

This Roadmap focuses on presenting and discussing research challenges, thus it does not aim to cover organisational, industrial, or educational aspects. No attempt is made to predict the future of research in MIR; we believe that this is not possible. The challenges have been identified by studying and using the current state of the art in MIR and related disciplines. We are very much aware that many of the great technological and scientific discoveries result from disruptive changes and developments, and these are impossible to predict using this approach.

The challenges have been grouped into four sections, each one reflecting a different emphasis and perspective: technological, user, sociocultural, and exploitation. The technological perspective is the more traditional one used in MIR, reflecting the core scientific and technical challenges. The other three sections aim to examine the field from non-traditional perspectives, thus emphasising important, though often ignored views, which can give us important insights into our research. Figure 1 shows the structure of the document in the form of a diagram indicating the relationships between the document sections and perspectives.

[1] http://smcnetwork.org/roadmap

[2] http://www.esf.org/human

[3] http://avmediasearch.eu/public/files/Chorus+_MusicThinkTank-Report_TheFutureOfMusicSearchAccessAndConsumption_final.pdf

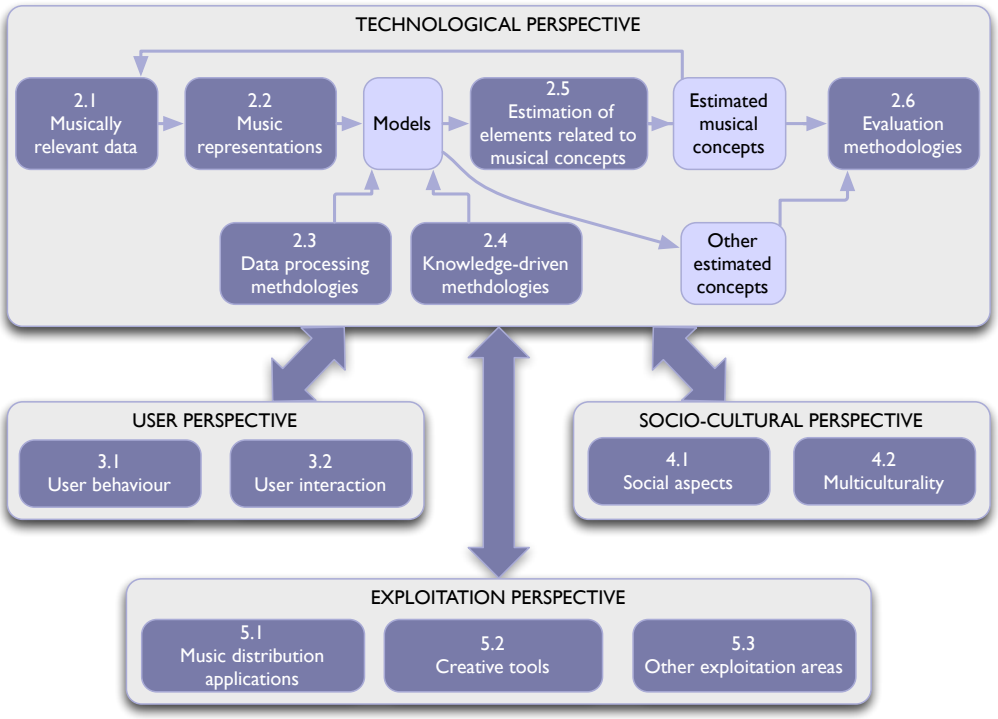
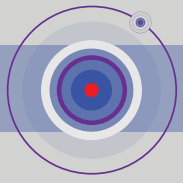


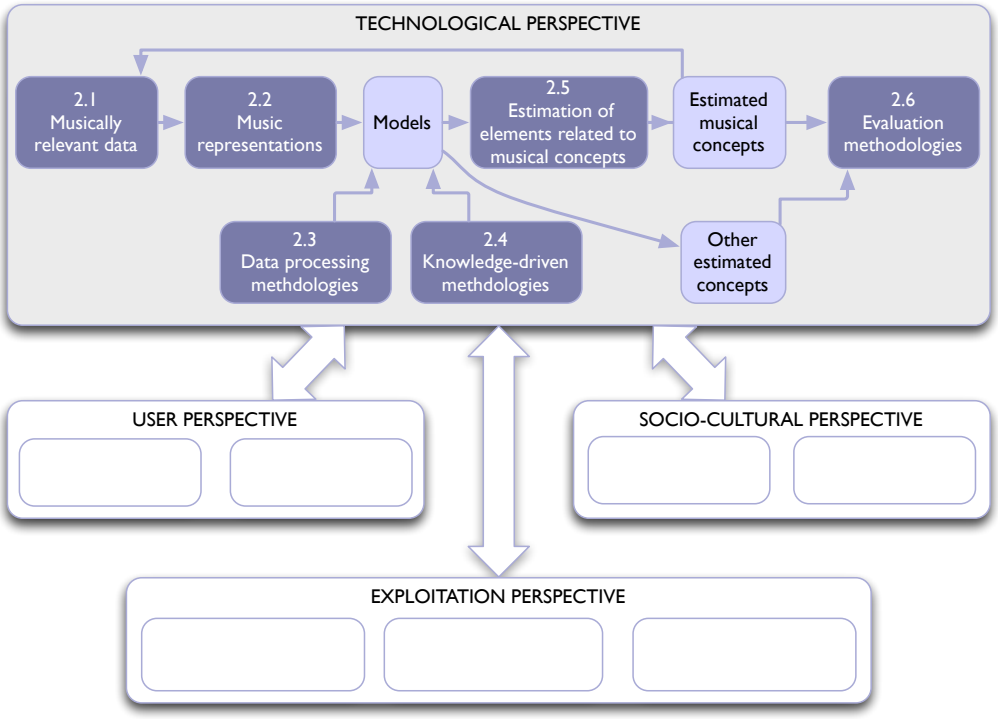

Figure 1: Diagram showing the relations between the different sections of the documents and MIR topics discussed in them.

# TECHNOLOGICAL PERSPECTIVE



CHAPTER 2
# Technological perspective

*Music Information Research focuses on the processing of digital data related to music. This includes gathering and organisation of machine-readable musical data, development of data representations, and methodologies to process and understand that data, taking into account domain knowledge and bringing expertise from relevant scientific and engineering disciplines.*



## 2.1 MUSICALLY RELEVANT DATA

We define "musically relevant data" as any type of machine-readable data that can be analysed by algorithms and that can give us relevant information for the development of musical applications. The main challenge is to gather musically relevant data of sufficient quantity and quality to enable music information research that respects the broad multi-modality of music. After all, music today is an all-encompassing experience that is an important part of videos, computer games, Web applications, mobile apps and services, specialised blogs, artistic applications, etc. Therefore we should be concerned with the identification of all sources of musically relevant data, the proper documentation of the process of data assembly and resolving of all legal and ethical issues concerning the data. Sufficient quantity and quality of data is of course the prerequisite for any kind of music information research. To make progress in this direction it is necessary that the research community works together with the owners of data, be they copyright holders in the form of companies or individual persons sharing their data. Since music information research is by definition a data intensive science, any progress in these directions will have immediate impact on the field. It will enable a fostering and maturing of our research and, with the availability of new kinds of musically relevant data, open up possibilities for new kinds of research and applications.

### 2.1.1 State of the art

Music Information Research (MIR) is so far to a large degree concerned with audio, neglecting many of the other forms of media where music also plays an important role. As recently as ten years ago, the main media concerned with music were represented by audio recordings on CDs, terrestrial radio broadcasts, music videos on TV, and printed text in music magazines. Today music is an all-encompassing experience that is an important part of videos, computer games, Web applications, mobile apps and services, artistic applications, etc. In addition to printed text on music there exist a vast range of web-sites, blogs and specialised communities caring and publishing about music. Therefore it is necessary for MIR to broaden its horizons and include a multitude of yet untapped data sources in its research agenda. Data that is relevant for Music Information Research can be categorised into four different subgroups: (i) audio-content is any kind of information computed directly from the audio signal; (ii) music scores is any type of symbolic notation that is normally used for music performance and that captures the musical intention of a composer; (iii) music-context is all information relevant to music which is not directly computable from the audio itself or the score, e.g. cover artwork, lyrics, but also artists' background and collaborative tags connected to the music; (iv) user-context is any kind of data that allows us to model the users in specific usage settings.

Let us start with the most prevalent source of data: audio content and any kind of information computed directly from the audio. Such information is commonly referred to as "features", with a certain consensus on distinguishing between low-level and high-level features (see e.g. [4]). Please see section 2.2 for an overview of different kinds of features. It is obvious that audio content data is by far the most widely used and researched form of information in our community. This can e.g. be seen by looking at the tasks of the recent "Music Information Retrieval Evaluation eXchange" (MIREX 2012 [1]). MIREX is the foremost yearly community-based framework for formal evaluation of MIR algorithms and systems. Out of the 16 tasks, all but one (Symbolic Melodic Similarity) deal with audio analysis including challenges like: Audio Classification, Audio Melody Extraction, Cover Song Identification, Audio Key Detection, Structural Segmentation and Audio Tempo Estimation. Concerning the availability of audio content data there are several legal and copyright issues. Just to give an example, the by far largest data set in MIR, the "Million Songs Dataset" [2], does not include any audio, only the derived features. In case researchers need to compute their own features they have to use services like "7-Digital" to access the audio. Collections that do contain audio as well are usually very small like e.g. the well known "GTzan" collection assembled by George Tzanetakis in 2002 consisting of 1000 songs freely available from the Marsyas webpage [3].

[1] http://www.music-ir.org/mirex/wiki/2012:Main_Page
[2] http://labrosa.ee.columbia.edu/millionsong
[3] http://marsyasweb.appspot.com/download/data_sets

The largest freely downloadable audio data set is the "1517 Artists" collection [4] consisting of 3180 songs from 1517 artists. There also exist alternative collaborative databases of Creative Commons Licensed sounds like Freesound [5].

An important source of information to start with is of course symbolic data, thus the score of a piece of music if it is available in a machine readable format, like MIDI, Music XML, sequencer data or other kinds of abstract representations of music. Such music representations can be very close to audio content like e.g. the score to one specific audio rendering but they are usually not fully isomorphic. Going beyond more traditional annotations, recent work in MIR [17] turned its attention to machine readable tablatures and chord sequences, which are a form of hand annotated scores available in non-standardised text files (e.g. "ultimate guitar" [6] contains more than 2.5 million guitar tabs). At the first MIR conference [7] a large part of the contributed papers were concerned with symbolical data. Almost ten years later this imbalance seems to have reversed with authors [8] lamenting that "ISMIR must rebalance the portfolio of music information types with which it engages" and that "research exploiting the symbolic aspects of music information has not thrived under ISMIR". Symbolic annotations of music present legal and copyright issues just like audio, but substantial collections (e.g. of MIDI files [8]) do exist.

Music context is all information relevant to a music item under consideration that is not extracted from the respective audio file itself or the score (see e.g. [22] for an overview). A large part of research on music context is strongly related to web content mining. Over the last decade, mining the World Wide Web has been established as another major source of music related information. Music related data mined from the Web can be distinguished into "editorial" and "cultural" data. Whereas editorial data originates from music experts and editors often associated with the music distribution industry, cultural data makes use of the wisdom of the crowd by mining large numbers of music related websites including social networks. Advantages of web based MIR are the vast amount of available data as well as its potential to access high-level semantic descriptions and subjective aspects of music not obtainable from audio based analysis alone. Data sources include artist-related Web pages, published playlists, song lyrics or blogs and twitter data concerned with music. Other data sources of music context are collaborative tags, mined for example from last.fm [16] or gathered via tagging games [25]. A problem with information obtained automatically from the Web is that it is inherently noisy and erroneous which requires special techniques and care for data clean-up. Data about new and lesser known artists in the so-called "long tail" is usually very sparse which introduces an unwanted popularity bias [5]. A list of data sets frequently used in Web-based MIR is provided by Markus Schedl [9]. The "Million Songs Dataset" [10] contains some web-related information like e.g. tag information provided by Last.fm.

A possibly very rich source of additional information on music content that has so far received little attention in our community is music videos. The most prominent source for music videos is YouTube [11], but alternatives like Vimeo [12] exist. Uploaded material contains anything from amateur clips to video blogs to complete movies, with a large part of it being music videos. Whereas a lot of the content on YouTube has been uploaded by individuals which may entail all kinds of copyright and legal issues, some large media companies have lately decided to also offer some of their content. There exists a lively community around the so-called TRECVid campaign [13], a forum, framework and conference series on video retrieval evaluation. One of the major tasks in video information retrieval is automatic labelling of videos, e.g. according to genre, which can be done either globally or locally [2]. Typical information extracted from videos are visual descriptors like color, its entropy and variance, hue, as well as temporal cues like cuts, fades, dissolves. Object-based features like the occurrence of faces or text and motion-based information like motion density and camera movement are also of interest. Text-based

[4] http://www.seyerlehner.info
[5] http://www.freesound.org
[6] http://www.ultimate-guitar.com
[7] http://ismir2000.ismir.net
[8] http://www.free-midi.org
[9] http://www.cp.jku.at/people/schedl/datasets.html
[10] http://labrosa.ee.columbia.edu/millionsong
[11] http://www.youtube.com
[12] http://www.vimeo.com
[13] http://trecvid.nist.gov

information derived from sub-titles, transcripts of dialogues, synopsis or user tags is another valuable source. A potentially very promising approach is the combined analysis of a music video and its corresponding audio, pooling information from both image and audio signals. Combination of general audio and video information is an established topic in the literature, see e.g. [26] for an early survey. There already is a limited amount of research explicitly on music videos exploiting both the visual and audio domain [9]. Although the TRECVid evaluation framework supports a "Multimedia event detection evaluation track" consisting of both audio and video, to our knowledge no data set dedicated specifically to music videos exists.

Another yet untapped source are machine readable tetxs on musicology that are available online (e.g. via Google Books [14]). Google books is a search engine that searches the full text of books if they have already been scanned and digitised by Google. This offers the possibility of using Natural Language Processing tools to analyse text books on music, thereby introducing MIR topics to the new emerging field of digital humanities.

As stated above, user-context data is any kind of data that allows us to model a single user in one specific usage setting. In most MIR research and applications so far, the prospective user is seen as a generic being for whom a generic one-for-all solution is sufficient. Typical systems aim at modeling a supposedly objective music similarity function which then drives music recommendation, play-listing and other related services. This however neglects the very subjective nature of music experience and perception. Not only do different people perceive music in different ways depending on their likes, dislikes and listening history, but even one and the same person will exhibit changing tastes and preferences depending on a wide range of factors: time of day, social situation, current mood, location, etc. Personalising music services can therefore be seen as an important topic of future MIR research.

Following recent proposals (see e.g. [23]), we distinguish five different kinds of user context data: (i) Environment Context, (ii) Personal Context, (iii) Task Context, (iv) Social Context, (v) Spatio-temporal Context. The environmental context is defined as all entities that can be measured from the surroundings of a user, like presence of other people and things, climate including temperature and humidity, noise and light. The personal context can be divided into the physiological context and the mental context. Whereas physiological context refers to attributes like weight, blood pressure, pulse, or eye color, the mental context is any data describing a user's psychological aspects like stress level, mood, or expertise. Another important form of physiological context data are recordings of gestures during musical performances with either traditional instruments or new interfaces to music. The task content should describe all current activities pursued by the user including actions and activities like direct user input to smart mobile phones and applications, activities like jogging or driving a car, but also interaction with diverse messenger and microblogging services. The latter is a valuable source for a user's social context giving information about relatives, friends, or collaborators. The spatio-temporal context reveals information about a userâĂŹs location, place, direction, speed, and time. As a general remark, the recent emergence of "always on" devices (e.g. smart phones) equipped not only with a permanent Web connection, but also with various built-in sensors, has remarkably facilitated the logging of user context data from a technical perspective. Data sets on the user context are still very rare but e.g. the "user - song - play count triplets" and the Last.fm tags of the "Million Song Dataset" [15] could be said to contain this type of personal information.

The proper documentation of the process of data assembly for all kinds of musically relevant data is a major issue which has not yet gained sufficient attention by the MIR community. In [21] an overview is provided of the different practices of annotating MIR corpora. Currently, several methodologies are used for collecting these data: - creating an artificial corpus [27], recording corpora [10] or sampling the world of music according to specific criteria (Isophonics [19], Salami [24], Billboard [3], MillionSong [1]). The data can then be obtained using experts (this is the usual manual annotation [19]), using crowd-sourcing [15] or so-called games with a purpose (Listen-Game [25], TagATune [14], MajorMiner [18]) or by aggregating other content (Guitar-Tab [20] MusiXMatch, Last.fm in the case of the MillionSong). As opposed to other domains, micro-working (such as Amazon Mechanical Turk) is not (yet) a common practice in the MIR field. These various methodologies for collecting data involve various costs: from the most expensive (traditional manual annotation) to the less

[14] http://books.google.com
[15] http://labrosa.ee.columbia.edu/millionsong

expensive (aggregation or crowd-sourcing). They also involve various qualities of data. This is related to the inter-annotator and intra-annotator agreement which is rarely assessed in the case of MIR. Compared to other fields, such as natural language processing or speech, music-related data collection or creation does not follow dedicated protocols. One of the major issues in the MIR field will be to better define protocols to make reliable annotated MIR corpora. Another important aspect is how our research community relates itself to initiatives aiming at unifying data formats in the world wide web. Initiatives that come to mind are e.g. linked data [16] which is a collection of best practices for publishing and connecting structured data on the Web and, especially relevant for MIR, MusicBrainz [17] which strives to become the ultimate source of music information or even the universal lingua franca of music. It should also be clear that the diverse forms of data important for MIR are very much "live data", i.e. many data sets are constantly changing over time and need to be updated accordingly. Additionally our community should strive to create data repositories which allow open access for the research community and possibly even the general public.

### 2.1.2 Specific Challenges

- **Identify all relevant types of data sources describing music.** We have to consider the all-encompassing experience of music in all its broad multi-modality beyond just audio (video, lyrics, scores, symbolic annotations, gesture, tags, diverse metadata from web-sites and blogs, etc.). To achieve this it will be necessary to work together with experts from the full range of the multimedia community and organise the data gathering process in a more systematic way compared to what has happened so far.
- **Guarantee sufficient quality of data (both audio and meta-data).** At the moment data available to our community stems from a wide range of very different sources obtained with very different methods often not documented sufficiently. We will have to come to an agreement concerning unified data formats and protocols documenting the quality of our data. For this a dialogue within our community is necessary which should also clarify our relation to more general efforts of unifying data formats.
- **Clarify the legal and ethical concerns regarding data availability as well as its use and exploitation.** This applies to the question what data we are allowed to have and what data we should have. The various copyright issues will make it indispensable to work together with owners of content, copyright and other stakeholders. All ethical concerns on privacy issues have to be solved. The combination of multiple sources of data poses additional problems in this sense.
- **Ascertain what data users are willing to share.** One of the central goals of future MIR will be to model the tastes, behaviors and needs of individual and not just generic users. Modelling of individual users for personalisation of MIR services presents a whole range of new privacy issues since it requires handling of very detailed and possibly controversial information. This is of course closely connected to policies of diverse on-line systems concerning privacy of user data. This is also a matter of system acceptance going far beyond mere legal concerns.
- **Make available a sufficient amount of data to the research community allowing easy and legal access to the data.** Even for audio data, which has been used for research right from the beginning of MIR, availability of sufficient benchmark data sets usable for evaluation purposes is still not a fully resolved issue. To allow MIR to grow from an audio-centered to a fully multi-modal science we will need benchmark data for all these modalities to allow evaluation and comparison of our results. Hence the already existing problem of data availability will become even more severe.
- **Create open access data repositories.** It will be of great importance for the advancement of MIR to create and maintain sustainable repositories of diverse forms of music related data. These repositories should follow open access licensing schemes.

[16] http://linkeddata.org
[17] http://musicbrainz.org

## References

[1] Thierry Bertin-Mahieux, Daniel P. W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In Klapuri and Leider [13], pages 591–596.

[2] Darin Brezeale and Diane J. Cook. Automatic video classification: A survey of the literature. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, pages 416–430, 2008.

[3] John Ashley Burgoyne, Jonathan Wild, and Fujinaga Ichiro. An expert ground-truth set for audio chord recognition and music analysis. In Klapuri and Leider [13], pages 633–638.

[4] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE*, 96(4):668–696, March 2008.

[5] Òscar Celma. *Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer, 2010.

[6] Roger Dannenberg, Kjell Lemström, and Adam Tindale, editors. *ISMIR 2006, 7th International Conference on Music Information Retrieval, Victoria, Canada, 8-12 October 2006, Proceedings*, 2006.

[7] Simon Dixon, David Bainbridge, and Rainer Typke, editors. *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007, Vienna, Austria, September 23-27, 2007*. Austrian Computer Society, 2007.

[8] J. Stephen Downie, Donald Byrd, and Tim Crawford. Ten years of ismir: Reflections on challenges and opportunities. In Hirata et al. [12], pages 13–18.

[9] O. Gillet, S. Essid, and G. Richard. On the correlation of automatic audio and visual segmentations of music videos. *IEEE Trans. Cir. and Sys. for Video Technol.*, 17(3):347–355, March 2007.

[10] Masataka Goto. Aist annotation for the rwc music database. In Dannenberg et al. [6], pages 359–360.

[11] Fabien Gouyon, Perfecto Herrera, Luis Gustavo Martins, and Meinard Müller, editors. *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012, Mosteiro S.Bento Da Vitória, Porto, Portugal, October 8-12, 2012*. FEUP Edições, 2012.

[12] Keiji Hirata, George Tzanetakis, and Kazuyoshi Yoshii, editors. *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009, Kobe International Conference Center, Kobe, Japan, October 26-30, 2009*. International Society for Music Information Retrieval, 2009.

[13] Anssi Klapuri and Colby Leider, editors. *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, Miami, Florida, USA, October 24-28, 2011*. University of Miami, 2011.

[14] Edith L. M. Law, Luis von Ahn, Roger B. Dannenberg, and Mike Crawford. Tagatune: A game for music and sound annotation. In Dixon et al. [7], pages 361–364.

[15] Mark Levy. Improving perceptual tempo estimation with crowd-sourced annotations. In Klapuri and Leider [13], pages 317–322.

[16] Mark Levy and Mark Sandler. A semantic space for music derived from social tags. In Dixon et al. [7], pages 411–416.

[17] Robert Macrae and Simon Dixon. Guitar tab mining, analysis and ranking. In Klapuri and Leider [13], pages 453–458.

[18] Michael Mandel and Daniel Ellis. A Web-Based Game for Collecting Music Metadata. *Journal of New Music Research*, 37(2):151–165, June 2008.

[19] M. Mauch, C. Cannam, M. Davies, S. Dixon, C. Harte, S. Kolozali, D. Tidhar, and M. Sandler. OMRAS2 metadata project 2009. In *Late-breaking session at the 10th International Conference on Music Information Retrieval, Kobe, Japan*, 2009.

[20] Matt McVicar and Tijl De Bie. Enhancing chord recognition accuracy using web resources. In *Proceedings of 3rd International Workshop on Machine Learning and Music*, MML '10, pages 41–44, New York, NY, USA, 2010. ACM.

[21] Geoffroy Peeters and Karën Fort. Towards a (better) definition of the description of annotated mir corpora. In Gouyon et al. [11], pages 25–30.

[22] Markus Schedl and Peter Knees. Context-based Music Similarity Estimation. In *Proceedings of the 3rd International Workshop on Learning the Semantics of Audio Signals (LSAS 2009)*, Graz, Austria, December 2 2009.

[23] Markus Schedl and Peter Knees. Personalization in Multimodal Music Retrieval. In *Proceedings of the 9th Workshop on Adaptive Multimedia Retrieval (AMR'11)*, Barcelona, Spain, July 2011.

[24] Jordan Bennett Louis Smith, John Ashley Burgoyne, Ichiro Fujinaga, David De Roure, and J. Stephen Downie. Design and creation of a large-scale database of structural annotations. In Klapuri and Leider [13], pages 555–560.

[25] Douglas Turnbull, Ruoran Liu, Luke Barrington, and Gert R. G. Lanckriet. A game-based approach for collecting semantic annotations of music. In Dixon et al. [7], pages 535–538.

[26] Hualu Wang, Ajay Divakaran, Anthony Vetro, Shih-Fu Chang, and Huifang Sun. Survey of compressed-domain features used in audio-visual indexing and analysis. *J. Visual Communication and Image Representation*, 14(2):150–183, 2003.

[27] Chunghsin Yeh, Niels Bogaards, and Axel Röbel. Synthesized polyphonic music database with verifiable ground truth for multiple f0 estimation. In Dixon et al. [7], pages 393–398.

## 2.2 MUSIC REPRESENTATIONS

Data representations impact the effectiveness of MIR systems in two ways: algorithms are limited by the types of input data they receive, and the user experience depends on the way that MIR systems present music information to the user. A major challenge is to provide abstractions which enable researchers and industry to develop algorithms that meet user needs and to present music information in a form that accords with users' understanding of music. The same challenge applies to content providers, who need to select appropriate abstractions for structuring, visualising, and sonifying music information. These abstractions include features describing music information, be it audio, symbolic, textual, or image data; ontologies, taxonomies and folksonomies for structuring music information; graphical representations of music information; and formats for maintaining and sonifying music data. The development of standard representations will advance MIR by increasing algorithm and system interoperability between academia and industry as well as between researchers working on MIR subtasks, and will provide a satisfactory user experience by means of musically and semantically meaningful representations.

### 2.2.1 State of the art

While audio recordings capture musical performances with a high level of detail, there is no direct relationship between the individual audio samples and the experience of music, which involves notes, beats, instruments, phrases or melodies (the musicological perspective), and which might give rise to memories or emotions associated with times, places or events where identical or similar music was heard (the user perspective). Although there is a large body of research investigating the relationship between music and its meaning from the philosophical and psychological perspectives [e.g. 2, 14, 21], scientific research has tended to focus more on bridging the "semantic gap" between audio recordings and the abstractions that are found in various types of musical scores, such as pitches, rhythms, melodies and harmonies. This work is known as semantic audio or audio content analysis (see section 2.5).

In order to facilitate the extraction of useful information from audio recordings, a standard practice is to compute intermediate representations at various levels of abstraction. At each level, features can describe an instant in time (e.g. the onset time of a note), a segment or time interval (e.g. the duration of a chord) or the whole piece (e.g. the key of a piece). Various sets of features and methods for evaluating their appropriateness have been catalogued in the MIR literature [11, 12, 15, 17, 18].

Low-level features relate directly to signal properties and are computed according to simple formulae. Examples are the zero-crossing rate, spectral centroid and global energy of the signal. Time-domain features such as the amplitude envelope and attack time are computed without any frequency transform being applied to the signal, whereas spectral features such as centroid, spread, flatness, skewness, kurtosis and slope require a time-frequency representation such as the short time Fourier transform (STFT), the constant-Q transform (CQT) [1] or the wavelet transform [9] to be applied as a first processing step. Auditory model-based representations [13] are also commonly used as a front-end for MIR research.

Mid-level features (e.g. pitches and onset times of notes) are characterised by more complex computations, where the algorithms employed are not always successful at producing the intended results. Typically a modelling step will be performed (e.g. sinusoidal modelling), and the choice of parameters for the model will influence results. For example, in Spectral Modelling Synthesis [24], the signal is explained in terms of sinusoidal partial tracks created by tracking spectral peaks across analysis frames, plus a residual signal which contains the non-sinusoidal content. The thresholds and rules used to select and group the spectral peaks determine the amount of the signal which is interpreted as sinusoidal. This flexibility means that the representation with respect to such a model is not unique, and the optimal choice of parameters is dependent on the task for which the representation will be used.

High-level features (e.g. genre, tonality, rhythm, harmony and mood) correspond to the terms and concepts used by musicians or listeners to describe aspects of music. To generate such features, the models employed tend to be more complex, and might include a classifier trained on a relevant data set, or a probabilistic model such as a

hidden Markov model (HMM) or dynamic Bayesian network (DBN). Automatic extraction of high-level features is not reliable, which means that in practice there is a tradeoff between the expressiveness of the features (e.g. number of classes they describe) and the accuracy of the feature computation.

It should also be noted that the classification of features into categories such as "high-level" is not an absolute judgement, and some shift in usage is apparent, resulting from the search for ever higher levels of abstraction in signal descriptors. Thus features which might have been described as high-level a decade ago might now be considered to be mid-level features. Also features are sometimes described in terms of the models used to compute them, such as psychoacoustic features (e.g. roughness, loudness and sharpness) which are based on auditory models. Some features have been standardised, e.g. in the MPEG7 standard [7]. Another form of standardisation is the use of ontologies to capture the semantics of data representations and to support automatic reasoning about features, such as the Audio Feature Ontology proposed by Fazekas [5].

In addition to the literature discussing feature design for various MIR tasks, another strand of research investigates the automatic generation of features [e.g. 17]. This is a pragmatic approach to feature generation, whereby features are generated from combinations of simple operators and tested on the training data in order to select suitable features. More recently, deep learning techniques have been used for automatic feature learning in MIR tasks [6], where they have been reported to be superior to the use of hand-crafted feature sets for classification tasks, although these results have not yet been replicated in MIREX evaluations. It should be noted however that automatically generated features might not be musically meaningful, which limits their usefulness.

Much music information is not in the form of audio recordings, but rather symbolic representations of the pitch, timing, dynamics and/or instrumentation of each of the notes. There are various ways such a representation can arise. First, via the composition process, for example when music notation software is employed, a score can be created for instructing the musicians how to perform the piece. Alternatively, a score might be created via a process of transcription (automatic or manual) of a musical performance. For electronic music, the programming or performance using a sequencer or synthesiser could result in an explicit or implicit score. For example, electronic dance music can be generated, recorded, edited and mixed in the digital domain using audio editing, synthesis and sequencing software, and in this case the software's own internal data format(s) can be considered to be an implicit score representation.

In each of these cases the description (or prescription) of the notes played might be complete or incomplete. In the Western classical tradition, it is understood that performers have a certain degree of freedom in creating their rendition of a composition, which may involve the choice of tempo, dynamics and articulation, or also ornamentation and sometimes even the notes to be played for an entire section of a piece (an improvised cadenza). Likewise in Western pop and jazz music, a work is often described in terms of a sequence of chord symbols, the melody and the lyrics; the parts of each instrument are then rehearsed or improvised according to the intended style of the music. In these cases, the resulting score can be considered to be an abstract representation of the underlying musical work. One active topic in MIR research is on reducing a music score to a higher-level, abstract representation [10]. However not all styles of music are based on the traditional Western score. For example, freely improvised and many non-Western musics might have no score before a performance and no established language for describing the performance after the fact.

A further type of music information is textual data, which includes both structured data such as catalogue metadata and unstructured data such as music reviews and tags associated with recordings by listeners. Structured metadata might describe the composers, performers, musical works, dates and places of recordings, instrumentation, as well as key, tempo, and onset times of individual notes. Digital libraries use metadata standards such as Dublin Core and models such as the Functional Requirements for Bibliographic Records (FRBR) to organise catalogue and bibliographic databases. To assist interoperability between data formats and promote the possibility of automatic inference from music metadata, ontologies have been developed such as the Music Ontology [19].

Another source of music information is image data from digitised handwritten or printed music scores. For preserving, distributing, and analysing such information, systems for optical music recognition (OMR) have been under development for several years [20]. As in audio recordings, intermediate representations at various abstraction levels are computed for digitised scores. The lowest-level representation consists of raw pixels from

a digitised grayscale score, from which low-level features such as staff line thickness and vertical line distance are computed. Mid-level features include segmented (but not recognised) symbols, while higher-level features include interpreted symbols and information about connected components or symbol orientation. In order to formalise these abstractions, grammars are employed to represent allowed combinations of symbols.

Looking beyond the conceptual organisation of the data, we briefly address its organisation into specific file formats, and the development and maintenance of software to read, write and translate between these formats. For audio data, two types of representations are used: uncompressed and compressed. Uncompressed (or pulse code modulated, PCM) data consists of just the audio samples for each channel, usually prepended by a short header which specifies basic metadata such as the file format, sampling rate, word size and number of channels. Compression algorithms convert the audio samples into model parameters which describe each block of audio, and these parameters are stored instead of the audio samples, again with a header containing basic metadata. Common audio file formats such as WAV, which is usually associated with PCM data, provide a package allowing a large variety of audio representations. The MP3 format (formally called MPEG-2 Audio Layer III) uses lossy audio compression and is common for consumer audio storage; the use of MP3 files in MIR research has increased in recent years due to the emergence of large-scale datasets. Standard open source software libraries such as libsndfile[18] are available for reading and writing common non-proprietary formats, but some file formats are difficult to support with open source software due to the license required to implement an encoder.

For symbolic music data, a popular file format is MIDI (musical instrument digital interface), but this is limited in expressiveness and scope, as it was originally designed for keyboard instrument sequencing. For scores, a richer format such as MusicXML or MEI (Music Encoding Initiative) is required, which are XML-based representations including information such as note spelling and layout. For guitar "tabs" (a generic term covering tablature as well as chord symbols with or without lyrics), free text is still commonly used, with no standard format, although software has been developed which can parse the majority of such files [8]. Some tab web sites have developed their own formats using HTML or XML for markup of the text files. Other text formats such as the MuseData and Humdrum kern format [23] have been used extensively for musicological analysis of corpuses of scores.

For structured metadata, formats such as XML are commonly used, and in particular semantic web formats for linked data such as RDFa, RDF/XML, N3 and Turtle are employed. Since these are intended as machine-readable formats rather than for human consumption, the particular format chosen is less important than the underlying ontology which provides the semantics for the data. For image data, OMR systems typically process sheet music scanned at 300 dpi resolution, producing output in expMIDI (expressive MIDI), MusicXML or NIFF (Notation Interchange File Format) formats.

Finally, although music exists primarily in the auditory domain, there is a long tradition of representing music in various graphical formats. Common Western music notation is a primary example, but piano-roll notation, spectrograms and chromagrams also present musical information in potentially useful formats. Since music is a time-based phenomenon, it is common to plot the evolution of musical parameters as a function of time, such as tempo and dynamics curves, which have been used extensively in performance research [3]. Simultaneous representations of two or more temporal parameters have been achieved using animation, for example the Performance Worm [4], which shows the temporal evolution of tempo and loudness as a trajectory in a two-dimensional space. Other visualisations include similarity matrices for audio alignment and structural segmentation [16] and various representations for analysis of harmony and tonality [e.g. 22].

### 2.2.2 Specific Challenges

- **Investigate more musically meaningful features and representations.** There is still a significant semantic gap between the representations used in MIR and the concepts and language of musicians and audiences. In particular, many of the abstractions used in MIR do not make sense to a musically trained user, as they ignore or are unable to capture essential aspects of musical communication. The challenge of

[18] http://www.mega-nerd.com/libsndfile

designing musically meaningful representations must be overcome in order to build systems that provide a satisfactory user experience. This is particularly the case for automatically generated features, such as those utilising deep learning techniques, where the difficulty is creating features well-suited for MIR tasks which are still interpretable by humans.

- **Develop more flexible and general representations.** Many representations are limited in scope and thus constrained in their expressive possibilities. For example, most representations have been created specifically for describing Western tonal music. Although highly constrained representations might provide advantages in terms of simplicity and computational complexity, it means that new representations have to be developed for each new task, which inhibits rapid prototyping and testing of new ideas. Thus there is a need to create representations and abstractions which are sufficiently adaptable, flexible and general to cater for the full range of music styles and cultures, as well as for unforeseen musical tasks and situations.
- **Determine the most appropriate representation for each application.** For some use cases it is not beneficial to use the most general representation, as domain- or task-specific knowledge might aid the analysis and interpretation of data. However, there is no precise methodology for developing or choosing representations, and existing "best practice" covers only a small proportion of the breadth of musical styles, creative ideas and contexts for which representations might be required.
- **Unify formats and improve system interoperability.** The wealth of different standards and formats creates a difficulty for service providers who wish to create seamless systems with a high degree of interoperability with other systems and for researchers who want to experiment with software and data from disparate sources. By encouraging the use of open standards, common platforms, and formats that promote semantic as well as syntactic interoperability, system development will be simpler and more efficient.
- **Extend the scope of existing ontologies.** Existing ontologies cover only a small fraction of musical terms and concepts, so an important challenge is to extend these ontologies to describe all types of music-related information, covering diverse music cultures, communities and styles. These ontologies must also be linked to existing ontologies within and outside of the MIR community in order to gain maximum benefit from the data which is structured according to the ontologies.
- **Create compact representations that can be efficiently used for large-scale music analysis.** It is becoming increasingly important that representations facilitate processing of the vast amounts of music data that exist in current and future collections, for example, by supporting efficient indexing, search and retrieval of music data.
- **Develop and integrate representations for multimodal data.** In order to facilitate content-based retrieval and browsing applications, representations are required that enable comparison and combination of data from diverse modalities, including audio, video and gesture data.

## References

[1] J.C. Brown. Calculation of a constant Q spectral transform. *Journal of the Acoustical Society of America*, 89(1):425–434, 1991.

[2] I. Cross and E. Tolbert. Music and meaning. In S. Hallam, I. Cross, and M. Thaut, editors, *The Oxford Handbook of Music Psychology*. Oxford University Press, 2009.

[3] P. Desain and H. Honing. Tempo curves considered harmful: A critical review of the representation of timing in computer music. In *International Computer Music Conference*, pages 143–149, Montreal, Canada, 1991.

[4] S. Dixon, W. Goebl, and G. Widmer. The Performance Worm: Real time visualisation of expression based on Langner's tempo-loudness animation. In *International Computer Music Conference*, pages 361–364, Gothenburg, Sweden, 2002.

[5] G. Fazekas. Audio features ontology, 2010. http://www.omras2.org/AudioFeatures.

[6] E.J. Humphrey, J.P. Bello, and Y. LeCun. Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. In *13th International Society for Music Information Retrieval Conference*, pages 403–408, Porto, Portugal, 2012.

[7] H.G. Kim, N. Moreau, and T. Sikora. *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Wiley and Sons, 2005.

[8] R. Macrae and S. Dixon. Guitar tab mining, analysis and ranking. In *12th International Society for Music Information Retrieval Conference*, pages 453–458, Miami, Florida, USA, 2011.

[9] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, USA, 1999.

[10] A. Marsden. Schenkerian analysis by computer: A proof of concept. *Journal of New Music Research*, 39(3):269–289, 2010.

[11] D. McEnnis, C. McKay, I. Fujinaga, and P. Depalle. JAudio: A feature extraction library. In *6th International Conference on Music Information Retrieval*, London, UK, 2005.

[12] M.F. McKinney and J. Breebaart. Features for audio and music classification. In *4th International Conference on Music Information Retrieval*, Baltimore, Maryland, USA, 2003.

[13] R. Meddis and M.J. Hewitt. Modeling the identification of concurrent vowels with different fundamental frequencies. *Journal of the Acoustical Society of America*, 91(1):233–245, 1992.

[14] M. Minsky. Music, mind, and meaning. *Computer Music Journal*, 5(3), Fall 1981.

[15] D. Mitrovic, M. Zeppelzauer, and C. Breiteneder. Features for content-based audio retrieval. *Advances in Computers*, 78:71–150, 2010.

[16] M. Müller, D.P.W. Ellis, A. Klapuri, and G. Richard. Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1088–1110, October 2011.

[17] F. Pachet and P. Roy. Analytical features: A knowledge-based approach to audio feature generation. *EURASIP Journal on Audio, Speech, and Music Processing*, Article ID 153017, 2009.

[18] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the Cuidado project. Cuidado project report, IRCAM, http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf, 2004.

[19] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The music ontology. In *8th International Conference on Music Information Retrieval*, pages 417–422, Vienna, Austria, 2007.

[20] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A.R.S. Marcal, C. Guedes, and J.S. Cardoso. Optical music recognition: State-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1:173–190, 2012.

[21] J. Robinson. *Music and Meaning*. Cornell University Press, New York, USA, 1997.

[22] C.S. Sapp. Visual hierarchical key analysis. *ACM Computers in Entertainment*, 3(4), 2012.

[23] E. Selfridge-Field. *Beyond MIDI: The Handbook of Musical Codes*. MIT Press, 1997.

[24] X. Serra and J. Smith. Spectral modeling synthesis: A sound analysis/synthesis based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4):12–24, Winter 1990.

## 2.3 DATA PROCESSING METHODOLOGIES

Since its origins, the MIR community has used and adapted data processing methodologies from related research fields like speech processing, text information retrieval, and computer vision. A natural consequential challenge is to more *systematically* identify potentially relevant methodologies from data processing disciplines and stay up-to-date with their latest developments. This exchange of data processing methodologies reduces duplication of research efforts, and exploits synergies between disciplines which are, at a more abstract level, dealing with similar data processing problems. It will become even more relevant as MIR embraces the full multi-modality of music and its full complexity as a cultural phenomenon. This requires a regular involvement and commitment of researchers from diverse fields of science as well as an effort of communication across disciplines, and possibly even the formulation of joint research agendas. Such a more organised form of exchange of methodologies is likely to have a boosting impact on all participating disciplines due to the joining of forces and combined effort.

### 2.3.1 State of the art

The origins of Music Information Research were multi-disciplinary in nature. At the first edition of the ISMIR conference series, in 2000 [19], although the number of research papers was significantly smaller than in later editions, papers drew ideas from a relatively large number of disciplines: digital libraries, information retrieval, musicology and symbolic music analysis, speech processing, signal processing, perception and cognition, image processing (with applications to optical music recognition), and user modeling. This initial conference also debated intellectual property matters and systematic evaluations.

Since then, the ISMIR conference has grown tremendously, as illustrated by the number of unique authors that underwent a 400% increase between 2000 and 2011. In the last 12 years, neighboring fields of science with a longer history have influenced this growth of the MIR community. From the initial diversity of backgrounds and disciplines, not all had equal influence in the growth of MIR. Looking back on the first 12 years of ISMIR shows a clear predominance of bottom-up methodologies issued from data-intensive disciplines such as Speech Processing, Text Retrieval and Computer Vision, as opposed to knowledge-based disciplines such as Musicology or (Music) Psychology. One possible reason for the relatively stronger influence of data-intensive disciplines over knowledge-based ones is that the initial years of ISMIR co-occur with phenomena such as industrial applications of audio compression research and the explosive growth in the availability of data through the Internet (including audio files) [8]. Further, following typical tasks from Speech Processing, Computer Vision and Text Retrieval, MIR research rapidly focused on a relatively small set of preferential tasks such as local feature extraction, data modeling for comparison and classification, and efficient retrieval. In the following, we will review data processing methods employed in the three above-mentioned disciplines and relate their domains to the music domain to point out how MIR could benefit from further cross-fertilisation with these disciplines.

The discipline of Speech Processing aims at extracting information from speech signals. It has a long history and has been influential in a number of MIR developments, namely transcription, timbre characterisation, source recognition and source separation.

Musical audio representations have been influenced by research in speech transcription and speaker recognition. It is common-place to start any analysis of musical audio by the extraction of a set of local features, typical of speech transcription and speaker recognition, such as Mel Frequency Cepstrum Coefficients (MFCCs) computed on short-time Fourier transforms. In speech processing, these features make up the basic building blocks of machine learning algorithms that map patterns of features to individual speakers or likely sequences of words in multiple stages (i.e. short sequences of features mapped to phones, sequences of phones mapped to words and sequences of words mapped to sentences). A prevalent technique for mapping from one stage to the next are Hidden Markov Model (HMMs). Similar schemes have been adapted to music audio data and nowadays form the basis of music signal classification in genres, tags or particular instruments.

[19] http://ismir2000.ismir.net

Research in speech processing has also addressed the problem of separating out a single voice from a recording of many people speaking simultaneously (known as the "cocktail party" problem). A parallel problem when dealing with music data is isolating the components of a polyphonic music signal. Source separation is easier if there are at least as many sensors as sound sources [18]. But in MIR, a typical research problem is the under-determined source separation of many sound sources in a stereo or mono recording. The most basic instantiation of the problem assumes that N source signals are linearly mixed into M < N channels, where the task is to infer the signals and their mixture coefficients from the mixed signal. To solve it, the space of solutions has to be restricted by making further assumptions, leading to different methods: Independent Component Analysis (ICA) assumes the sources to be independent and non-Gaussian, Sparse Component Analysis (SCA) assumes the sources to be sparse, and Non-negative Matrix Factorisation (NMF) assumes the sources, coefficients and mixture to be nonnegative. Given that speech processing and content-based MIR both work in the audio domain, local features can be directly adopted – and in fact, MFCCs have been used in music similarity estimation from the very beginning of MIR [10]. HMMs have also been employed for modeling sequences of audio features or symbolic music [9]. Several attempts have been made to apply source separation techniques to music, utilising domain-specific assumptions on the extracted sources to improve performance: [28] assume signals to be harmonic, [27] assumes continuity in time, and [3] incorporates instrument timbre models.

Text Retrieval has also had a great influence on MIR, particularly the tasks of *document retrieval* (in a given collection, find documents relevant to a textual query in the form of search terms or an example document) and *document classification* (assign a given document to at least one of a given set of classes, e.g., detect the topic of a news article or filter spam emails). Both problems require some abstract model for a document. The first system for document classification [17] represented each document as a word count vector over a manually assembled vocabulary of "clue words", then applied a Na'ive Bayes classifier to derive the document's topic, neither regarding the order nor co-occurrence of words within the document. Today, documents are still commonly represented as a word count vector – or Bag of Words (BoW) – for both classification and retrieval, but improvements over [17] have been proposed on several levels, namely stemming, term weighting [22], topic modeling [7], semantic hashing [12], word sense disambiguation [19], and N-gram models. Some of these techniques have been applied to find useful abstract representations of music pieces as well, but their use implies that a suitable equivalent to words can be defined for music. Some authors tried to apply vector quantisation ("stemming") to frame-wise audio features ("words") to form a BoW model for similarity search [24]. [21] additionally employ TF/IDF term weighting of their so-called "audio-words". [13] successfully applied HDP topic models for similarity estimation, albeit modeling topics as Gaussian distributions of MFCCs rather than multinomials over discrete words.

Finally, three typical Computer Vision problems have been particularly influential in MIR research, namely *scene recognition* (classifying images of scenery), *multiple object detection* (decomposing a complex image into a set of known entities and their locations) and *image retrieval by example*. Again, in Computer Vision, all these tasks require abstract representations of images or image parts to work with, and researchers have developed a wealth of image-specific local features and global descriptors (see [6], pp.17-24 for a review). A common framework has been inspired by Text Retrieval: [29] regard images as documents composed of "keyblocks", in analogy to text composed of keywords. Keyblocks are vector-quantised image patches extracted on a regular grid, forming a 2-dimensional array of "visual words", which can be turned into a Bag of Visual Words (BoVW) by building histograms. Several improvements have since been proposed, namely regarding visual words [26], Pooling [2], Spatial pyramids [15], Topic modeling [25], Generative image models [14], Learning invariances [11], Semantic hashing [14]. As for Speech and Text processing, some of these techniques have been adopted for the processing of music audio features. Examples include [1] who employs sparse coding of short spectrogram excerpts of harpsichord music, yielding note detectors. [4] use Haar-like feature extractors inspired from object detection to discriminate speech from music. [20] apply horizontal and vertical edge detectors to identify harmonic and percussive elements. [16] apply Convolutional RBMs for local feature extraction with some success in genre classification. [23] learn local image features for music similarity estimation. Additionally, as music pieces can be represented directly as images by using e.g. images of spectrograms, several authors directly applied image processing techniques to music: [5] extract features for genre classification with oriented difference of Gaussian filters. Recent improvements on using image features for music classification can be found in [5].

### 2.3.2 Specific Challenges

- **Systematise cross-disciplinary transfer of methodologies.** Early breakthroughs in MIR came from a relatively limited number of external fields, mainly through the contributions of individual researchers working in neighboring fields -e.g. Speech Processing- and applying their methodologies to music. Being more systematic about this implies two challenges for the MIR community: first, to stay up-to-date with latest developments in disciplines that were influential in some points of MIR evolution, and second to define ways to *systematically* identify potentially relevant methodologies from neighboring disciplines.
- **Take advantage of the multiple modalities of music data.** Music exists in many diverse modalities (audio, text, video, score, etc.) which in turn call for different processing methodologies. Given a particular modality of interest -e.g. audio-, in addition to identifying promising processing methodologies from neighboring fields dealing with the same modality -e.g. speech processing-, an effort will have to be made to apply methodologies *across* modalities. Further, as music exists simultaneously in diverse modalities, another challenge for MIR will be to include methodologies from cross-modal processing, i.e. using *joint* representations/models for data that exists, and can be represented, simultaneously in diverse modalities.
- **Adopt recent Machine Learning techniques.** As exemplified above, MIR makes a great use of machine learning methodologies, in particular many tasks are formulated according to a batch learning approach where a fixed amount of annotated training data is used to learn models which can then be evaluated with similar data. However, music data can now be found in very large amounts (e.g. in the scale of hundreds of thousands of items for music pieces in diverse modalities, or in the scale of tens of millions in the case of e.g. tags), music is increasingly existing in data streams rather than in data *sets*, and the characterisation of music data can evolve with time (e.g. tag annotations are constantly evolving, sometimes even in an adverse way). These data characteristics (i.e. very large amounts, streaming, non-stationarity) -Big Data characteristics- imply a number of challenges for MIR, such as data acquisition, dealing with weakly structured data formats, scalability, online (and real-time) learning, semi-supervised learning, iterative learning and model updates, learning from sparse data, learning with only positive examples, learning with uncertainty, etc. (see e.g. Yahoo! Labs "key scientific challenges" in Machine Learning [20] and the White Paper "Challenges and Opportunities with Big Data" published by the Computing Community Consortium [21]).

## References

[1] Samer A. Abdallah. *Towards Music Perception by Redundancy Reduction and Unsupervised Learning in Probabilistic Models*. PhD thesis, King's College London, London, UK, 2002.

[2] Y-Lan Boureau, Jean Ponce, and Yann LeCun. A theoretical analysis of feature pooling in visual recognition. In *Proc. of ICML (International Conference on Machine Learning)*, pages 111–118, 2010.

[3] Juan José Burred. *From Sparse Models to Timbre Learning: New Methods for Musical Source Separation*. PhD thesis, Technical University of Berlin, Berlin, Germany, September 2008.

[4] Norman Casagrande, Douglas Eck, and Balázs Kégl. Frame-level speech/music discrimination using AdaBoost. In *Proc. ISMIR (International Symposium on Music Information Retrieval)*, 2005.

[5] Y. Costa, L. Oliveira, A. Koerich, F. Gouyon, and J. Martins. Music genre classification using LBP textural features. *Signal Processing*, 92(11):2723–2737, 2012.

[6] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008.

[7] Scott Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6):391–407, 1990.

[20] http://labs.yahoo.com/ksc/Machine_Learning
[21] http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf

[8] J. Stephen Downie, Donald Byrd, and Tim Crawford. Ten years of ISMIR: Reflections on challenges and opportunities. In *Proc. ISMIR (International Society for Music Information Retrieval Conference)*, pages 13–18, 2009.

[9] Arthur Flexer, Elias Pampalk, and Gerhard Widmer. Hidden markov models for spectral similarity of songs. In *Proc. DAFx (Int. Conference on Digital Audio Effects)*, 2005.

[10] J. Foote. Content-based retrieval of music and audio. In *Proc. Multimedia Storage and Archiving Systems*, pages 138–147, 1997.

[11] Geoffrey Hinton, Alex Krizhevsky, and Sida Wang. Transforming auto-encoders. In *Proc. ICANN (International Conference on Artificial Neural Networks)*, pages 44–51, 2011.

[12] Geoffrey Hinton and Ruslan Salakhutdinov. Discovering binary codes for documents by learning deep generative models. *Topics in Cognitive Science*, pages 1–18, 2010.

[13] Matthew D. Hoffman, David M. Blei, and Perry R. Cook. Content-based musical similarity computation using the hierarchical Dirichlet process. In *Proc. ISMIR (International Society for Music Information Retrieval Conference)*, pages 349–354, 2008.

[14] Alex Krizhevsky and Geoffrey Hinton. Using very deep autoencoders for content-based image retrieval. In *Proc. European Symposium on Artificial Neural Networks (ESANN)*, 2011.

[15] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. CVPR (IEEE Computer Society Conference on Computer Vision and Pattern Recognition)*, pages 2169–2178, 2006.

[16] Honglak Lee, Peter T. Pham, Yan Largman, and Andrew Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Proc. NIPS (Advances in Neural Information Processing Systems)*, pages 1096–1104, 2009.

[17] M.E. Maron. Automatic indexing: An experimental inquiry. *Journal of the Association for Computing Machinery*, 8(3):404–417, 1961.

[18] Nikolaos Mitianoudis. *Audio Source Separation using Independent Component Analysis*. PhD thesis, Queen Mary University of London, April 2004.

[19] Roberto Navigli. Word sense disambiguation: A survey. *ACM Computing Surveys*, 41(2):10:1–10:69, 2009.

[20] Tim Pohle, Peter Knees, Klaus Seyerlehner, and Gerhard Widmer. A high-level audio feature for music retrieval and sorting. In *Proc. DAFx (International Conference on Digital Audio Effects)*, 2010.

[21] Matthew Riley, Eric Heinen, and Joydeep Ghosh. A text retrieval approach to content-based audio hashing. In *Proc. ISMIR (International Society for Music Information Retrieval Conference)*, pages 295–300, 2008.

[22] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513–523, 1988.

[23] Jan Schluter and Christian Osendorfer. Music similarity estimation with the mean-covariance restricted boltzmann machine. In *Proc. ICMLA (International Conference on Machine Learning and Applications)*, pages 118–123, 2011.

[24] Klaus Seyerlehner, Gerhard Widmer, and Peter Knees. Frame-level audio similarity: A codebook approach. In *Proc. DAFx (International Conference on Digital Audio Effects)*, 2008.

[25] Josef Sivic, Bryan C. Russell, Alexei A. Efros, Andrew Zisserman, and William T. Freeman. Discovering objects and their localization in images. In *Proc. ICCV (International Conference on Computer Vision)*, pages 370–377, 2005.

[26] Jan van Gemert, Jan-Mark Geusebroek, Cor J. Veenman, and Arnold W. M. Smeulders. Kernel codebooks for scene categorization. In *Proc. ECCV (European Conference on Computer Vision)*, pages 696–709, 2008.

[27] Tuomas Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3):1066–1074, 2007.

[28] Tuomas Virtanen and Anssi Klapuri. Separation of harmonic sounds using linear models for the overtone series. In *Proc. ICASSP (International Conference on Acoustics, Speech, and Signal Processing)*, pages 1757–1760, 2002.

[29] Lei Zhu, Aibing Rao, and Aidong Zhang. Theory of keyblock-based image retrieval. *ACM Trans. Inf. Syst.*, 20(2):224–257, 2002.

## 2.4 KNOWLEDGE-DRIVEN METHODOLOGIES

For a long time, the MIR community has been focusing on a range of bottom-up approaches, addressing the kinds of data we use and the types of algorithms we apply to it. A major challenge is to complement this focus and explore other methodologies and fields of science which approach music in a more integrated way. After all, music information research is just one of many sciences that centre on and care about music, which include musicology, psychology, sociology and neuroscience. Over decades of research, each of these fields has aggregated knowledge concerning music which can inform the process of music information research. The focus here is on gaining domain knowledge from outside of MIR as opposed to borrowing methodologies or algorithms. This will require that researchers from different disciplines engage in a dialogue on all aspects of music. The potential impact is that all participating disciplines benefit from the diverse and differing views on the phenomenon of music, in all its aspects and forms.

### 2.4.1 State of the art

In what follows, we briefly review the already existing and potential relations between MIR and musicology, psychology, sociology and neuroscience, which we identified as particularly relevant for our field of research.

#### Musicology

Musicology is fundamental to MIR, and building bridges between disciplines is at the very core of research in musicology [4]. Musicologists have taken an active role in the ISMIR community, for instance, musicology has always been considered a key topic in the ISMIR call for papers (see, e.g. research areas related to computational musicology, computational ethnomusicology explicitly considered at ISMIR 2012 [22]). Moreover, the conference on Interdisciplinary Musicology (CIM [23]) has included papers on computational modeling in the program, and there is a special edition of this conference on the topic of "Technology" that is planned for 2014 [24]. There are also some relevant journals in this intersection (e.g. Journal of Mathematics and Music [25]) and the Special Issue on Computational Ethnomusicology in the Journal of New Music Research [26]. An overview on the relationship between MIR and musicology is provided in the Musicology tutorial presented at ISMIR 2011 by Volk & Wiering [27], and a guide to the use of MIR technology in musicology is given in [11].

Although musicological studies in MIR have traditionally focused on the symbolic domain, recent developments in music transcription and feature extraction technologies from audio signals have opened new research paths at the intersection of musicology and signal processing. Key research topics in this area have been, among others, melodic similarity, key estimation and chord tracking. Musicological and MIR research have been contrasted [28] in terms of, among others, data sources, repertoires and methodologies, and some opportunities for future research have been pointed out. MIR technologies can contribute with tools and data that are useful for musicological purposes, and Musicology can provide relevant research problems and use cases that can be addressed through MIR technologies. A mutual influence is starting to take place, although there is still a need for more collaboration between musicologists and technicians to create a truly interdisciplinary research area and contribute with truly music-rooted models and technologies. Only by this collaboration can we address the current gap between feature extractors and expert analyses and make significant contributions to existing application needs, e.g. version identification, plagiarism detection, music recommendation, and to study how the relationship between people and music changes with the use of technology (e.g. "Musicology for the Masses" project [29]).

[22] http://ismir2012.ismir.net/authors/call-for-participation
[23] http://www.uni-graz.at/~parncutt/cim
[24] http://www.sim.spk-berlin.de/cim14_919.html
[25] http://www.tandfonline.com/toc/tmam20/current
[26] http://www.tandfonline.com/toc/nnmr20/current
[27] http://ismir2011.ismir.net/tutorials/ISMIR2011-Tutorial-Musicology.pdf
[28] http://ismir2011.ismir.net/tutorials/ISMIR2011-Tutorial-Musicology.pdf
[29] http://www.elec.qmul.ac.uk/digitalmusic/m4m

### Psychology of Music

Music is created and experienced by humans, and the ultimate goal of MIR is to produce results that are helpful and interesting for humans. Therefore it is only natural to care about how humans perceive and create music. Music psychology tries to explain both musical behavior and musical experience with psychological methods. Its main instrument therefore is careful experimentation involving human subjects engaged in some kind of musical activity. Research areas span the whole spectrum from perception to musical interaction in large groups. Research questions concern the perception of sound or sound patterns, as well as perception of more musically meaningful concepts like harmony, pitch, rhythm, melody and tonality. The emotions associated with personal music experience are a part of music psychology, as are personal musical preferences and how they are influenced through peer groups and family, and musical behaviors from dancing to instrument playing to the most sophisticated interaction within whole orchestras.

Therefore music psychology should be able to provide valuable knowledge for MIR researchers in a whole range of sub-fields. Indeed there already is a certain exchange of knowledge between music psychology and MIR. Just to give a few examples, Carol L. Krumhansl, an eminent figure in music psychology, was an invited speaker at the Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010), in Utrecht, Netherlands [30] talking about "Music and Cognition: Links at Many Levels". Her monograph on "cognitive foundations of musical pitch" [10] is still seen as one of the standard texts on the subject. Gerhard Widmer, who has been an important contributor to MIR early on, was a keynote speaker at the "12th International Conference on Music Perception and Cognition (ICMPC)" [31], which is one of the most important conferences in the field of music psychology. At last year's major conference in the MIR field (ISMIR 2012 [32]) there was a joint presentation of an MIR researcher and a psychologist elaborating on the sometimes complicated dialog of the two disciplines [1, 2].

### Sociology of Music

Social psychology and the sociology of music focus on individuals as members of groups and on how groups and shared cultural codes influence music-related attitudes and activities. This point of view allows one to ask and answer important questions like: How do individuals and groups use music? How is the collective production of music made possible? How does music relate to broader social distinctions, especially class, race, and gender?

Although it is evident that such a sociology of music should be able to provide important insights not only for the field of MIR, many authors have suggested that research over recent decades has largely ignored the social functions of music at the expense of its cognitive and emotional functions (see e.g. [8]). [7] concluded that music serves three social functions: it is used by individuals to help manage their moods, self-identity [5], and interpersonal relationships. [14] elaborated this idea, showing that a sample of 13- to 14-year-olds listened to music to portray a social image to others, and to fulfill their emotional needs. Similarly, [16] showed that American and English adolescents listened to music to satisfy both emotional and social needs, as well as for reasons of self-actualisation. [12] remarked that listening to music was "a social activity", which offered an opportunity for participants "to socialise with friends" (e.g., dancing, sharing live music). Music has a stronger social component for teenagers and young people than for seniors but it still keeps some powers to strengthen social bonds and to provide memory aids when brain functions decline. In this respect, life-span and elderly-centred applications are yet to be fully explored and exploited [13]. How MIR can benefit from these and other results concerning the sociology of music is still a largely open question which opens up new and promising areas of research.

### Neuroscience

All music psychological questions raised above could of course also be examined with neuroscientific methods. Instead of measuring the subject's behavior in music psychological experiments or directly asking subjects about their experiences concerning music it is possible to measure various signals from the human brain during such experiments. Possible signals range from electro-encephalography (EEG) to magneto-encephalography (MEG)

[30] http://ismir2010.ismir.net
[31] http://icmpc-escom2012.web.auth.gr
[32] http://ismir2012.ismir.net

or functional magnetic resonance imaging (fMRI). Each of the signals has its own characteristic strengths and weaknesses. E.g. EEG has a very good temporal but poor spatial resolution where fMRI is just the opposite. No matter what brain signals are being used, the fundamental question is always what parts of the brain contribute in what way to a subject's experience or creation of music. It is not immediately clear what MIR could gain from such a knowledge about brain structures involved in perception and production of music that could go beyond knowledge obtained from psychological experiments not utilising neuroscientific methods. The biggest contribution might concern problems where humans have difficulty self-assessing their performance and experience. One example is the experience of emotions when listening to music. Neuroscientific methods might be able to provide a more quantitative and maybe more accurate picture than human self-assessment (see e.g. [3], [15]). Differences in brain structure and function between skilled musicians and non-musicians is another well researched subject (see e.g. [6], [9]). The same holds for the study of the neuronal processes during performance of music where the sensorimotor interplay is at the center of interest (see [17] for a recent review).

### 2.4.2 Specific Challenges

- **Integrate insights from disciplines relevant to MIR and make them useful for our research.** This requires mutual understanding and exchange of results and researchers. The challenge is to integrate research agendas through the formulation of common interests and goals as well as a common vocabulary and dedicated communication paths. This will be important for both MIR and all other disciplines caring about music since there is a mutual benefit to be gained from this.
- **Develop richer musical models incorporating musicological knowledge.** MIR has been focusing on a limited number of musical concepts, which are modelled at a shallower depth than they are treated by musicologists. Enriching these concepts will help bridge the gap between low-level MIR representations and higher-level semantic concepts.
- **Extend and strengthen existing links to music psychology.** An example for a joint interest is the clearer formulation and understanding of the notion of "music similarity" with the help of music psychological results and proper experimentation. This requires that music psychologists be informed about MIR models and methods to compute music similarity and that MIR researchers are being educated about how music psychologists access subjective notions and cognitive aspects of music similarity in humans. Expected outcomes are improved models and algorithms to compute music similarity as well as computer aided selection of research stimuli for the psychological experiments.
- **Give due attention to the social function of music in our research.** This makes it necessary that MIR cares about groups of individuals and their interaction instead of about disconnected individuals. Taste formation, preference and music selection are a combined function of personal and group variables, and we currently do not know how to weight both aspects to achieve good predictive models. Research and technologies that help to understand, modify, increase or make possible group cohesion, improvements on self-image, or strengthen collective bonds could have a strong impact, especially on disfavoured, problem-prone and marginal groups. The final challenge here would be to be able to shift the increasing trend of enjoying music as an individual, isolated, activity, making social ways to search, share, listen to, and re-create the otherwise "personal" collections of music possible.
- **Learn, understand and eventually integrate neuro-scientific results concerning music.** The question of how music influences emotions of listeners is a good example which is of great interest to MIR and where a growing body of neuro-scientific results on the basics of emotional experience exists. Comprehension of these results could enable better and richer MIR models of emotion in music. On the other hand, education of neuroscience researchers in MIR technology might help design of brain studies on music (e.g. in producing generative musical research stimuli).

## References


[1] Jean-Julien Aucouturier and Emmanuel Bigand. Mel Cepstrum and Ann Ova: The Difficult Dialog Between MIR and Music Cognition. In Fabien Gouyon, Perfecto Herrera, Luis Gustavo Martins, and Meinard Müller, editors, *ISMIR*, pages 397–402. FEUP Edições, 2012.

[2] Jean-Julien Aucouturier and Emmanuel Bigand. Seven problems that keep MIR from attracting the interest of the natural sciences. *Journal of Intelligent Information Systems*, in print, 2013.

[3] A. J. Blood and R. J. Zatorre. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci U S A*, 98(20):11818–11823, September 2001.

[4] Ewa Dahlig-Turek, Sebstain Klotz, Richard Parncutt, and Frans Wiering, editors. *Musicology (Re-) Mapped*. Standing Committee for the Humanities. European Science Foundation, 2012.

[5] Tia DeNora. Music as a technology of the self. In *Poetics*, volume 27, pages 31–56, Niederlande, 1999. Elsevier.

[6] Christian Gaser and Gottfried Schlaug. Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, 23(27):9240–5, October 2003.

[7] D. J. Hargreaves and A. C. North. The functions of music in everyday life: Redefining the social in music psychology. *Psychology of Music*, 27(1):71–83, 1999.

[8] David J. Hargreaves and Adrian C. North. *The social psychology of music / edited by David J. Hargreaves and Adrian C. North*. Oxford University Press, 1997.

[9] T. Krings, R. Topper, H. Foltys, S. Erberich, R. Sparing, K. Willmes, and A. Thron. Cortical activation patterns during complex motor tasks in piano players and control subjects. A functional magnetic resonance imaging study. *Neuroscience Letters*, 278(0304-3940 SB - IM):189–193, January 2000.

[10] Carol L Krumhansl. *Cognitive foundations of musical pitch*. Oxford University Press, New York, 1990.

[11] Daniel Leech-Wilkinson. *The Changing Sound of Music: Approaches to Studying Recorded Musical Performance*. CHARM: London, http://www.charm.kcl.ac.uk/studies/chapters/intro.html (Accessed 11/04/2012), 2009.

[12] Adam J. Lonsdale and Adrian C. North. Why do we listen to music? A uses and gratifications analysis. *British journal of psychology (London, England : 1953)*, 102(1):108–134, February 2011.

[13] Wendy L. Magee, Michael Bertolami, Lorrie Kubicek, Marcia LaJoie, Lisa Martino, Adam Sankowski, Jennifer Townsend, Annette M. Whitehead-Pleaux, and Julie Buras Zigo. Using music technology in music therapy with populations across the life span in medical and educational programs. *Music and Medicine*, 3(3):146–153, 2011.

[14] A. C. North, D. J. Hargreaves, and S. A. O'Neill. The importance of music to adolescents. *British Journal of Educational Psychology*, 70 (Pt 2)(2):255–272, 2000.

[15] Louis A. Schmidt and Laurel J. Trainor. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion*, 15(4):487–500, 2001.

[16] Mark Tarrant, Adrian C. North, and David J. Hargreaves. English and American Adolescents' Reasons for Listening to Music. *Psychology of Music*, 28(2):166–173, October 2000.

[17] Robert J Zatorre, Joyce L Chen, and Virginia B Penhune. When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7):547–558, 2007.

## 2.5 ESTIMATION OF ELEMENTS RELATED TO MUSICAL CONCEPTS

By musical concept extraction we refer to the estimation of the elements of a notation system from the audio signal and the estimation of higher-level semantic information from these elements. These elements belong to a vocabulary and are assembled according to a grammar specific to a culture. The challenge here is to automatically derive musical concepts from audio signals or from commonly available symbolic data, such as MIDI or scores. Extracting musical concepts from audio signals is technically a very difficult task and new ways to perform this still need to be found. More specifically, we need to develop better source separation algorithms; develop methodologies for joint estimation of music content parameters; and use symbolic information plus audio data to extract higher level semantic concepts. For this reason, this challenge does not only involve researchers (in signal processing, machine learning, cognition and perception, and musicology), but also content providers (record companies), who could help by delivering material for research (such as the separate audio tracks of multi-track recordings). Enabling the description of data in terms of musical concepts can help improve the understanding of the content and hence develop better use or new uses of this content. Having access to separate source elements or to accurate score information may have a huge impact on the creative industries (game industry, music e-learning . . . ) and the music distribution industry (better access to music), as well as facilitating large-scale musicological analyses.

### 2.5.1 State of the art

In MIR, Audio Content Analysis (ACA) aims at extracting musical concepts using algorithms applied to the audio signal. One of its goals is to estimate the score of a music track (melody, harmony, rhythm, beat and downbeat positions, overall structure) from the audio signal. ACA has been a major focus of research in the MIR community over the past decade. But how can algorithm performance be further improved in this field?

Most ACA algorithms aim at estimating two kinds of concepts: subjective or application-oriented concepts (such as genre, mood, user tags and similarity), and musical concepts (such as pitch, beat, meter, chords and structure). Whatever the task, ACA algorithms first need to extract meaningful information from the signal and then map it to the concept. ACA therefore involves research related to signal processing (extracting better audio features or creating better signal models such as the sinusoidal model performing better source separation), and to knowledge encoding and discovery (how to encode/acquire knowledge in/with an algorithm) including therefore machine-learning (SVM, AdaBoost, RandomForest). Considering that subjective concepts are hard to define, their estimation is usually performed using examples, hence using machine-learning to acquire the knowledge from these examples. Musical concepts can be defined explicitly or by example, hence ACA algorithms either acquire the knowledge through predefined models (such as a musical grammar to define chord transition probabilities) or trained. A last trend concerns the estimation of subjective concepts using estimated musical concepts as information (for example inferring the genre from the estimated tempo and/or chord progression).

"Musical concepts" denotes the parameters related to written music. Since the MIR community is largely made up of Western researchers, written music refers to the music notation system originating from European classical music, consisting of notes with an associated position and duration inside a bar, in the context of a meter (hence providing beat positions inside the bars), a clef (indicating the octaves assigned to the notes), a key signature (series of sharps or flats) organised by instrument (or hands) into parallel staffs, and finally organised in a large structure corresponding to musical sections. An extension of common notation summarises groups of simultaneously occurring notes using chord symbols. ACA aims at retrieving this music notation from the observation of an audio music track (realisation of a generative music process). Since the audio signal represents a realisation of the music notation it exhibits variations in terms of interpretation (not all the notes are played, pitches vary over time, and musicians modify timing). ACA algorithms estimate pitches with associated starting and ending times which are then mapped to the [pitch-height, clef, key, metrical position and duration] system. All this makes music transcription a difficult problem to solve. Moreover, until recently, from an application point-of-view, the market place was considered limited (to users with musical training). Today, with the success of applications such as Melodyne (multi-pitch estimation), Garage-Band, the need for search using Query-by-

Humming (dominant melody extraction), mobile applications such as Tonara (iPad) and online applications such as Songs2See [33], information related to music transcription is now reaching everyday people. For the estimation of music transcription two major trends can be distinguished.

### Non-informed estimation (estimation-from-scratch)

These approaches attempt to estimate the various music score concepts from scratch (without any information such as score or chord-tabs). In this category, approaches have been proposed for estimating the various pitches, the key, the sequence of chords, the beat and downbeat positions and the global structure.

Multi-pitch estimation is probably the most challenging task since it involves being able to identify the various pitches occurring simultaneously and estimating the number of sources playing at any time. According to [37], most multi-pitch algorithms follow three main principles closely related to mechanisms of the auditory system: harmonicity, spectral smoothness, and synchronous amplitude evolution within a given source. From these principles a number of approaches are derived: solving the problem using a global optimisation scheme such as NMF [35], harmonic temporal structured clustering [13], iterative optimisation [14] or a probabilistic framework [31]. Considering the fact that the performance obtained in the past years in the related MIREX task ( 69% note-accuracy for simple music materials) remains almost constant, it seems that a glass ceiling has been reached in this domain and that new approaches should be studied. A sub-problem of multi-pitch estimation can be found in the simpler "melody extraction" problem (which is also related to the "lyric recognition/ alignment" described below). The principles underlying melody extraction methods [32] are similar, but only one pitch needs to be estimated, which is usually the most pre-dominant hence easily detected in the signal. Because of this, much better performance has been achieved for this task (up to 85% in MIREX-2011).

Key and chord estimation are two closely related topics. They both aim at assigning a label chosen from a dictionary (a fixed set of 24 tonalities, or the various triads with possible extensions) to a segment of time. Given that the estimation of key and chords from estimated multi-pitch data is still unreliable algorithms rely for the most part on the extraction of Chroma or Harmonic Pitch Class Profiles [8] possibly including harmonic/pitch-enhancement or spectrum whitening. Then, a model (either resulting from perceptual experiments, trained using data or inspired by music theory) is used to map the observations to the labels. In this domain, the modeling of dependencies (with HMMs or Bayesian networks) between the various musical parameters is a common practice: dependencies between chords and key [26] between successive chords, between chord, metrical position and bass-note [20], or between chord and downbeat [23]. Key and chord estimation is the research topic that relies the most on music theory.

While music scores define the temporal grid at multiple metrical levels, most research focuses on the beat level (named tactus). In this field, methods can be roughly subdivided into a) audio-to-symbolic or onset-based methods and b) energy-variation-based methods [33]. The periodicities can then be used to infer the tempo directly or to infer the whole metrical structure (tatum, tactus, measure, systematic time deviations such as swing factor [16]) through probabilistic or multi-agent models. Other sorts of front-ends have also been used to provide higher-level context information (chroma-variation, spectral balance [15] [28]). Given the importance of correct estimation of the musical time-grid provided by beat and downbeat information, this field will remain active for some time. A good overview can be found in [9].

Research on the estimation of Music Structure from audio started at the end of the '90s with the work of Foote [6] (co-occurrence matrix) and Logan [18]. By "structure" the various works mean detection of homogeneous parts (state approach [27]) or repetitions of sequences of events, possibly including transpositions or time-stretching (sequence approach [27]). Both methods share the use of low-level features such as MFCC or Chroma/PCP as front-end. In the first case, methods are usually based on time-segmentation and various clustering or HMM techniques [17]. Sequence approaches usually first detect repetitions in a self-similarity matrix and then infer the structure from the detected repetitions using heuristics or fitness approaches [24]. A good overview of this topic can be found in [25].

[33] http://www.songs2see.com

### Informed estimation (alignment and followers)

These approaches use previous information (such as given by a score, a MIDI file or a text-transcription) and align it to an audio file hence providing inherently its estimation. This method is currently applied to two fields for which estimation-from-scratch remains very complex: scores and lyrics.

Score alignment and score following are two closely related topics in the sense that the latter is the real-time version of the first. They both consist in finding a time-synchronisation between a symbolic representation and an audio signal. Historically, score following was developed first with the goal of allowing interactions between a computer and a musician ([5], [34]) using MIDI or fingering information and not audio because of CPU limitations. This work was later extended by Puckette [29] to take into account pitch estimation from audio and deal with polyphonic data. Given the imperfect nature of observations, [10] introduced statistical approaches. Since 1999, Hidden Markov Model/ Viterbi seems to have been chosen as the main model to represent time dependency [30]. The choice of Viterbi decoding, which is also used in dynamic time warping (DTW) algorithms, is the common point between Alignment and Followers [22]. Since then, the focuses of the two fields have been different. Alignment focuses on solving computational issues related to DTW and Follower on anticipation (using tempo or recurrence information [3]). While formerly being the privilege of a limited number of people, today score following is now accessible to a large audience through recent applications such as Tonara (iPad) or Songs2See (web-based).

Automatic transcription of the lyrics of a music track is another complex task. It involves first locating the signal of the singer in the mixed audio track, and then recognising the lyrics conveyed by this signal (large differences between the characteristics of the singing voice and speech make standard speech transcription systems unsuitable for the singing voice). Work on alignment started with the isolated singing voice [19] and was later extended to the singing voice mixed with other sources. Usually systems first attempt to isolate the singing voice (e.g. using the PreFest dominant melody detection algorithm [7], then estimate a Voice Activity Criterion and then decode the phoneme sequence using a modified HMM topology (filler model in [7]), adapting the speech phoneme model to singing. Other systems also exploit the temporal relationships between the text of the lyrics and the music. For example, the system Lyrically [36] uses the specific assumption that lyrics are organised in paragraphs as the music is organised in segments. The central segment, the chorus, serves as an anchor-point. Measure positions are used as the anchor-point for lines.

### Deriving musical information from symbolic representations

Research related to the extraction of higher-level music elements from symbolic representations has always been at the heart of MIR, with research centred around systems such as Humdrum [12], MuseData, Guido/MIR [11], jSymbolic [21], Music21 [4] or projects such as Wedel Music [1] and MUSART [2]. Most of the higher-level elements are the same as those targeted by audio description (for example, the symbolic tasks run at MIREX: genre, artist, similarity, cover song, chord, key, melody identification, meter estimation), but some, due to current limitations of audio processing, are still specific to symbolic processing (e.g. recognition of motives and cadences).

## 2.5.2 Specific Challenges

- **Separate the various sources of an audio signal.** The separation of the various sources of an audio track (source separation) facilitates its conversion to a symbolic representation (including the score and the instrument names). Conversely, the prior knowledge of this symbolic information (score and/or instruments) facilitates the separation of the sources. Despite the efforts made over the last decades, efficient source separation and multi-pitch estimation algorithms are still lacking. Alternative strategies should therefore be exploited in order to achieve both tasks, such as collaborative estimation.
- **Jointly estimate the musical concepts.** In a music piece, many of the different parameters are inter-dependent (notes often start on beat or tatum positions, pitch most likely belongs to the local key). Holistic/joint estimation should be considered to improve the performance of algorithms and the associated computational issues should be solved.

- **Develop style-specific musical representations and estimation algorithms**. Depending on the music style, different types of representation may be used (e.g. full score for classical music and lead sheets for jazz). Based on previous knowledge of the music style, a priori information may be used to help the estimation of the relevant musical concepts.
- **Consider non-Western notation systems.** Currently, most analyses are performed from the point of view of Western symbolic notation. Dependence of our algorithms on this system should be made explicit. Other notation systems, other informative and user-adapted music representations, possibly belonging to other music cultures, should be considered, and taken into account by our algorithms.
- **Compute values for the reliability of musical concept estimation.** Many musical concepts (such as multi-pitch or tempo) are obtained through "estimation" (as opposed to MFCC which is a cascade of mathematical operators). Therefore the values obtained by these estimations may be wrong. The challenge is to enable algorithms to compute a measure of the reliability of their estimation ("how much the algorithm is sure about its estimation"). From a research point of view, this will allow the use of this "uncertainty" estimation in a higher-level system. From an exploitation point of view, this will allow the use of these estimations for automatically tagging music without human intervention.
- **Take into account reliability in systems.** Estimations of musical concepts (such as multi-pitch or beat) can be used to derive higher-level musical analysis. We should study how the uncertainty of the estimation of these musical concepts can be taken into account in higher-level algorithms.
- **Develop user-assisted systems.** If it is not possible to estimate the musical concepts fully automatically, then a challenge is to study how this can been done interactively with the user (using relevance feedback).

## References

[1] Jérôme Barthélemy and Alain Bonardi. Figured bass and tonality recognition. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 129–136, Bloomington, Indiana, USA, 2001.

[2] William P. Birmingham, Roger B. Dannenberg, Gregory H. Wakefield, Mark Bartsch, David Bykowski, Dominic Mazzoni, Colin Meek, Maureen Mellody, and William Rand. MUSART: Music Retrieval via Aural Queries. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 73–81, Bloomington, Indiana, USA, 2001.

[3] Arshia Cont. ANTESCOFO: Anticipatory synchronization and control of interactive parameters in computer music. international computer music conference. In *Proc. of ICMC (International Computer Music Conference)*, Belfast, Ireland, 2008.

[4] Michael Scott Cuthbert, Christopher Ariza, and Lisa Friedland. Feature extraction and machine learning on symbolic music using the music21 toolkit. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 387–392, Miami, Florida, USA, 2011.

[5] Roger B. Dannenberg. An On-Line Algorithm For Real-Time Accompaniment. In *Proc. of ICMC (International Computer Music Conference)*, pages 193–198, Paris, France, 1984. Computer Music Association.

[6] Jonathan Foote. Visualizing music and audio using self-similarity. In *Proc. of ACM Multimedia*, pages 77–80, Orlando, Florida, USA, 1999.

[7] Hiromasa Fujihara, Masataka Goto, Jun Ogata, and Hiroshi G. Okuno. LyricSynchronizer: Automatic Synchronization System between Musical Audio Signals and Lyrics. *Selected Topics in Signal Processing, IEEE Journal of*, 5(6):1252–1261, 2011.

[8] Emilia Gomez. Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing, Special Cluster on Computation in Music*, 18(3), 2006.

[9] Fabien Gouyon and Simon Dixon. A review of rhythm description systems. *Computer Music Journal*, 29(1):34–54, 2005.

[10] Lorin Grubb and Roger B. Dannenberg. A stochastic method of tracking a vocal performer. In *Proc. of ICMC (International Computer Music Conference)*, pages 301–308, Thessaloniki, Greece, 1997.

[11] Holger H. Hoos, Kai Renz, and Marko Gorg. GUIDO/MIR An experimental musical information retrieval system based on GUIDO music notation. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, Bloomington, Indiana, USA, 2001.

[12] David Huron. Music information processing using the humdrum toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26(2):11–26, 2002.

[13] Hirokazu Kameoka, Takuya Nishimoto, and Shigeki Sagayama. A multipitch analyzer based on harmonic temporal structured clustering. *Audio, Speech and Language Processing, IEEE Transactions on*, 15(3):982–994, 2007.

[14] Anssi Klapuri. Multipitch analysis of polyphonic music and speech signals using an auditory model. *Audio, Speech and Language Processing, IEEE Transactions on*, 16(2):255–266, 2008.

[15] Anssi Klapuri, Antti Eronen, and Jaakko Astola. Analysis of the meter of acoustic musical signals. *Audio, Speech and Language Processing, IEEE Transactions on*, 14(1):342–355, 2006.

[16] Jean Laroche. Efficient tempo and beat tracking in audio recordings. *JAES (Journal of the Audio Engineering Society)*, 51(4):226–233, 2003.

[17] Mark Levy and Mark Sandler. Structural segmentation of musical audio by constrained clustering. *Audio, Speech and Language Processing, IEEE Transactions on*, 16(2):318–326, 2008.

[18] Beth Logan and Stephen Chu. Music summarization using key phrases. In *Proc. of IEEE ICASSP (International Conference on Acoustics, Speech, and Signal Processing)*, volume II, pages 749–752, Istanbul, Turkey, 2000.

[19] Alex Loscos, Pedro Cano, and Jordi Bonada. Low-delay singing voice alignment to text. In *Proc. of ICMC (International Computer Music Conference)*, page 23, Bejing, China, 1999.

[20] Matthias Mauch and Simon Dixon. Simultaneous estimation of chords and musical context from audio. *Audio, Speech and Language Processing, IEEE Transactions on*, 18(6):1280 – 1289, 2010.

[21] Cory McKay and Ichiro Fujinaga. jSymbolic: A feature extractor for MIDI files. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 302–5, Victoria, BC, Canada, 2006.

[22] Nicola Orio and Diemo Schwarz. Alignment of monophonic and polyphonic music to a score. In *Proc. of ICMC (International Computer Music Conference)*, La Havana, Cuba, 2001.

[23] Hélène Papadopoulos and Geoffroy Peeters. Joint estimation of chords and downbeats from an audio signal. *Audio, Speech and Language Processing, IEEE Transactions on*, 19(1):138 – 152, January 2010.

[24] Jouni Paulus and Anssi Klapuri. Music structure analysis using a probabilistic fitness measure and a greedy search algorithm. *Audio, Speech and Language Processing, IEEE Transactions on*, 17(6):1159–1170, 2009.

[25] Jouni Paulus, Meinard Müller, and Anssi Klapuri. Audio-based music structure analysis. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, Utrecht, The Netherlands, 2010.

[26] Johan Pauwels and Jean-Pierre Martens. Integrating musicological knowledge into a probabilistic system for chord and key extraction. In *Proc. AES 128th Convention*, London, UK, 2010.

[27] Geoffroy Peeters. *Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation: Sequence and State Approach*, pages 142–165. Lecture Notes in Computer Science. Springer-Verlag Berlin Heidelberg 2004, 2004.

[28] Geoffroy Peeters and Hélène Papadopoulos. Simultaneous beat and downbeat-tracking using a probabilistic framework: theory and large-scale evaluation. *Audio, Speech and Language Processing, IEEE Transactions on*, 19(6):1754–1769, August 2011.

[29] Miller Puckette. Explode: A user interface for sequencing and score following. In *Proc. of ICMC (International Computer Music Conference)*, pages 259–261, 1990.

[30] Christopher Raphael. Automatic segmentation of acoustic musical signals using hidden markov models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(4):360–370, 1999.

[31] Matti P. Ryynänen and Anssi P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32(3):72–86, 2008.

[32] Justin Salamon and Emilia Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *Audio, Speech and Language Processing, IEEE Transactions on*, 20(6):1759–1770, 2012.

[33] Eric Scheirer. Tempo and beat analysis of acoustic musical signals. *JASA (Journal of the Acoustical Society of America)*, 103(1):588–601, 1998.

[34] Barry Vercoe. The synthetic performer in the context of live performance. In *Proc. of ICMC (International Computer Music Conference)*, pages 199–200, Paris, France, 1984.

[35] Emmanuel Vincent, Nancy Berlin, and Roland Badeau. Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription. In *Proc. of IEEE ICASSP (International Conference on Acoustics, Speech, and Signal Processing)*, pages 109–112, Las Vegas, Nevada, USA, 2008. IEEE.

[36] Ye Wang, Min-Yen Kan, Tin Lay Nwe, Arun Shenoy, and Jun Yin. Lyrically: automatic synchronization of acoustic musical signals and textual lyrics. In *Proc. of ACM Multimedia*, pages 212–219. ACM, 2004.

[37] Chunghsin Yeh, Axel Roebel, and Xavier Rodet. Multiple fundamental frequency estimation and polyphony inference of polyphonic music signals. In *Audio, Speech and Language Processing, IEEE Transactions on*, volume 18, page 6, 2010.

## 2.6 EVALUATION METHODOLOGIES

It is paramount to MIR that independent researchers build upon previous research, and an overarching challenge in MIR is to define and implement research evaluation methodologies that effectively contribute to creation of knowledge and general improvements in the field. In many scientific disciplines dealing with data processing, significant improvements over the long term have been achieved by empirically defining evaluation methodologies via several iterations of an experimental "loop" including formalisation, implementation, experimentation, and finally validity analysis. In MIR, evaluation initiatives have played an increasing role in the last 10 years, and the community is presently facing the validity analysis issue: that is, finding the most appropriate way to build upon its own legacy and redefine the evaluation methodologies that will better lead to future improvements, the resolution of which will in turn entail further technical challenges down the line (i.e., down the "loop"). This will require above all a deeper involvement of more MIR researchers in the very definition of the evaluation methodologies, as they are the individuals with the best understanding of relevant computational issues. Importantly, this will also require the involvement of the music industry (via e.g. proposing evaluations of relevance to them), and content providers (in order for researchers to have access to data). Effective MIR evaluations will impact in a fundamental manner the very way MIR research is done, it will positively affect the width and depth of the MIR research, and it will increase the relevance of MIR to other research fields.

### 2.6.1 State of the art

Many experimental disciplines have witnessed significant improvements over the long term thanks to community-wide efforts in systematic evaluations. This is the case for instance of (text-based) Information Retrieval with the TREC initiative (Text REtrieval Conference see [34]) and the CLEF initiative (Cross-Language Evaluation Forum [35]), Speech Recognition [8], Machine Learning [5], and Video and Multimedia Retrieval with e.g. the TRECVID [36] and VideoCLEF initiatives (the latter later generalised to the "MediaEval Benchmarking Initiative for Multimedia Evaluation" [37]).

Although evaluation "per se" has not been a traditional focus of pioneering computer music conferences (such as the ICMC) and journals (e.g. Computer Music Journal), recent attention has been given to the topic. In 1992, the visionary Marvin Minsky declared: "the most critical thing, in both music research and general AI research, is to learn how to build a common music database" [6], but it was not until a series of encounters, workshops and special sessions organised between 1999 and 2003 by researchers from the newly-born Music Information Retrieval community that the necessity of conducting rigorous and comprehensive evaluations was recognised [3].

The first public international evaluation benchmark took place at the ISMIR Conference 2004 [2], where the objective was to compare state-of-the-art audio algorithms and systems relevant for some tasks of music content description. This effort has then been systematised and continued via the yearly Music Information Retrieval Evaluation eXchange (MIREX). MIREX has widened the scope of the evaluations and now covers a broad range of tasks, including symbolic data description and retrieval [4].

The number of evaluation endeavors issued from different communities (e.g. Signal Processing, Data Mining, Information Retrieval), yet relevant to MIR, has recently increased significantly. For instance, the Signal Separation Evaluation Campaign (SiSEC [38]) was started in 2008, and deals with aspects of source separation in signals of different natures (music, audio, biomedical, etc.). A Data Mining contest was organised at the 19th International Symposium on Methodologies for Intelligent Systems (ISMIS) with two tracks relevant to MIR research (Tunedit): Music Genre recognition and Musical Instrument recognition [39]. The CLEF initiative (an IR evaluation forum)

[34] http://trec.nist.gov
[35] http://www.clef-initiative.eu
[36] http://www-nlpir.nist.gov/projects/trecvid
[37] http://multimediaeval.org
[38] http://sisec.wiki.irisa.fr
[39] http://tunedit.org/challenge/music-retrieval

extended its scope to MIR with the MusiCLEF initiative [40] [7]. The ACM Special Interest Group on Knowledge Discovery and Data Mining organises a yearly competition, the KDD Cup [41], focusing on diverse Data Mining topics every year, and in 2011, the competition focused on a core MIR topic: Music Recommendation [42]. In 2012, the MediaEval Benchmarking Initiative for Multimedia Evaluation [43] organised a music-related task for the first time. Also in 2012, the Million Song Dataset challenge appeared, a music recommendation challenge incorporating many different sorts of data (user data, tags . . . ) [44].

The establishment of an annual evaluation forum (MIREX), accepted by the community, and the appearance of relevant satellite forums in neighbouring fields have undoubtedly been beneficial to the MIR field. However, a lot of work is still necessary to reach a level where evaluations will have a systematic and traceable positive impact on the development of MIR systems and on the creation of new knowledge in MIR. For about 10 years, meta-evaluation methodologies have been instrumental in advancement of the Text Information Retrieval field; they need to be addressed in MIR too [12]. The special panel and late-breaking news session held at ISMIR 2012 addressed the various methodologies used in the MIR field and compared those to the ones used in other fields such as Media-Eval [9].

### Reproducible Research

Much computational science research is conducted without regard to the long-term sustainability of the outcomes of the research, apart from that which appears in journal and conference publications. Outcomes such as research data and computer software are often stored on local computers, and can be lost over time as projects end, students graduate and equipment fails and/or is replaced. Enormous effort is invested in the production of these outputs, which have great potential value for future research, but the benefit of this effort is rarely felt outside of the research group in which it took place. Arguments for sustainability begin with the cost-savings that result from re-use of software and data, but extend to other issues more fundamental to the scientific process. These are enunciated in the "reproducible research" movement [1, 13], which promotes the idea that, along with any scientific publication, there should be a simultaneous release of all software and data used in generating the results in the publication, so that results may be verified, comparisons with alternative approaches performed, and algorithms extended, without the significant overhead of reimplementing published work.

Various practical difficulties hinder the creation of long-term sustainable research outputs. The research software development process is usually gradual and exploratory, rather than following standard software engineering principles. This makes code less robust, so that it requires greater effort to maintain and adapt. Researchers have varying levels of coding ability, and may be unwilling to publicise their less-than-perfect efforts. Even when researchers do make code available, their priority is to move on to other research, rather than undertake the additional software engineering effort that might make their research more usable. Such software engineering efforts might be difficult to justify in research funding proposals, where funding priority is given to work that is seen to be "research" over "development" efforts. Also, research career progression tends to be awarded on the basis of high-impact papers, while software, data and other outputs are rarely considered. Another perceived difficulty is that public release of software might compromise later opportunities for commercialisation, although various licenses exist which allow both to occur [11].

To these general problems we may add several issues specific to the music information research community. The release of data is restricted by copyright regulations, particularly relating to audio recordings, but this is also relevant for scores, MIDI files, and other types of data. The laws are complex and vary between countries. Many researchers, being unsure of the legal ramifications of the release of data, prefer the safer option of not releasing data. Reliance on specific hardware or software platforms also makes code difficult to maintain in the longer term. One solution for obsolete hardware platforms is the use of software emulation, as addressed by the EU projects PLANETS and KEEP. For music-related research, such general-purpose emulation platforms might not be sufficient to reproduce audio-specific hardware [10].

[40] http://clef2012.org/resources/slides/MusiClef.pdf
[41] http://www.sigkdd.org/kddcup/index.php
[42] http://www.kdd.org/kdd2011/kddcup.shtml
[43] http://www.multimediaeval.org
[44] http://www.kaggle.com/c/msdchallenge

In the MIR community, great effort has been expended to provide a framework for the comparison of music analysis and classification algorithms, via the MIREX evaluations, as well as the more recent MusiClef and MSD challenges (c.f. section 2.6). More recently, the Mellon-funded NEMA project [45] attempted to develop a web service to allow researchers to test their algorithms outside of the annual MIREX cycle. Although there are a growing number of open-access journals and repositories for software and data, there are obstacles such as publication costs and lack of training which hinder widespread adoption. Addressing the training aspect are the Sound Software [46] and Sound Data Management Training [47] projects, the SHERPA/RoMEO project, which contains information on journal self-archiving and open access policies, and the spol initiative [48] for reproducable research.

### 2.6.2 Specific Challenges

- **Promote best practice evaluation methodology within the MIR community.** The MIR community should strive to promote within itself, at the level of individual researchers, the use of proper evaluations, when appropriate.
- **Define meaningful evaluation tasks.** Specific tasks that are part of large-scale international evaluations define *de facto* the topics that new contributors to the MIR field will work on. The very definition of such tasks is therefore of utmost importance and should be addressed according to some agreed criteria. For instance, tasks should have a well-defined community of users for whom they are relevant, e.g. while audio onset detection is only marginally relevant for industry, it is very relevant to research. The MIR research community should also open up to tasks defined by the industry, e.g. as the Multimedia community does with the "Grand Challenges" at the ACM Multimedia conference.
- **Define meaningful evaluation methodologies.** Evaluation of algorithms should effectively contribute to the creation of knowledge and general improvements in the MIR community. Effectively building upon MIR legacy and providing meaningful improvements call for a constant questioning of all aspects of the evaluation methodology (metrics, corpus definition, etc.). For instance, evaluation metrics are currently useful for quantifying each system's performance; a challenge is that they also provide *qualitative* insights on how to improve this system. Also, data curation is costly and time-consuming, which implies a challenge to aggregate, for evaluation purposes, data and metadata with the quality of a curated collection, and to preserve provenance.
- **Evaluate whole MIR systems.** While evaluation of basic MIR components (estimators for beat, chords, fundamental frequency, etc.) is important, the MIR community must dedicate more effort to evaluation of whole MIR systems, e.g. music recommendation systems, music browsing systems, etc. Such evaluations will lead to insights with regards to which components are relevant to the system and which not.
- **Promote evaluation tasks using multimodal data**. Most MIR systems are concerned with audio-only or symbolic-only scenarios. A particular challenge is to target the evaluation of multimodal systems, aggregating information from e.g. audio, text, etc.
- **Implement sustainable MIR evaluation initiatives.** An important challenge for MIR evaluation initiatives is to address their *sustainability* in time. The MIR community must dedicate more effort to its legacy in terms of evaluation frameworks. This implies many issues related for example to general funding, data availability, manpower, infrastructure costs, continuity, reproducibility, etc.
- **Target long-term sustainability of Music Information Research.** Focusing on the sustainability of MIR evaluation initiatives is only part of the general challenge to target long-term sustainability of MIR itself. In particular, consistent efforts should be made to foster reproducible research through papers,

[45] http://nema.lis.illinois.edu/?q=node/12
[46] http://www.soundsoftware.ac.uk
[47] http://rdm.c4dm.eecs.qmul.ac.uk/category/project/sodamat
[48] spol-discuss@list.ipol.im

software and data that can be reused to verify or extend published work. Also training will be necessary to effect the reorientation of the MIR community to adopt research practices for ensuring reproducibility, from the code, data, and publication perspectives. Any progress towards creating reproducible research will have an immediate impact not only on the MIR field, but also towards the application of MIR technologies in other research fields.

## References

[1] J. Buckheit and D. L. Donoho, editors. *Wavelets and Statistics*, chapter Wavelab and reproducible research. Springer-Verlag, Berlin, New York, 1995.

[2] P. Cano, E. Gómez, F. Gouyon, P. Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich, and N. Wack. *ISMIR 2004 Audio Description Contest.* Music Technology Group Technical Report, University Pompeu Fabra, MTG-TR-2006-02, 2006.

[3] J. Stephen Downie. *The MIR/MDL Evaluation Project White Paper Collection: Establishing Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation Frameworks : Preliminary Foundations and Intrastructures.* J. Stephen Downie, Graduate School of Library and Information Science, University of Illinois, 2003.

[4] J. Stephen Downie. The Music Information Retrieval Evaluation eXchange (MIREX). *D-Lib Magazine*, Dec. 2006.

[5] Isabelle Guyon, Steve R. Gunn, Asa Ben-Hur, and Gideon Dror. Result analysis of the NIPS 2003 feature selection challenge. In *Proc. NIPS (Advances in Neural Information Processing Systems)*, 2004.

[6] Marvin Minsky and Otto E. Laske. A conversation with Marvin Minsky. *AI Magazine*, 13(3):31–45, 1992.

[7] Nicola Orio, David Rizo, Riccardo Miotto, Markus Schedl, Nicola Montecchio, and Olivier Lartillot. MusiCLEF: A benchmark activity in multimodal music information retrieval. In *Proc. ISMIR (International Society for Music Information Retrieval Conference)*, pages 603–608, 2011.

[8] David Pearce and Hans-Gunter Hirsch. The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *Proc. INTERSPEECH (Conference of the International Speech Communication Association)*, pages 29–32, 2000.

[9] Geoffroy Peeters, Julián Urbano, and Gareth J. F. Jones. Notes from the ISMIR12 late-breaking session on evaluation in music information retrieval. In *Proc. ISMIR (International Society for Music Information Retrieval Conference)*, 2012.

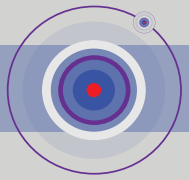
[10] Bruce Pennycook. Who will turn the knobs when I die? *Organised Sound*, 13(3):199–208, December 2008.

[11] Victoria Stodden. The legal framework for reproducible scientific research: Licensing and copyright. *Computing in Science and Engineering*, 11(1):35–40, 2009.

[12] Julián Urbano, Markus Schedl, and Xavier Serra. Evaluation in music information retrieval. *Journal of Intelligent Information Systems*, in print, 2013.

[13] Patrick Vandewalle, Jelena Kovacevic, and Martin Vetterli. Reproducible research in signal processing - what, why, and how. *IEEE Signal Processing Magazine*, 26(3):37–47, May 2009.

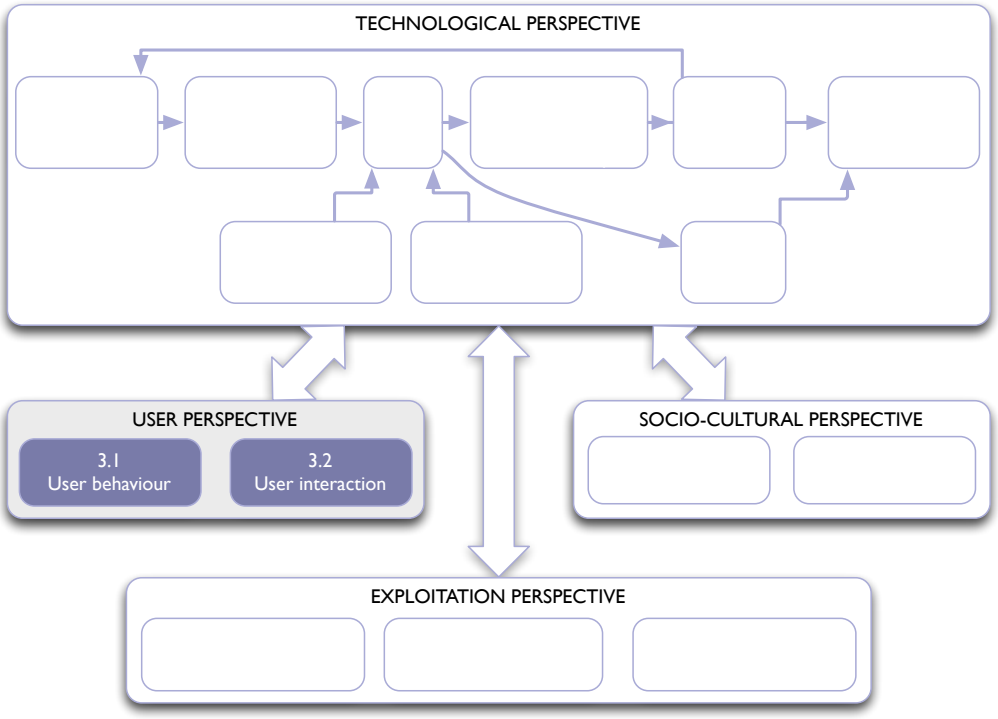
# USER PERSPECTIVE



CHAPTER 3

# User perspective

*Music Information Research considers the user perspective, both in order to understand the user roles within the music communication chain and to develop technologies for the interaction of these users with music data. MIR aims to capture, process and model the data gathered through user interaction and develop methodologies for the design of new musical devices in order to enable new interaction possibilities between users and these devices.*



## 3.1 USER BEHAVIOUR

Music is listened to, performed and created by people. It is therefore essential to consider the user as central to the creation of user scenarios, hence to the development of technologies. Developing user applications involves analysing the user needs in respect of novel scenarios and the user behaviour in respect of existing ones, thus enabling the creation of the user-specification-development loop. Taking into account user needs applies to all stages of the development loop, however the analysis of user behaviour must be carefully conducted by a specialist. Gathering feedback from users is a research field in itself and shouldn't be done without carefully designed methods. Considering user needs through the analysis of user behaviour will have a great impact on the usability of the developed MIR technologies.

### 3.1.1 State of the art

Activities related to music can be roughly grouped into (i) listening (to recorded media or live performances; review/discussion of what was heard), (ii) performing (interpretation, improvisation, rehearsal, recording, live performance) and (iii) creating (composition, recording, studio production, improvisation). Other activities are concerned with researching, studying (education, musicology), sharing, worship and dance (see part 5.3).

Within each group, MI research can relate to the analysis of practices or to the proposal of tools to help the practice.

#### Listening.

Among these categories, research presented in conferences such as ISMIR mainly focus on the listening scenario and propose tools to help people access (listen to) music. But little attention is paid to analysing user practices. As pointed out by [24], a focus on the user has repeatedly been identified as a key requirement for future MIR research, yet empirical user studies have been relatively sparse in the literature, the overwhelming research attention in MIR remaining systems-focused. [13] proposes an overview of user studies performed so far in the MIR field and propose explanation why their impact on the field have been weak so far: lack of findability, dominance of small scaled studies that are difficult to generalize. Important questions related to the user are: What are its requirements and information needs? How do people organise their music? How would they like to see, access, search through digital libraries? What is the influence of the listening context? What is the role of social relations? Given that one of the Grand-Challenges in MIR is the creation of a full-featured system [7], these questions should be answered in order to make the system useful for users. This is especially true considering that the results provided by the little research done on this topic yielded unexpected results. For example [12] showed that some of the users are seeking new music without specific goals in mind, possibly just to update and expand their musical knowledge or for the sheer pleasure of searching. With this in mind, systems should support various browsing approaches. [3] highlight user needs for use tagging (scenarios in which a given piece of music might be relevant), a subject currently largely under-studied. [11] identifies the changes in musical taste according to social factors and [4] suggest support for collaborative play-list creation. [23] conclude that textual queries for melodic content are too difficult to be used by ordinary users. The various possibilities to design music recommendation systems that take user into account are summarized in [20]. According to [10], landscape representations or geographic views of music collections have certain disadvantages and that users seem to have preferences for simple and clean interfaces. A recent survey made within the Chorus+ EU project [14], also highlights important points such as the prevalence of YouTube as the most-used music service (among participants to the survey). It also highlights the fact that most people search using artist, composer, song title, album or genre but the search possibilities enabled by new technologies (taste, mood or similarity) appear less prevalent.

#### Performing.

If few papers relate to the listener-behaviour, this is not the case for performers and performances (in terms of music concerts, opera, theatre, dance) or interactions (interactive installations or instruments). A large community has been studying the subject of performance from the pioneer works of [21]. In this, a performer is considered

as the essential mediator between composer and listener. These studies show the impact of the performer, the performances, the large-structure and micro-structure, and the intentional mood on the choice of tempo, timing, loudness, timbre and articulation [8, 19]. First experiments were made using piano analysis (for ease of event-recoding) [18], but today they are extended to saxophone [16], cello [2] and singing voice. Understanding the process of performance has several goals: a better understanding of what makes a great interpretation (the Horowitz or Rachmaninov factors [25]); music education; and automatic expressive perfomances (KTH model of [22] and Rendering Contest (Rencon) [1]). Tools to visualise performance interpretation have also been proposed [6]. According to Delgado [5], different research strategies can be distinguished: (a) analysis-by-measurement (based on acoustic and statistical analysis of performances); (b) analysis-by-synthesis (based on interviewing expert musicians); and (c) inductive machine learning applied to large database of performances. An example of the use of MIR research for inductive machine learning is given by [2]. Considering that performance is not limited to the instrumentalists, the conductor is also studied [15], and research includes studies on interaction and gesture ([9], [1]). The large number of related contributions at conferences such as ISPS (International Symposium on Performance Science) [2] shows that this domain is very active. As another example of the activity in this field, the current SIEMPRE EU [3] project aims at developing new theoretical frameworks, computational methods and algorithms for the analysis of creative social behaviour with a focus on ensemble musical performance.

#### Composing.

While historical musicology aims at studying composition once published, hence not considering the composition practice, research groups such as the one of Barry Eaglestone [17] at the Information Systems and the Music Informatics research groups, or new projects such as MuTec2 [4] aim at following composers during their creative project (using sketches, drafts, composer interviews, and considering composer readings). Related to this new field, the conference TCPM-2011 "Tracking the Creative Process in Music" [5] has been created.

### 3.1.2 Specific Challenges

- **Analyse user needs and behaviour carefully.** Gathering feedback from users is actually a research field in itself and shouldn't be done without carefully designed methods.
- **Develop tools and technologies that take user needs and behaviour into account.** Much work in MIR is technology-driven, rather than being user, context or application driven. Thus the challenge is to step into the shoes of users and understand their world-view in order to produce useful applications. User studies must be considered right from the beginning of a research project. Appropriate tools and technologies must be developed to cater for different types of users who perform the same task in different contexts.
- **Identify and study new user roles related to music activities.** The aforementioned user-types (listener, performer and composer) are prototypes and not orthogonal (guitar-hero involves both listening and performing). Moreover users can have different expertise for the same role (common people, musicologist). Development of MIR tools will also create new user-profiles that need to be identified and taken into account.
- **Develop tools that automatically adapt to the user.** According to the role, profile and context the tools must be personalised. Those profiles are therefore dynamic, multidimensional. Those are also fuzzy given the nature of the input provided by the user. An alternative to the personalisation of tools is the use of a companion (the "music guru").

[1] http://renconmusic.org
[2] http://www.performancescience.org
[3] http://siempre.infomus.org
[4] http://apm.ircam.fr/MUTEC
[5] http://tcpm2011.meshs.fr/?lang=en

## References


[1] Frédéric Bevilacqua, Norbert Schnell, and Sarah Fdili Alaoui. Gesture Capture: Paradigms in Interactive Music/Dance Systems. In Transcript Verlag, editor, *Emerging Bodies: The Performance of Worldmaking in Dance and Choreography*, pages 183–193. Gabriele Klein and Sandra Noeth, 2011.

[2] Magdalena Chudy and Simon Dixon. Towards music performer recognition using timbre. In *Proc. of SysMus10 (Third International Conference of Students of Systematic Musicology)*, Cambridge, UK, 2010.

[3] Sally Jo Cunningham, Matt Jones, and Steve Jones. Organizing digital music for use: an examination of personal music collections. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 447– 454, Barcelona (Spain), 2004.

[4] Sally Jo Cunningham and David M. Nichols. Exploring social music behaviour: An investigation of music selection at parties. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 26–30, Kobe, Japan, 2009.

[5] Miguel Delgado, Waldo Fajardo, and Miguel Molina-Solana. A state of the art on computational music performance. *Expert Systems with Applications*, 38(1):155–160, 2011.

[6] Simon Dixon, Werner Goebl, and Gerhard Widmer. The performance worm: Real time visualisation of expression based on langners tempo-loudness animation. In *Proc. of ICMC (International Computer Music Conference)*, Göteborg, Sweden, 2002.

[7] J. Stephen Downie, Donald Byrd, and Tim Crawford. Ten years of ISMIR: Reflections on challenges and opportunities. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 13–18, Kobe, Japan, 2009.

[8] Alf Gabrielsson. Music performance research at the millennium. *Psychology of music*, 31(3):221–272, 2003.

[9] Sergi Jorda. Interactive music systems for everyone exploring visual feedback as a way for creating more intuitive, efficient and learnable instruments. In *Proc. of SMAC (Stockholm Music Acoustics Conference)*, Stockholm, Sweden, 2003.

[10] Philipp Kolhoff, Jacqueline Preuß, and Jörn Loviscach. Content-based icons for music files. *Computers and Graphics*, 32(5):550–560, 2008.

[11] Audrey Laplante. The role people play in adolescents music information acquisition. In *Proc. of WOMRAD (Workshop on Music Recommendation and Discovery)*, Barcelona, Spain, 2010.

[12] Audrey Laplante and J. Stephen Downie. Everyday life music information-seeking behaviour of young adults. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 381–382, Victoria, BC, Canada, 2006.

[13] Jin Ha Lee and Sally Jo Cunningham. The impact (or non-impact) of user studies in music information retrieval. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, Porto, Portugal, 2012.

[14] Thomas Lidy and Pieter van der Linden. Think-tank on the future of music search, access and consumption. In European Community's Seventh Framework Programme (FP7/2007-2013), editor, *MIDEM*, Cannes, France, 2011.

[15] Geoff Luck, Petri Toiviainen, and Marc R. Thompson. Perception of expression in conductors' gestures: A continuous response study. *Music Perception*, 28(1):47–57, 2010.

[16] Esteban Maestre, Antonio Pertusa, and Rafael Ramirez. Identifying saxophonists from their playing styles. In *Proc. of the 30th AES International Conference on Intelligent Audio Environments*, Saariselkä, Finland, 2007.

[17] Ralf Nuhn, Barry Eaglestone, Nigel Ford, Adrian Moore, and Guy Brown. A qualitative analysis of composers at work. In *Proc. of ICMC (International Computer Music Conference)*, Göteborg, Sweden, 2002. Citeseer.

[18] Richard Parncutt. Accents and expression in piano performance. *Perspektiven und Methoden einer Systemischen Musikwissenschaft*, pages 163–185, 2003.

[19] John Rink. *The Practice Of Performance: Studies In Musical Interpretation.* Cambridge University Press Cambridge, 1995.

[20] Markus Schedl and Arthur Flexer. Putting the user in the center of music information retrieval. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, Porto, Portugal, 2012.

[21] Carl E. Seashore. *Psychology of music.* New York: McGraw Hill Book Company Inc., 1938.

[22] Johan Sundberg, Lars Fryden, and Anders Askenfelt. *What tells you the player is musical? An analysis-by-synthesis study of music performance*, volume 39, pages 61–75. Stockholm, Sweden: Publication issued by the Royal Swedish Academy of Music, 1983.

[23] Alexandra L. Uitdenbogerd and Yaw Wah Yap. Was Parsons right? An experiment in usability of music representations for melody-based music retrieval. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 75–79, Baltimore, Maryland, USA, 2003.

[24] David M. Weigl and Catherine Guastavino. User studies in the music information retrieval literature. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, Miami, Florida, USA, 2011.

[25] Gerard Widmer, Simon Dixon, Werner Goebl, Elias Pampalk, and Asmir Tobudic. In search of the horowitz factor. *AI Magazine*, 24(3):111–130, 2003.

## 3.2 USER INTERACTION

The grand challenge of user interaction is how to design MIR systems that put the user at the centre of the system. This applies to the whole interaction loop, including visualisation, input devices, manipulation metaphors, and also system adaptation to user behaviour. This challenge is relevant because it contributes to both the user's and to the researcher's (e.g. system designer's) understanding of the system's features and components, the overall purpose of the system, and the contribution the system can make to the user's activities. The benefit to users is more productive workflows and systems which better serve the users' needs. The researchers stand to benefit from the feedback loop which enables them to fine-tune and develop systems with greater accuracy. Effective user-oriented research will have a major impact on the usability of MIR systems and their wider deployment.

### 3.2.1 State of the art

In the last decade, Human Computer Interaction (HCI) research has witnessed a change in focus from conventional ways to control and communicate with computers (keyboards, joysticks, mice, knobs, levers, buttons, etc.) to more intuitive uses of non-conventional devices such as gloves, speech recognition, eye trackers, cameras, and tangible user interfaces. As a result of technological advances and the desire to surpass the WIMP (window, icon, menu, pointing device) limitations, interaction research has progressed beyond the desktop and the ubiquitous graphical user interface (GUI) into new physical and social contexts. Since terms such as "multi-touch" and gestures like "two-finger pinch and zoom" have become part of the users' daily life, novel research areas such as "tangible interaction" have finally entered the mainstream. However, aside from the ongoing research explicitly focused towards real-time musical performance which typically falls under the New Interfaces for Musical Expression (NIME[6]) discipline, not much of this research has yet been devoted to novel interface and interaction concepts in the field of MIR.

#### The use of HCI and related methodologies in MIR

The Association for Computing Machinery defines human-computer interaction (HCI) as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them."[7] HCI involves the study, planning, and design of the interaction between people (users) and computers. It is often regarded as the intersection of computer science, behavioural sciences, design and several other fields of study. Interaction between users and computers occurs at the interface which is the result of particular affordances of a given combination of software and hardware. The basic and initial goal of HCI is therefore to improve the interactions between users and computers by making computers more usable and responsive to the user's needs. For decades HCI has mostly focused on making interaction more efficient, though more recently the emphasis has shifted to the user's *Quality of Experience*, highlighting the benefits of beauty and fun, and the intrinsic values of the experience and its outcomes [e.g. 24, 27]. The human component in HCI is therefore highly relevant from the cultural, psychological and physiological perspectives.

MIR could benefit from knowledge inherited from HCI and other related disciplines such as User Experience (UX), and Interface and Interaction Design studies. These methodologies could bring benefits not only to the conception of MIR systems at earlier design stages, but also for the evaluation and subsequent iterative refinement of these systems. While the evaluation of MIR systems is traditionally and historically conceived to provide categorically correct answers (e.g finding or identifying a known target song), new evaluation challenges are presented by open systems which leave users room for interpretation [e.g. 30], include more subjective aspects (e.g. the users' emotions, perceptions and internal states [e.g. 10]), or encourage contextual engagement [e.g. 9][9]. Furthermore, beyond the evaluation of User Experience, another MIR component that would directly benefit from HCI-related knowledge would be research into open and holistic systems for the creation of MIR systems and tools.

[6] http://www.nime.org
[7] ACM SIGCHI Curricula for Human-Computer Interaction: [8]
[9] The current SOA in MIR evaluation of research results is covered in section 2.6

## Music Search Interaction

Over the past 12 years a few projects from the MIR community have contributed to the development of interfaces for music search and discovery. In the field of data visualisation, there is an extensive bibliography on the representation of auditory data. In the particular case of the visual organisation of musical data, solutions often consist of extracting feature descriptors from data files, and creating a multidimensional feature space that will be projected onto a 2D surface, using dimensionality reduction techniques (e.g. *Islands of Music* [28]; *SOM: Self Organizing Map* [17]; *SOMeJB* [19] and [5]). Beyond 2D views, the advantage of a topological metaphor has been applied to facilitate users' exploration of big data collections in *nepTune*, an interactively explorable 3D version of *Islands of Music*, which supports spatialised sound playback [16], and the *Globe of Music* which places a collection on a spherical surface to avoid any edges or discontinuities [18]. More recently, *MusicGalaxy* [32] implements an adaptive zoomable interface for exploration that makes use of a complex non-linear multi-focal zoom lens and introduces the concept of facet distances representing different aspects of music similarity. *Musicream* [7] uses the "search by example" paradigm, representing the songs with dynamic coloured circles which fall from the top of the screen and when selected show their title and can be used to 'fish' for similar ones.

In terms of developing a user-oriented visual language for screen-based music searches, the interactive aspect of most commercial library music applications has resorted to the metaphor of spreadsheets (e.g. *iTunes*) or has relied on searching for music by filling a set of forms and radio buttons (e.g. *SynchTank*). Innovative approaches from the MIR community suggested visually mapping sound clusters into abstract "islands" (e.g. [28]); collaborative mapping onto real geographical visual references (e.g. *Freesound* [10]); and tangible tabletop abstract symbols (e.g. *SongExplorer* [14]). Visual references have included control panels used in engineering (e.g. *MusicBox* [20]); gaming platforms (*Musicream* [7]); lines of notation (e.g. *Sonaris* and *mHashup* [23]); or turntables (*Songle* [8]).

A few MIR-driven search interfaces have addressed different user contexts. *Mediasquare* [4] addresses social interaction in 3D virtual space where users are impersonated by avatars enabling them to browse and experience multimedia content by literally walking through it. *decibel 151* [22] enables multi-user social interaction in physical space by turning each user into a "walking playlist", creating search environments for social networking in real time. Special visual interfaces have addressed poorly described or less familiar music to the user (e.g. field recordings; ethnomusicological collections) to both educate and allow music discovery in an entertaining way (e.g. *Songlines* 2010 and [21]). User contexts however remain vastly under-researched and remain a major challenge for the MIR community.

Some of the above interfaces have adopted HCI research methods which consider MIR-driven search systems holistically, not only as visual representations of data, but focusing on the user Quality of Experience. This resulted from a coherent system design approach which creates a feedback loop for an iterative research and innovation process between the interactive front end and the data processing back end of the application. Further research challenges are presented by a holistic approach to MIR user-oriented system design in the context of novel devices and modalities, real-time networks, collaborative platforms, open systems, physical experiences and tangible interfaces.

## Tangible and Tabletop Interaction

Tangible User Interfaces (TUI), which combine control and representation in a single physical device emphasise tangibility and materiality, physical embodiment of data, bodily interaction and the embedding of systems in real spaces and contexts. Although several implementations predate this concept, the term Tangible User Interface was coined at the MIT MediaLab in 1997 [33] to define interfaces which augment the real physical world by coupling digital information to everyday physical objects and environments. Such interfaces contribute to the user experience by fusing the representation and control of digital data with physical artefacts thus allowing users to literally "grasp data" with their own hands.

Within the domain of Tangible Interaction, Tabletop Interaction constitutes a special research field which uses the paradigm of a horizontal surface meant to be touched and/or manipulated via the objects placed on it. In contrast to the mouse and keyboard interface model which restricts the user's input to an ordered sequence of

[10] http://www.freesound.org

events (click, click, double click, etc.), this type of interface allows multiple input events to enter the system at the same time, enabling any action at any time or position, by one or several simultaneous users. The implicit ability of tabletop interfaces to support physical tracked objects with particular volume, shape and weight properties, expands the bandwidth and richness of the interaction beyond the simple idea of multi-touch. Such objects can represent abstract concepts or real entities; they can relate to other objects on the surface; they can be moved and turned around on the table surface, and these spatial changes can affect their internal properties and their relationships with neighbouring objects. The availability of open-source, cross-platform computer vision frameworks that allow the tracking of fiducial markers combined with multi-touch finger tracking (e.g. *reacTIVision*, which was developed for the *Reactable* project [1]), have become widely used among the tabletop developers community (both academic and industrial), and have increased the development of tabletop applications for educational and creative use (e.g. [15];[6].

There is a growing interest in applying Tabletop Interfaces to the music domain. From the *Audiopad* [29] to the *Reactable* [13], music performance and creation has become the most popular and successful application field in the entire lifetime of this interaction paradigm. Tabletop interfaces developed using MIR have specifically focused on interacting with large music collections. *Musictable* [31], takes a visualisation approach similar to the one chosen in Pampalk's *Islands of Music*, for creating a two dimensional map that, when projected on a table, is used to make collaborative decisions to generate playlists. Hitchner [11] uses a SOM to build the map visually represented by a low-resolution mosaic, enabling the users to redistribute the songs according to their preferences. *Audioscapes* is a framework enabling innovative ways of interacting with large audio collections using touch-based and gestural controllers [26]. The MTG's *SongExplorer* [14] uses high-level descriptors of musical songs applied to N-Dimensional navigation on a 2D plane, thus creating a coherent 2D map based on similarity with specially designed tangible pucks for more intuitive interaction with the tabletop visual interface. Tests comparing the system with a conventional GUI interface controlling the same music collection, showed that the tabletop implementation was a much more efficient tool for discovering new, valuable music to the users. Thus the specific affordances of tabletop interfaces (support of collaboration and sharing of control; continuous, real-time interaction with multidimensional data; support of complex, expressive and explorative interaction [12]), together with the more ubiquitous and easily available individual multi-touch devices, such as tablets and smart-phones, can bring novel approaches to the field of MIR, not only for music browsing but particularly for the more creative aspects related to MIR music creation and performance.

The physical embodiment of data, bodily interaction and the embedding of systems in real spaces and contexts is particularly present in recent research into gestural and spatial interaction. The Real-Time Musical Interactions team at IRCAM has been working with motion sensors embedded within everyday objects to explore concepts of physical and gestural interaction which integrate performance, gaming and musical experience. Their *Interlude* project [11] combined interactivity, multimodal modelling, movement tracking and machine learning to explore new means for musical expression ([2], [3] and [25]). The results included the *Urban Musical Game* which breaks down some of the boundaries between audience and musician by producing a sound environment through the introduction of a musical ball; *Mogees* which uses piezo sensors coupled with gesture recognition technology for music control allowing users to easily transform any surface into a musical interface; and *MOs (Modular Musical Objects)* which represent one of the pioneering attempts to answer the challenges of tangible, behaviour-driven musical objects for music creation. This project has demonstrated the huge potential of research into physical and gestural interfaces for MIR within the context of future internet applications for the Internet of Things.

### 3.2.2 Specific Challenges

- **Develop open systems which adapt to the user.** HCI research has shown that systems which leave users room for interpretation, include more subjective aspects or encourage contextual engagement, contribute to an improved Quality of Experience for the user.
- **Design MIR-based systems more holistically.** A System Design approach must include user experience,

[11] http://interlude.ircam.fr

and not only focus on the engine or the algorithms of a given system. Front and back-ends cannot be interchanged without consequences: a given algorithmic mechanism will probably favour a particular type of interface or interaction methods.

- **Develop interfaces to better address collaborative, co-creative and sharing multi-user applications.** Most MIR interfaces have been developed for a single user. In the context of open social networks, multi-user MIR applications present opportunities for enhanced co-creation and sharing or music.
- **Develop interfaces which make a broader range of music more accessible and "edutain" audiences.** Many users find that new (to them) styles of music are inaccessible. Interfaces which elucidate structure, expression, harmony, etc. can contribute to "enhanced listening" offering both education and entertainment at the same time.
- **Expand the MIR systems interaction beyond the multi-touch paradigm.** Physical tracked objects with particular volume, shape and weight properties, can considerably expand the bandwidth and richness of MIR interaction.
- **Consider the context in the design of MIR systems.** MIR methods or applications should take into account the context and device in which they will be used, e.g., a multi-user spatial environment is not simply an augmented geographic interface; interaction methodologies for a tabletop application cannot be simply transferred from those used on a smartphone or mobile screen-based device.
- **Design MIR system interfaces for existing and novel modalities for music creation.** "Real-time MIR" interface and interaction research can successfully bridge the gap between MIR and NIME (New Interfaces for Musical Expression). Physical and gestural interaction can integrate performance, gaming and musical experience.

## References


[1] R. Bencina, M. Kaltenbrunner, and S. Jordà. Improved Topological Fiducial Tracking in the reacTIVision System. In *Proceedings of the IEEE International Workshop on Projector-Camera Systems*, 2005.

[2] Frédéric Bevilacqua, Norbert Schnell, and Sarah Fdili Alaoui. Gesture Capture: Paradigms in Interactive Music/Dance Systems. In Transcript Verlag, editor, *Emerging Bodies: The Performance of Worldmaking in Dance and Choreography*, pages 183–193. Gabriele Klein and Sandra Noeth, 2011.

[3] Frédéric Bevilacqua, Norbert Schnell, Nicolas Rasamimanana, Bruno Zamborlin, and Fabrice Guédy. Online gesture analysis and control of audio processing. In *Musical Robots and Interactive Multimodal Systems*, pages 127–142. Springer, 2011.

[4] M. Dittenbach, H. Berge, R. Genswaider, A. Pesenhofer, A. Rauber, T. Lidy, and W. Merkl. Shaping 3D multimedia environments: The media square. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, Amsterdam, Netherlands, July 2007.

[5] P. Hlavac E. Pampalk and P. Herrera. Hierarchical organization and visualization of drum sample libraries. In *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx'04)*, pages 378–383, Naples, Italy, 2004.

[6] D. Gallardo, C.F. Julià, and S. Jordà. TurTan: A tangible programming language for creative exploration. In *3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems*, pages 89–92. IEEE, 2008.

[7] M. Goto and T. Goto. Musicream: New music playback interface for streaming, sticking, sorting, and recalling musical pieces. In *ISMIR 2005: Proceedings of the 6th International Conference on Music Information Retrieval*, pages 404–411, 2005.

[8] M. Goto, J. Ogata, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. PodCastle and Songle: Crowdsourcing-Based Web Services for Spoken Content Retrieval and Active Music Listening. In *Proceedings of the 2012 ACM Workshop on Crowdsourcing for Multimedia (CrowdMM 2012)*, pages 1–2, October 2005.

[9] M. Hassenzahl and N. Tractinsky. Online gesture analysis and control of audio processing, musical robots and interactive multimodal systems. *Springer Tracts in Advanced Robotics*, 74, 2011. Springer Verlag.

[10] P. Hekkert. Design aesthetics: Principles of pleasure in product design. *Psychology Science*, 48(2):157–172, 2006.

[11] S. Hitchner, J. Murdoch, and G. Tzanetakis. Music browsing using a tabletop display. In *Conference on Music Information Retrieval ISMIR*, 2007.

[12] S. Jordà. On Stage: the reactable and other musical tangibles go real. *International Journal of Arts and Technology*, 1(34):268–287, 2008.

[13] S. Jordà, M. Kaltenbrunner, and R. Bencina. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 139–146. ACM, 2007.

[14] C.F. Julià and S. Jordà. Songexplorer: A tabletop application for exploring large collections of songs. In *10th International Society for Music Information Retrieval Conference*, 2009.

[15] M. Khandelwal and A. Mazalek. Teaching Table: a tangible mentor for pre-k math education. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 191–194. ACM, 2009.

[16] P. Knees, T. Pohle, M. Schedl, and G. Widmer. Exploring Music Collections in Virtual Landscapes. *IEEE MultiMedia*, 14(3):46–54, 2007.

[17] T. Kohonen. *Self–Organizing Maps*. Springer, 2001.

[18] S. Leitich and M. Topf. Globe of Music: Music Library Visualization Using GEOSOM. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, 2007.

[19] T. Lidy and A. Rauber. Genre-oriented Organisation of Music Collections using the SOMeJB System: An Analysis of Rhythm Patterns and Other Features. In *Proc. of the DELOS Workshop on Multimedia Contents in Digital Libraries*, 2003.

[20] A.S. Lillie. Musicbox: Navigating the space of your music. Master's thesis, Massachussets Institute of Technology, 2008.

[21] M. Magas and P. Proutskova. A location-tracking interface for ethnomusicological collections. *JNMR Special Issue on Computational Musicology*, 2013. NNMR-2012-0044.R1.

[22] M. Magas, R. Stewart, and B. Fields. B.. decibel 151: Collaborative spatial audio interactive environment. In *ACM SIGGRAPH*, 2009.

[23] M. Magas, M.and Casey and C. Rhodes. mHashup: fast visual music discovery via locality sensitive hashing. In *ACM SIGGRAPH 2008 New Tech Demos*, 2008.

[24] J. McCarthy and P. Wright. Technology as experience. *Interactions*, 1(5):42–43, 2004.

[25] F. Guédy N. Rasamimanana N. Schnell, F. Bevilacqua. *Gemessene Interpretation - Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen, Klang and Begriff*, volume 4, chapter Playing and Replaying âĂŞ Sound, Gesture and Music Analysis and Re-Synthesis for the Interactive Control and Re-Embodiment of Recorded Music. Schott Verlag, Mainz, 2011.

[26] S. R. Ness and G. Tzanetakis. Audioscapes: Exploring surface interfaces for music exploration. In *Proceedings of the International Computer Music Conference (ICMC 2009)*, Montreal, Canada, 2009.

[27] D. Norman. *Emotional Design: Why We Love (Or Hate) Everyday Things*. Basic Books, 2004.

[28] E. Pampalk. Islands of Music: Analysis, Organization, and Visualization of Music Archives. *Journal of the Austrian Soc. for Artificial Intelligence*, 24(4):20–23, 2003.

[29] Recht B. Patten, J. and H. Ishii. Audiopad: A tag-based interface for musical performance. In *Proceedings of the 2002 conference on New interfaces for musical expression*, pages 1–6, 2002.

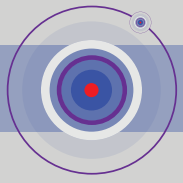
[30] P. Sengers and B. Gaver. Staying open to interpretation: Engaging multiple meanings in design and evaluation. In *Proceedings of the 6th conference on Designing Interactive Systems*, 2006.

[31] Stavness, J. Gluck, L. Vilhan, and S. Fels. The MUSICtable: A map-based ubiquitous system for social interaction with a digital music collection. Technical report, Lecture Notes In Computer Science, 2005. 3711:291.

[32] Sebastian Stober and Andreas Nürnberger. MusicGalaxy: A multi-focus zoomable interface for multi-facet exploration of music collections. In Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet, and Kristoffer Jensen, editors, *Exploring Music Contents*, volume 6684 of *LNCS*, pages 273–302, Berlin / Heidelberg, 2011. Springer Verlag. extended paper for post-proceedings of 7th International Symposium on Computer Music Modeling and Retrieval (CMMR'10).

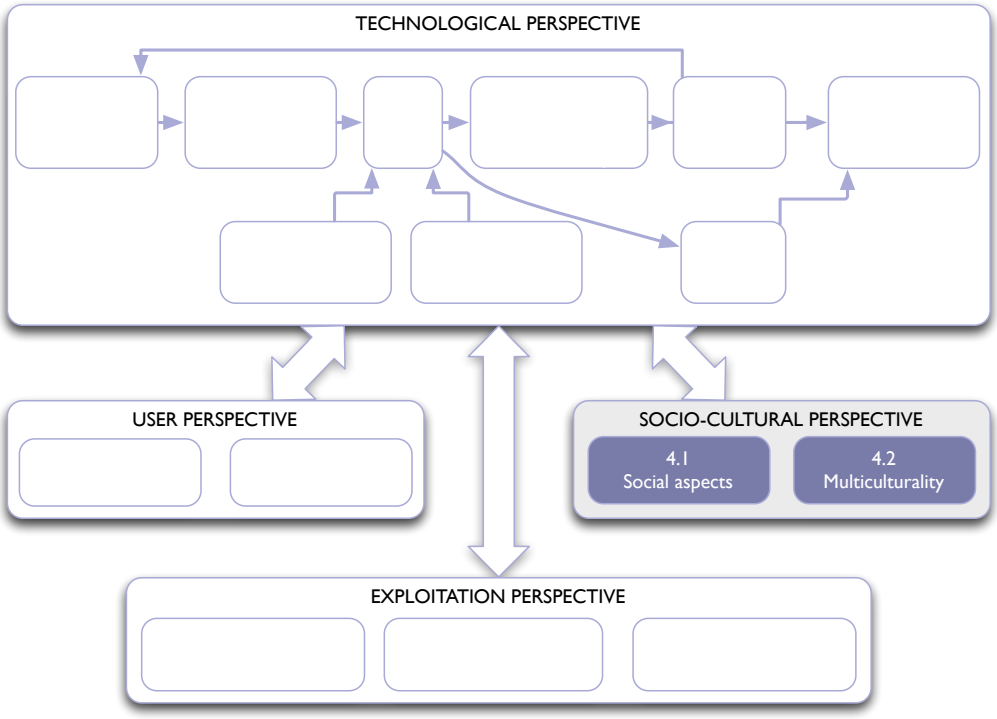
[33] B. Ullmer and H. Ishii. *Human-Computer Interaction in the New Millenium*, chapter Emerging Frameworks for Tangible User Interfaces. Addison-Wesley, Victoria, Canada, 2001. 579-601.1–382.

# SOCIO-CULTURAL PERSPECTIVE



CHAPTER 4
# Socio-cultural perspective

*Music Information Research involves the understanding and modeling of music-related data in its full contextual complexity. Music is a communication phenomenon that involves people and communities immersed in specific social and cultural context. MIR aims at processing musical data that captures the social and cultural context and at developing data processing methodologies with which to model the whole musical phenomenon.*



## 4.1 MUSIC-RELATED COLLECTIVE INFLUENCES, TRENDS AND BEHAVIORS

Music is a social phenomenon, thus its understanding and modeling requires the inclusion of this dimension. Social interaction is a driving force of music listening, categorisation, preference, purchasing behavior, etc. Additionally, teams or crowds are usually able to achieve feats that go beyond what individuals accomplish, and this is especially relevant for annotation and other collaborative scenarios. Finally, scattered in different virtual places, formats and time-scales, there is much data available that contains implicit information about music-related social factors, which could make possible the understanding and prediction of trends and other collective behaviours related to music. To carry out this research, which would complement the other, more traditional, approaches to music description, we need to involve people working in Social Computing, Sociologists and experts in Dynamic Systems and Complex Networks. Social computing will help to gather massive annotations and obtain knowledge on the key actors and factors of collective-mediated processes in musical choice, interaction and conceptualisation. Human dynamics will make possible massive-scale predictions about trends, ways and moments to listen to music, and provide pointers to the best locations and conditions for commercial and marketing activities (e.g., Buzzdeck[1]). The main obstacle to promising advances is the scarcity of open data and the privacy issues associated with access to data. Contrastingly, there are also issues to be solved when managing and analysing extremely large amounts of data of this kind.

### 4.1.1 State of the art

Even though most of the XXth Century technologies have made possible different modes of experiencing music individually, if we consider all the cultures in the world, music is still mostly experienced and valued in a social context. Even in the Western culture individual listening becomes a social activity as the experience is frequently, afterwards or simultaneously, interpreted and shared with other people. Hence, the value of music as social mediator and the social dynamics it makes possible have yet to be properly addressed by researchers. In addition to a traditional view corresponding to the social psychology of music/sociology of music (see section 2.4) we consider two research perspectives on social aspects of music: human dynamics and social computing.

How has MIR addressed, supported or capitalised on the social aspects of music? What is still to be done? As an orientation, the word "social" can be found in more or less 100 papers presented in the past 12 ISMIR editions but in most of the cases it is just a passing word, or part of a somewhat shallow expression like "social tags" or "social networks". In the bunch of papers that really deal with social aspects, social psychology and social computing are dominant perspectives, whereas human dynamics has been, up to now, absent.

Most basic research on social aspects of music has focused on individuals with relation to significant groups (i.e., peers, family, gang, nation), as we have summarised above. Alternatively, social behavior can be considered globally, nearly getting rid of the individual (we cannot avoid the link to Asimov's "psychohistory"[2], like researchers on social animals (especially insects) usually do. A global understanding of the flow patterns of spread, influence and consumption/enjoyment of a specific musical agent or content calls for new techniques such as complex network analysis or human dynamics [1]. Our knowledge of the interplay between individual activity and social network is limited, partly due to the difficulty in collecting large-scale data that record, simultaneously, dynamical traces of individual behaviors, their contexts and social interactions. This situation is changing rapidly, however, thanks to the pervasive use of mobile phones and portable computing devices. Indeed, the records of mobile communications collected by telecommunication carriers provide extensive proxy of individual symbolic and physical behaviors and social relationships. The high penetration of mobile phones implies that such data captures a large fraction of the population of an entire country. The availability of these massive CDRs (Call Detail Records) has made possible, for instance, the empirical validation in a large-scale setting of traditional social network hypotheses [18]. Taking advantage of them for music-related purposes is still pending because massive geo-temporally tagged data is still one of the bottlenecks for MIR researchers. We are still lacking knowledge

[1] http://buzzdeck.com
[2] http://en.wikipedia.org/wiki/Psychohistory_(fictional)

about listening patterns and how they are modulated by interaction with peers, by sharing of musical information with peers, or by geographical and environmental conditions (e.g., weather, time of the day) [6]. In order to study massive concurrent behavior patterns we only have available a large dataset of last.fm scrobblings harvested by òscar Celma [3]. It is interesting to note that the most recent "Million Song Dataset" [4] does not include any geo-temporal information. Telecommunication service companies should then be targeted by researchers and research project managers in order to make some progress along this line.

The social computing view, on the other hand, addresses either the creation of social conventions and context by means of technology (i.e., wikis, bookmarking, networking services, blogs), or the creation of data, information and knowledge in a collective and collaborative way (e.g., by means of collaborative filtering, reputation assignment systems, tagging [8], game playing [17], collaborative music creation tools, etc.). It is usually assumed that social computation, sometimes also called social information processing, will be more effective and efficient than individual or disconnected efforts [16]. When information is created socially, it is not independent of people, but rather is significant precisely because it links to people, who are in turn associated with other people [3]. Games with a purpose (GWAP) are a paradigmatic example of social computation for annotation of different knowledge domains. Major Miner [5], The Listen Game, TagATune [6], MagnaTagATune [7] [9], Moodswings [7], Mooso, HerdIt [2], etc., have been successfully used for gathering massive ground-truth "annotations" of music excerpts or for generating data about music preference or relatedness (see above section Collecting music related data). A further step in generating knowledge consists in building ontologies from tagging and writing behavior inside a delimited social network [10]; [12]. A unified model of social networks and semantics where social tagging systems can be modeled as a tripartite graph with actors, concepts and instances (e.g., songs or files) makes possible, by analysing the relations between concepts both on the basis of co-occurrence in instances and common usage by actors (users), the emergence of lightweight ontologies from online communities [11]. A completely different approach to community knowledge extraction for the design of ontologies is the implementation of Web portals with collaborative ontology management capabilities [19]. It has been recently reported on these strategies related to the Freesound community [4]. In addition to games and tag-related activity, collective musical knowledge can be generated by means of musical activity itself (and not just by tags or texts). Collective generation of playlists has been studied under different perspectives [14] [15]. Precisely in this category Turntable.fm [8] (unavailable in many European countries) is one of the recent successful musical apps for the iPhone (but see also Patent US7603352 [9], or just the collective playlist creation function as available in Spotify). Mashups [13] are another contemporary type of music content that benefits from music audio and context analysis technologies [5] although it is still pending to study how collective knowledge emerges inside communities that are focused on them. To conclude, a proper multidisciplinary forum to discuss music social computation would be the "International Conference on Social computing, behavioural modeling and prediction" [10] (held since 2008).

### 4.1.2 Specific Challenges

- **Promote formal techniques and methodologies for modeling music-related social and collective behavior**. Conference tutorials, keynotes and promotion of special sessions or workshops should be good vehicles for that.
- **Adopt and adapt complex networks and dynamic systems perspectives, techniques and tools.** Temporal sequences of descriptors can be considered as spanning a complex network. Semantic descriptors of a given file constitute networks too, so there are some opportunities to reframe existing research with network methodologies (e.g., diffusion models). In addition, decision processes about music items (in

[3] http://www.dtic.upf.edu/~ocelma/MusicRecommendationDataset/index.html
[4] http://labrosa.ee.columbia.edu/millionsong/lastfm
[5] http://majorminer.org/info/intro
[6] http://www.gwap.com/gwap/gamesPreview/tagatune
[7] http://tagatune.org/Magnatagatune.html
[8] http://blog.turntable.fm/
[9] http://www.google.com/patents/US7603352
[10] http://sbp.asu.edu

playlists or purchases, for example) can be addressed as specific cases of burst models.

- **Analyse interaction and activity in social music networks.** The roles, functions and activities of peers in digitally-mediated music recommendation and music engagement networks can be formally characterised by using specific analysis techniques. Trends, "infections" and influences in groups can be modelled mathematically and this can provide additional "contextual" information to understand activities related to music information.
- **Characterise the interplay between physical space, time, network structures and musical contents and context**. This requires a big data perspective where many disparate data sources and huge amounts of data can be integrated and mined (in some cases in real time). As some of these data are only available from business companies providing music, communication or geolocation services, strategic research coalitions with them have to be searched for.
- **Develop tools for social gaming and social engagement with music.** This will provide a "new" way to experience music and to create new knowledge and awareness about it. Sharing our music learning and experiencing processes may make them more robust and effective. Can we make typical teenage awe for music last until the very end of our lives by taking advantage of engaging activities with family, friends and colleagues? Can we revert the 20th Century trend of making music listening an isolationist activity?
- **Develop technologies for collective music-behaviour self-awareness.** Collective and simultaneous awareness/sensing is the target here. Personal tools for self-quantification are to be used to track and evidence collective synchronicities (e.g. entrainment, synchronous listening from remote places, sharing mood in a concert). It is easy to see that we, as members of a multitude, are clapping or rocking at the same time, and this has the ability to modify our external and internal states. In order to intensify such modifications we could use other signals than open behavior, and more contexts than music concerts (e.g., games, tweeting, blogging, listening to music . . . ).

## References


[1] A.L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435:207–211, 2005.

[2] Luke Barrington, Damien O'Malley, Douglas Turnbull, and Gert Lanckriet. User-centered design of a social game to tag music. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09, pages 7–10, New York, NY, USA, 2009. ACM.

[3] Thomas Erickson. *Social Computing. In: Soegaard, Mads and Dam, Rikke Friis (eds.)"Encyclopedia of Human-Computer Interaction".* The Interaction-Design.org Foundation, Aarhus, Denmark, 2011.

[4] F. Font, G. Roma, P. Herrera, and X. Serra. Characterization of the Freesound online community. In *Third International Workshop on Cognitive Information Processing*, 2012.

[5] G. Griffin, Y. E. Kim, and D. Turnbull. Beat-syncmash-coder: A web application for real-time creation of beat-synchronous music mashups. In *Proc. of the IEEE Conf. on Acoustics, Speech, and Signal Processing*, 2010.

[6] P. Herrera, Z. Resa, and M. Sordo. Rocking around the clock eight days a week: an exploration of temporal patterns of music listening. In *First Workshop On Music Recommendation And Discovery (WOMRAD), ACM RecSys*, Barcelona, Spain, 2010.

[7] Y. E. Kim, E. Schimdt, and L. Emelle. Moodswings: a collaborative game for music mood label collection. In *Proceedings of the 2008 International Conference on Music Information Retrieval*, Philadelphia, PA, 2008. ISMIR.

[8] P. Lamere. Social tagging and music information retrieval. *Journal of New Music Research: Special Issue: From Genres to Tags: Music Information Retrieval in the Age of Social Tagging*, 37(2):101–114, 2008.

[9] E. Law, K. West, M. Mandel, M. Bay, and J. S. Downie. Evaluation of algorithms using games: the case of music tagging. In *Proc. ISMIR 2009*, pages 387–392, 2009.

[10] Mark Levy and Mark Sandler. A semantic space for music derived from social tags. In *Proc. of International Conference on Music Information Retrieval*, pages 411–416, Vienna, Austria, 2007 2007.

[11] P. Mika. Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5(1):5–15, 2007.

[12] J. Z. Pan, S. Taylor, and E. Thomas. MusicMash2: Mashing Linked Music Data via An OWL DL Web Ontology. In *Proceedings of the WebSci'09: Society On-Line*, Athens, Greece, March 2009.

[13] A. Sinnreich. *Mashed Up: Music, Technology, and the Rise of Configurable Culture*. University of Massachusetts, 2010.

[14] David Sprague, Fuqu Wu, and Melanie Tory. Music selection using the partyvote democratic jukebox. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '08, pages 433–436, New York, NY, USA, 2008. ACM.

[15] S. Stumpf and S. Muscroft. When users generate music playlists: When words leave off, music begins? In *Proc. ICME*, pages 1–6, 2011.

[16] J. Surowiecki. *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business economies societies and nations.* New York: Doubleday, 2004.

[17] L. von Ahn. Games with a purpose. *IEEE Computer Magazine*, 39(6):92–94, 2006.

[18] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.L. Barabási. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, doi:10.1145/2020408.2020581, pages 1100–1108, 2011.

[19] A. V. Zhdanova. Community-driven ontology construction in social networking portals. *Web Intelligence and Agent Systems: An International Journal*, 6:93–121, 2006.

## 4.2 MULTICULTURALITY

Most music makes very little sense unless we experience it in its proper cultural context, thus the processing of music information has to take into account this cultural context. Most of MIR has focused on the mainstream popular Western music of the past few decades and thus most research results and technologies have a cultural bias towards that particular cultural context. The challenge is to open up our view on music, to develop technologies that take into account the existing musical diversity and thus the diverse musical cultural contexts. To approach the multicultural aspects of MIR there is a need to involve researchers from both engineering disciplines (Signal Processing, Machine Learning) and humanities (Musicology, Cultural Studies), and to involve people belonging to the specific cultures being studied. This approach will offer the possibility to identify new MIR problems and methodologies that could impact the whole MIR field. At the same time the development of Information Technologies that reflect diversity should help preserve the cultural richness of our world, which is threatened by the globalisation and homogenisation of the IT infrastructures. This is a topic that has started to be addressed by the MIR community but that will require much bigger efforts, not just by the research community but by political and industrial bodies.

### 4.2.1 State of the art

MIR uses a variety of methodologies, but the most common approximations are based on using signal processing and machine learning methods that treat musical data as any other machine readable data, thus without much domain knowledge. On the other hand the research done within the fields of Computational Musicology and Computational Ethnomusicology puts a special emphasis on the musical and cultural aspects, thus incorporating domain knowledge that we want to emphasise here. These two research areas have been growing in the last few years and their influence in the MIR community has been increasing (see section 2.4.1).

The term Computational Musicology comes from the research tradition of musicology, a field that has focused on the study of the symbolic music representations (scores) of the classical western music tradition [2]. This research perspective takes advantage of the availability of scores in machine-readable format and of all the musicological research that has been done on this music tradition. Music theoretical models are very much followed and current research focuses on the understanding and modeling of different musical facets such as melody, harmony or structure of western classical music. This research can be followed in the yearly journal Computing in Musicology [16]. From these references it can be observed that this field has been opening up, approaching other types of music, such as popular western music or different world music traditions, and it has started to use other types of data sources, such as audio recording.

In [15] the concept of Computational Ethnomusicology was introduced to refer to the use of computer tools to assist in ethnomusicological research. The emphasis is on the study of folk and popular music traditions that are outside the western classical and pop cultures, thus cultures that tend to be based on oral traditions and that have been mainly studied through audio recordings. Since the article was published there has been an increasing number of research articles related to Computational Ethnomusicology. For instance, according to [4], the percentage of papers on this area at the annual ISMIR conference increased from 4.8% in 2002 to 8.1% in 2008. A year later, in 2009, ISMIR hosted an oral session devoted to the analysis of folk music, sociology and ethnomusicology. After this event, a group of researchers working on MIR and ethnomusicology started the EthnoMIR discussion group which has organised a yearly workshop on Folk Music Analysis (2011 in Athens, 2012 in Seville, and 2013 in Amsterdam) with the purpose of gathering researchers who work in the area of computational music analysis of music from different cultures, using symbolic or signal processing methods, to present their work, discuss and exchange views on the topic. At the ISMIR 2011 there was a session dedicated to "non-western music" and at ISMIR 2012 there was a session on "musical cultures" and a larger than ever amount of contributions related to different musical traditions. In 2011 the European Research Council funded a project entitled "CompMusic: Computational Models for the discovery of the world's music" [11] that is studying five art

[11] http://compmusic.upf.edu

music traditions (Hindustani, Carnatic, Turkish-makam, Andalusi, and Beijing Opera) from an MIR perspective.

Recent studies on non-western music show the need to expand or even rethink some of the MIR methodologies. Some papers deal with specific musical facets, such as timbre/instrumentation (e.g. [14]), rhythm (e.g. [7]), motives (e.g. [10] [3]), tuning and scale (e.g. [5] [11]), melody (e.g. [17] [12]), or performance variations (e.g. [13] [6]). From these references it becomes clear that many of the musical concepts used in MIR need to be rethought and new approaches developed if we want to take a multicultural perspective. Concepts like tuning, rhythm, melody, scale, chord, tonic, . . . are very culture specific, and need to be treated as such. Among the non-western music repertoires that have been most studied from this perspective are the Turkish-makam (e.g. [1]) and the art traditions of India (e.g. [9] [8]).

### 4.2.2 Specific Challenges

- **Identify and characterise music cultures that can be studied from a data driven perspective.** For MIR purposes a musical culture can be considered a combination of a user community plus a musical repertoire that can be characterised computationally. Thus we can extract data from the trace left online by a user community, such as online social networks, and from different music data repositories created by that community, especially audio and scores. With this type of data we can then study quite a few aspects of a given musical culture. The challenge is to identify musical cultures that can be studied like this.
- **Gather and make available culturally relevant data for different music cultures.** Gather different data sources (audio, audio descriptors, editorial metadata, expert data, user commentaries, ...) with which to study and characterise the community+repertoire of the selected cultures. This data has to be made available to the research community.
- **Identify specific music characteristics for each culture.** Identify particular semantic music concepts and characteristics that are specific to each culture. These should be the aspects that allow us to differentiate the different musical cultures.
- **Develop methodologies for culture specific problems.** Develop knowledge based data processing approaches that can take advantage of the specificities of each culture, thus modeling the characteristics of each user community and music repertoire.
- **Develop specific applications of relevance for each cultural context.** The members of each user community might have specific needs and thus the applications to be developed for them should target their context and interests.
- **Carry out comparative studies using computational approaches.** These comparative studies should be done from the research results obtained in the characterisation and modeling of specific music traditions and repertoires.

## References


[1] Bariş Bozkurt. An Automatic Pitch Analysis Method for Turkish Maqam Music. *Journal of New Music Research*, 37(1):1–13, March 2008.

[2] Lelio Camilleri. Computational Musicology A Survey on Methodologies and Applications. *Revue Informatique et Statistique dans les Sciences humaines*, pages 51–65, 1993.

[3] Darrell Conklin and Christina Anagnostopoulou. Comparative Pattern Analysis of Cretan Folk Songs. *Journal of New Music Research*, 40(2):119–125, 2010.

[4] Olmo Cornelis, Micheline Lesaffre, Dirk Moelants, and Marc Leman. Access to Ethnic Music: Advances and Perspectives in Content-Based Music Information Retrieval. *Signal Processing*, 90(4):1008–1031, 2010.

[5] Ali Gedik and Bariş Bozkurt. Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music. *Journal of New Music Research*, 38(2):103–116, June 2009.

[6] L. Henbing and Marc Leman. A Gesture-Based Typology of Sliding-Tones in Guqin Music. *Journal of New Music Research*, pages 61–82, 2007.

[7] Andre Holzapfel and Yannis Stylianou. Scale Transform in Rhythmic Similarity of Music. *Audio, Speech and Language Processing, IEEE Transactions on*, 19(1):176–185, January 2011.

[8] Gopala Krishna Koduri, Sankalp Gulati, Preeti Rao, and Xavier Serra. Rãga Recognition based on Pitch Distribution Methods. *Journal of New Music Research*, 41(4):337–350, 2012.

[9] Arvindh Krishnaswamy. Melodic Atoms for Transcribing Carnatic Music. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 1–4, Barcelona (Spain), 2004.

[10] Olivier Lartillot and Mondher Ayari. Motivic Pattern Extraction in Music, and Application to the Study of Tunisian Modal Music. *South African Computer Journal*, 36:16–28, 2006.

[11] Dirk Moelants, Olmo Cornelis, and Marc Leman. Exploring African tone scales. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 489–494, Kobe, Japan, 2009.

[12] Joaquín Mora, Francisco Gómez, Emilia Gómez, Francisco Escobar-Borrego, and José Miguel Díaz-Báñez. Characterization and Melodic Similarity of A Cappella Flamenco Cantes ! In *Proc. of ISMIR (International Society for Music Information Retrieval)*, Utrecht, The Netherlands, 2010.

[13] Meinard Müller, Peter Grosche, and Frans Wiering. Automated analysis of performance variations in folk song recordings. In *Proceedings of the International Conference on Multimedia Information Retrieval - MIR '10*, New York, New York, USA, 2010. ACM Press.

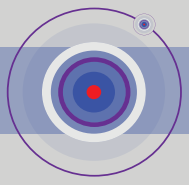
[14] Polina Proutskova and Michael Casey. You Call That Singing? Ensemble Classification for Multi-Cultural Collections of Music Recordings. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, number Ismir, pages 759–764, Kobe, Japan, 2009.

[15] George Tzanetakis, Ajay Kapur, W.A. Schloss, and Matthew Wright. Computational Ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2):1–24, 2007.

[16] Walter B Hewlett and Eleanor Selfridge-Field. Computing in Musicology. In *Center for Computer Assisted Research in the Humanities*. The MIT Press, 2006.

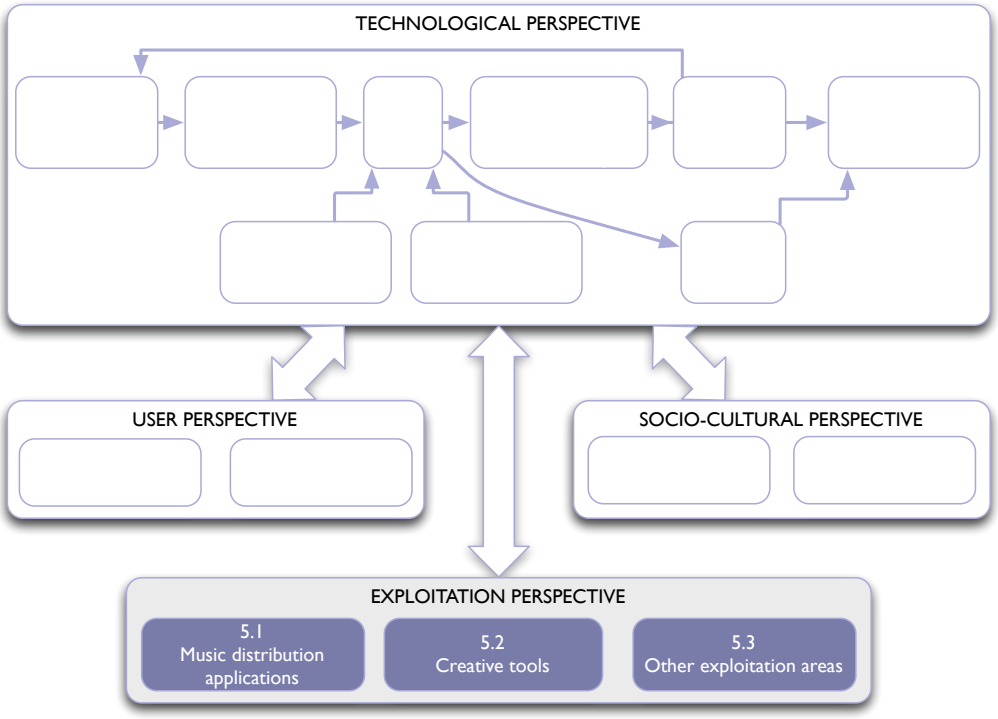
[17] Frans Wiering, Remco C Veltkamp, Jörg Garbers, Anja Volk, Peter van Kranenburg, and Louis P Grijp. Modelling Folksong Melodies. *Interdisciplinary Science Reviews*, 34(2):154–171, September 2009.

# EXPLOITATION PERSPECTIVE



CHAPTER 5
# Exploitation perspective

*Music Information Research is relevant for producing exploitable technologies for organising, discovering, retrieving, delivering, and tracking information related to music. These technologies should enable improved user experience and commercially viable applications and services for digital media stakeholders.*



## 5.1 MUSIC DISTRIBUTION APPLICATIONS

MIR is fundamental for developing technologies to be used in the music distribution ecosystem. The stakeholders in the music value chain are music services, record companies, performing rights organisations, music tech companies, music device and equipment manufacturers, and mobile carriers. The main challenge is to develop scalable technologies that are relevant to both the services that organise and distribute the music and also those services that track what is being distributed. These technologies span from music search and recommendation to audio identification both for recordings and compositions among others. By fully addressing the music distribution challenges, the MIR Community will establish closer ties with the industry which will help accessing resources (such as actual music data) and alternative ways of funding. On its side, the Music Distribution industry will have access to technologies more targeted to actual end-user scenarios which will give them an edge in the global market. Incidentally, it will help reducing innovation cycles from research to development and exploitation which, in turn, will have a clear impact on competitiveness and hence music distribution companies' profitability.

### 5.1.1 State of the art

A number of topics on the future of electronic music distribution have been addressed. This includes music search and discovery of music catalogues, the music rights industry-related technologies and other more transversal topics such as scalability and metadata cleaning.

As could be witnessed over the last few years, music is being produced and published at a faster rate than ever before: estimates range form yearly 11,000 (nonclassical) major label albums averaging some ten songs per album ([16], p. 261) up to 97,751 albums released in the United States in 2009, as reported by Nielsen SoundScan [1]. In the physical world, record shops were de-facto intermediaries that preselected music due to the physical constraints of storing music records and cd's. Digital technologies have changed this situation in at least two respects: digital music distribution channels such as iTunes, Amazon or Spotify can provide quick access to millions of music pieces at very low cost, hence they are less strictly preselected, and, with the abandonment of physical records, they shifted granularity from albums to single tracks, making it even harder for potential customers to make a choice. To fill this gap of missing preselections, automatic music recommendation systems supporting search and discovery have been developed attempting to provide an improved and manageable access to the music of the world.

Amazon [2] suggests albums or songs based on what has been purchased in the same order or by the same customers as items one searched for or bought. This is a form of collaborative filtering [8], which assumes that users who have agreed in the past (in their purchase decisions) will also agree in the future (by purchasing the same items). Collaborative filtering generally suffers from two related problems: the coldstart problem and the popularity bias. The coldstart problem is the fact that albums that have not yet been purchased by anybody can never be suggested. The popularity bias is the problem that for any given item, popular albums are more likely to have been purchased in conjunction with it than unpopular ones, and so have a better chance of being recommended. In consequence, collaborative filtering alone is incapable of suggesting new music releases. An additional problem specific to Amazon is that users may purchase items for somebody else (e.g., as a present), which might flaw the recommendations generated both for them and for other users of allegedly the same taste. Spotify [3], a music streaming service, bases its recommendations [4] [5] on its users' listening behavior, analysing which artists are often played by the same listeners. While this may potentially result in better suggestions than analysing sparse data such as record purchases, it is again subject to the cold-start problem and popularity bias. Furthermore, Spotify only recommends related artists and not songs, which is rather unspecific. Genius is a

[1] http://www.businesswire.com/news/home/20100106007077/en/2009-U.S.-Music-Purchases-2.1-2008-Music
[2] http://www.amazon.com
[3] http://www.spotify.com
[4] http://vimeo.com/57900625
[5] http://www.slideshare.net/erikbern/collaborative-filtering-at-spotify-16182818

function in Apple iTunes [6] which generates playlists and song recommendations by comparing music libraries, purchase histories and playlists of all its users, possibly integrating external sources of information. Assuming such external information does not play a major role, this system is again based mainly on collaborative filtering. Last.fm [7] combines information obtained from users' listening behavior and user-supplied tags (words or short expressions describing a song or artist). Tags can help to make recommendations transparent to users, e.g. a user listening to a love song may be recommended other tracks that have frequently been tagged as 'slow' and 'romantic'. But they are also inherently erroneous due to the lack of carefulness of some users, and require a range of counter measures for data cleaning. Tags are also affected by the cold-start problem and popularity bias. Pandora [8], another music streaming service, recommends songs from its catalogue based on expert reviews of tracks with respect to a few hundred genre-specific criteria. This allows for very accurate suggestions of songs that sound similar to what a user listens to, including sophisticated explanations for why a song was suggested (e.g., a track may be recommended because it is in a 'major key', features 'acoustic rhythm guitars', 'a subtle use of vocal harmony' and exhibits 'punk influences'). Such expert reviews incur high costs in terms of time and money which makes it impossible to extend the catalog at a rate that can keep up with new releases. This has a limiting effect on the selection of music available to users.

Most approaches described so far rely on some form of meta-information: user's listening or purchasing behavior, statistics about artists and genres in music collections, user defined tags etc. Another option is to actually analyse the audio content trying to model what is important for the perceived similarity between songs: instrumentation, tempo, rhythm, melody, harmony, etc. While many research prototypes of recommendation systems that use content-based audio similarity have been described in the literature (e.g., [12], [11], [10], [9], to name just a few), very little has been reported about successful adoption of such approaches- without combination with other methods- to real-life scenarios. Content based recommendation is used to some extent by a number of music companies like Mufin [9], echonest [10] or BMAT [11] amongst others. An exhaustive view on Music Recommendation systems can be found at [5].

In a landscape where the music industry is facing difficult times with income from physical sales shrinking, the music rights revenues are increasing worldwide. According to [12] the author's society royalty collections were 7.5€ billion in 2010 (climbing a 5,5% year-on-year) and [1] announced that the global performance rights reached the 905 US$ millions in 2011 (an increase of 4,9% from the previous year). These positive numbers are due to the increase of the number of media paying royalties and an improvement of the collecting methods of these societies. Hence, it is important to address the needs of the music rights business; i.e. the process of paying the owners of these rights (authors, performers, labels. . . ) for the usage of the music they have created and performed [13].

The rights organisations get most of their revenue not only from television, radio stations and those industries whose services are based on music, like clubs or venues, but also from a lot of other companies and associations from shops or dentists to school plays, basically anyone who aims at using somebody else's music creation [14]. In recent years, the music rights revenues coming from the digital world have also grown in importance [15]. All this rights money is collected through the royalty collection societies, which are divided in three kinds depending on the rights they represent: Authors, Performance and Master. Most of authors' societies worldwide are associated with the CISAC [16] while the master societies are associated with the IFPI [17]. The societies collect music rights and distribute them among their associates. At this point, a lot of controversy arises due to the different processes

[6] http://www.apple.com/itunes
[7] http://last.fm
[8] http://www.pandora.com
[9] http://www.mufin.com
[10] http://echonest.com
[11] http://www.bmat.com
[12] http://www.cisac.org/CisacPortal/initConsultDoc.do?idDoc=22994#pressrelease
[13] http://ascap.com/licensing/licensingfaq.aspx
[14] http://www.bmi.com/creators/royalty/how_we_pay_royalties/detail
[15] http://mediadecoder.blogs.nytimes.com/2012/01/17/digital-notes-royalties-for-streaming-music-grew-17-in-2011
[16] http://www.cisac.org
[17] http://www.ifpi.org

they use for such distribution and questions are raised about how to make this process as fair as possible [18].

Ideally, every right owner should be paid for the use of their music but in practice it is difficult and expensive to control all the media and all potential venues where music could eventually be used. The solutions that have been found vary depending on the country, the society and the type of source. Some years ago, the societies used to distribute based on the results of the top selling charts which created huge inequalities between artists. Later some other systems and technologies appeared:

- Cue sheets: Media companies are obliged to fill cue sheets, the list of music broadcast, explaining their use. However, this tends to be inaccurate because, while generating the cue sheets represents lots of work, media companies don't benefit from the accuracy of those [19].
- Watermarking: It consists in embedding an extra signal into a digital music work so this signal can be detected when the work is reproduced. Watermarking requires the use of watermarked audio references when broadcasting which is very rare. Also, the extra signal can easily be removed from original audio. [20] Fingerprinting: It consists in an algorithm that extracts the main features of an audio piece making a so-called fingerprint of the track. This fingerprint may easily be matched against an audio database which may comprise recordings from television, radio or internet radio broadcasts. [21] [4] [17]
- Clubs: The collecting societies track music played in all types of venues by sending a specialist who recognises music and writes down a cue sheet or by installing recording stations in Dj boards. [22]
- Online: Some of the most used music channels on the Internet as streaming or peer-to-peer services are extremely difficult to monitor. Nowadays the music monitored online is based on crawling millions of webs pages to detect their music usage. [23]
- Social Networks: A particular case of online music tracking is finding phylogenetic relationships between music objects spread on social networks. The type of relationships may include: "is the same song as", "contains snippet of", "includes", "remixes", "similar", "are the same song with different durations", "is the live version of", "is a cover version", "is a radio edit of" and so forth. This hasn't been addressed by MIR but is documented in other fields. [7] [6]
- Music vs Non Music discrimination: TV channels have normally blanket fees contracts with performing rights organisations according to which they pay royalties proportionally to the percentage of music broadcast. As this data is usually inaccurate, the PROs tend to outsource statistical estimation of this percentage which is also inaccurate. Although there has been quite some research on speech/music discrimination [13][14], generic music vs non music discrimination- robust to speech overlap- is a challenge for the industry.

While the research and engineering problems of simple audio identification use cases have practically been solved; for other real industry use cases, such as background music detection (over voice), in noisy backgrounds and with edited music, there are no robust technical solutions. In this business niche, a number of players share the market: Tunesat [24] in the USA, BMAT [25] in Spain, kollector [26] in Europe, Monitec [27] in Southamerica, Soundmouse [28] in the UK and yacast [29] in France.

[18] http://www.younison.eu/downloads/get/23
[19] http://www.editorsguild.com/v2/magazine/Newsletter/SepOct03/sepoct03_music_cue_sheets.html
[20] http://www.musictrace.de/technologies/watermarking.en.htm
[21] http://www.musictrace.de/technologies/fingerprinting.en.htm
[22] http://www.bemuso.com/musicbiz/musicroyaltycollectionsocieties.html
[23] http://www.musicrow.com/2012/01/tunesat-debuts-exclusive-internet-monitoring-technology
[24] http://tunesat.com/
[25] http://www.bmat.com
[26] http://www.kollector.com
[27] http://www.monitec.com
[28] http://www.soundmouse.com/aboutus/about_us.html
[29] http://www.yacast.fr/fr/index.html

A major challenge a new technology must face when it is to be applied in viable commercial products is scalability; i.e. the ability of the technology to handle massive amounts of data and the ability to handle that data's eventual growth in a cost effective manner. The problem is twofold. Firstly, some techniques are simply neither deployed nor tested since it's computationally impossible due to the size of datasets. Secondly, assuming the technique is scalable from a non-functional point of view, applying it to multi-million datasets may reveal problems which were not obvious in the first place. Beyond the problem of handling "big data", granting research access to huge music-related datasets may generate beneficial by-products for the music information research world. First, in large collections, certain phenomena may become discernible and lead to novel discoveries. Secondly, a large dataset can be relatively comprehensive, encompassing various more specialised subsets. By having all subsets within a single universe, we can have standardised data fields, features, etc. Lastly, a big dataset available to academia greatly promotes the interchange of ideas and results leading to, yet again, novel discoveries. A good example here is the "Million Songs Dataset" [2], which contains user tags provided by Last.Fm.

Systems that are able to automatically recommend music (as described above) are one of the most commercially relevant outcomes from the MIR community. For such recommender systems it is especially important to be able to cope with very large - and growing - collections of music. The core technique driving automatic music recommendation systems is the modelling of music similarity which is one of the central notions of MIR. Proper modelling of music similarity is at the heart of every application allowing automatic organisation and processing of music databases. Scaling up sublinearly the computation of music similarity to the millions is therefore an essential concern of MIR. Scalable music recommendation systems have been the subject of a number of publications. Probably one of the first content-based music recommendation systems working on large collections (over 200.000 songs) was published by [3]. Although latest results (see e.g. [15]) enable systems to answer music similarity queries in about half a second on a standard desktop CPU on a collection of 2.5 million music tracks yet, the system performs in a linear fashion.

The issue of scalability clearly also affects other areas of MIR: music identification meaning both pure fingerprinting technologies and cover detection, multimodal music recommendation and personalisation (using contextual and collaborative filtering Information).

### 5.1.2 Specific Challenges

- **Demonstrate better exploitation possibilities of MIR technologies.** The challenge is to convince stakeholders of the value of the technology provided by the MIR community and help them find new revenue streams from their digital assets which are additive and non-cannibalising to existing revenue channels. For these technologies to be relevant they should re-valorise the digital music product, help reduce piracy, streamline industry processes, and reduce inefficiencies.
- **Develop systems that go beyond recommendation, towards discovery.** Systems have to go beyond simple recommendation and playlisting by supporting discovery and novelty as opposed to predictability and familiarity. This should be one way of making our systems interesting and engaging for prospective users.
- **Develop music similarity methods for particular applications and contexts.** This means that results produced by computers have to be consistent with human experience of music similarity. Therefore it will be necessary to research methods of personalising our systems to individual users in particular contexts instead of providing one-for-all services.
- **Develop systems which cater to the scale of Big Data.** The data sets might be songs, users or any other music related elements. From a non-functional perspective, the algorithms and tools themselves should be fast enough to run with sublinear performance on very large datasets so they can easily enable solutions for streaming and subscription services. Beyond raw performance such as processing speed, from a functional view, the algorithms have to be designed to handle the organisation of large music catalogues and the relevance weighting of rapidly increasing quantities of music data mined from crowd-sourced tagging and

social networks. Applying algorithms to those big datasets may reveal problems and new research scenarios which were not obvious in the first place.

- **Develop large scale robust identification methods for recordings and works.** Performing rights organisations and record companies are shifting towards fingerprinting technologies for complete solutions for tracking their affiliates'/partners' music and for managing their music catalogues. While music fingerprinting has been around for years and it has been widely used, new use cases which require extensive R&D are arising: copyright enforcement for songs and compositions in noisy and live environments and music metadata autotagging among others. Also, finding phylogenetic relationships between songs/performances available on the web, such as "is a remix of" or "is the live version of", may unlock new application scenarios based on music object relationship graphs such as multimodal trust and influence metering in social networks.
- **Develop music metadata cleaning techniques.** One common feedback from all industry stakeholders such as record companies, music services, music distributor and PROs is the lack of so-called "clean music databases". The absence of clean music databases causes broken links between data from different systems and incorrect editorial metadata tagging for music recordings, which ultimately affects the perceived end-user quality of the applications and services relying on MIR technologies. We encourage the MIR community to address music metadata cleaning by using music analysis and fingerprinting methods as well as text-based techniques borrowed from neighbouring research fields such as text information retrieval and data management among others.
- **Develop music detection technology for broadcast audio streams.** The media industry is lacking the means for accurately detecting when music (including background music) has been broadcast, in order to transparently handle music royalty payments. This technology should go beyond music vs speech discrimination and address real life use cases such as properly discriminating music vs generic noise.

## References


[1] Recording industry in numbers – 2012 edition. In *Recording Industry in Numbers – 2012 Edition*, 2012.

[2] Thierry Bertin-Mahieux, Daniel PW Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In *Proc. of ISMIR (International Society for Music Information Retrieval)*, pages 591–596, Miami, Florida, USA, October 2011.

[3] P. Cano, M. Koppenberger, and N. Wack. An industrial-strength content-based music recommendation system. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 673–673. ACM, 2005.

[4] Pedro Cano. *Content-based audio search from fingerprinting to semantic audio retrieval.* Phd thesis, Pompeu Fabra University, Barcelona, Spain, 2007.

[5] Oscar Celma. *Music recommendation and discovery: The long tail, long fail, and long play in the digital music space.* Springer, 2010.

[6] Z. Dias, A. Rocha, and S. Goldenstein. Video phylogeny: Recovering near-duplicate video relationships. In *Information Forensics and Security (WIFS), 2011 IEEE International Workshop on*, pages 1–6. IEEE, 2011.

[7] Z. Dias, A. Rocha, and S. Goldenstein. Image phylogeny by minimal spanning trees. *Information Forensics and Security, IEEE Transactions on*, 7(2):774–788, 2012.

[8] J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 230–237. ACM, 1999.

[9] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer. A music search engine built upon audio-based and web-based similarity measures. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 447–454. ACM, 2007.

[10] P. Lamere and D. Eck. Using 3d visualizations to explore and discover music. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, pages 173–174, 2007.

[11] R. Neumayer, M. Dittenbach, and A. Rauber. Playsom and pocketsomplayer, alternative interfaces to large music collections. In *Proc. of ISMIR*, volume 5. Citeseer, 2005.

[12] E. Pampalk. Islands of music: Analysis, organization, and visualization of music archives. Master's thesis, Vienna University of Technology, Vienna, Austria, 2001.

[13] C. Panagiotakis and G. Tziritas. A speech/music discriminator based on RMS and zero-crossings. *Multimedia, IEEE Transactions on*, 7(1):155–166, 2005.

[14] E. Scheirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In *Proc. of IEEE ICASSP (International Conference on Acoustics, Speech, and Signal Processing)*, volume 2, pages 1331–1334. IEEE, 1997.

[15] D. Schnitzer, A. Flexer, and G. Widmer. A fast audio similarity retrieval method for millions of music tracks. *Multimedia Tools and Applications*, pages 1–18, 2012.

[16] H.L. Vogel. *Entertainment industry economics: a guide for financial analysis*. Cambridge University Press, 2010.

[17] A. Wang et al. An industrial strength audio search algorithm. In *Proc. Int. Conf. on Music Info. Retrieval ISMIR*, volume 3, 2003.

## 5.2 CREATIVE TOOLS

Creative practitioners produce, transform and reuse music materials. The MIR challenge is how to develop tools that process music information in a way that enhances creative production. Tools for automatically extracting relevant information from audio materials could be developed for purposes such as content-based manipulation, generativity, synchronisation with other media, or real-time processing. Moreover, the large volume of available data requires efficient data manipulation systems that enable new methods of manipulation for creative purposes. This challenge requires collaborative research between music information researchers and the actors, including artists, performers and creative industries professionals. The impact of this research is in generating more creative possibilities and enabling production efficiency in a variety of creative contexts including music performance, music and sound production and post-production, sound engineering applications for audiovisual production, art installations, creative marketing, mobile apps, gaming, commercial installations, environmental installations, indoor and outdoor events. Collaborative research between creative practitioners and music information researchers contributes to bridging the gap between the arts and the sciences, and introduces novel practices and methodologies. It extends the paradigm of user-centred research to creative-input research, where the feedback loop between creative practitioner and researcher is an iterative, knowledge-building process, supported by adaptive modelling of research environments and resulting in more versatile creative tools.

### 5.2.1 State of the art

Music Information Research offers multiple possibilities for supporting musical creation or for inspiring the creation of multimedia art pieces. The creative possibilities of MIR can be studied and classified according to many different criteria, such as *off-line tools* (for composition or editing) vs. *real-time tools* for interaction, which in turn can be divided into applications for live performance and for art installations, or categories of tools designed for *professional* vs. tools designed for *novice* end-users, the latter including all type of applications promoting different models of "active listening".

#### Content-based sound processing

Content-based sound processing consists in using high-level information of the audio signal in order to process it. It includes processes controlled by high-level parameters and processes based on the decomposition of the audio content, through operations such as segmentation, source separation and transcription, into elements that can be processed independently. The goal is to provide expert end-users (e.g. musicians, sound designers) with intuitive tools, controlled through parameters relevant from the viewpoint of human cognition of music and sound, and also to enhance the quality of existing processes by selecting appropriate processes and parameter sets according to the nature of the extracted elements. For instance, a better subjective quality for slowing down a sound by time-stretching is obtained if the transient parts are separated from the sustained ones and preserved in the time-scale change process [13].

*Sound editing:* Sound editing refers to offline tools using pre-recorded audio contents. Some commercial products have started to implement such features. These include Celemony's *Melodyne* [30] and Roland's *R-Mix* [31], which provide studio production tools for pitch recognition and correction, tempo and timing alteration and spectrum visualisation. IRCAM's *Audiosculpt* [32], targeting expert users, enables to compute various kinds of analyses (segmentation and beat tracking, pitch and spectral envelope analysis) and use them as inputs for high-quality audio processes. Apple's *GarageBand* [33] is a good example of content-based processing application aimed at mass-market end-users: it automatically processes the content of an *AppleLoop* imported into a sequence by adapting its tempo and pitch scale to the sequence musical context. Most existing tools efficiently implement content-based editing for monophonic signals, however they also demonstrate the current limitations of the

[30] http://www.celemony.com/cms/index.php?id=products_editor
[31] http://www.rolandconnect.com/product_2011-09.php?p=r-mix
[32] http://forumnet.ircam.fr/product/audiosculpt
[33] http://www.apple.com/support/garageband

state-of-the-art of the research on the analysis of polyphonic recordings. A significant advance in this direction is the integration of a polyphonic transcription (audio to MIDI) in the *Live 9* application by Ableton [34] issued early 2013.

*Computer-aided composition:* Software environments for computer-aided composition such as *OpenMusic* [35], CommonMusic [36] or PWGL [37] are not only used for computing instrumental scores from user-defined algorithms, but also for controlling various kinds of sound syntheses from symbolic music representations [1]. In these environments the availability of audio analysis modules extracting musical information in the form of symbolic structures enables composers to elaborate scores with parameters in relation to the content of input sound files, and also to control sound synthesis from processing the extracted information at the symbolic level. The unlimited computing possibilities of these music languages allow expert musicians to adapt all the available analysis parameters to a broad variety of aesthetic approaches.

### Use of audio databases for music and sound production

The advancement of audio database technologies enables new applications for sound and music production, not only for content-based management of audio samples but also for the development of new methods for sound synthesis and music composition.

*Content-based management of audio samples:* MIR techniques can be very convenient for finding suitable loops or sound files to fit a particular composition or mix. The *MuscleFish* [18] and *Studio Online* [17] systems developed at the end of the 90s were the very first applications of content-based search in audio sample databases that have been further elaborated in the CUIDADO European project [16]. More recently, the availability of large public and free sound databases and repositories such as *Freesound* [38] has become mainstream. Using repositories and APIs such as EchoNest's *Remix Python* API [39] or MTG's *Essentia* [40], developers and hackers are creating a panoply of imaginative remix applications, many of them being developed during *Music Hack Day* events, which lately appear to be a very productive place for MIR based creation. However, even though the use of large audio sample banks is now mainstream in music production, the existing products, such as Native Instrument's *Kontakt* [41], MOTU's *MachFive* [42] or *Vienna Symphonic Library* [43] are not yet exploiting the full potential of MIR technologies for the content-based management audio databases.

*Corpus-based synthesis and musaicing:* One of the most obvious MIR applications for real-time music creation is that of "concatenative synthesis" [8, 15], "musaicing" [19] or mashup. These three terms approximately relate to the same idea, creating new music by means of concatenating short fragments of sound or music recordings to "approximate" the sound of a target piece. More precisely, an existing music piece or musical fragment is substituted with small, similar sounding music fragments, leading to a similarly structured result. The duration of these sound units can vary depending on the techniques employed and the desired aesthetic results, but are roughly in the range of 10 milliseconds up to several seconds or several musical bars. While manual procedures could be used with longer fragments (i.e. of several seconds), the use of shorter fragments inevitably leads to automatised MIR analysis and recovery techniques, in which a "target" track or sound is analysed, its descriptors extracted for every small fragment, and these fragments substituted with the best candidates from a large database of sound snippets. When using a pre-analysed sound repository and a compact feature representation, these techniques can be efficiently applied in real-time. Janer and de Boer [7] describe a method for real-time voice-driven audio mosaicing synthesis. *BeatJockey* [10] is a system aimed at DJs, which integrates audio mosaicing, beat-tracking and machine learning techniques and brings them into the Reactable musical tabletop. Several commercial tools

[34] https://www.ableton.com/en/live/new-in-9
[35] http://forumnet.ircam.fr/product/openmusic
[36] http://commonmusic.sourceforge.net
[37] http://www2.siba.fi/PWGL
[38] http://www.freesound.org
[39] http://echonest.github.com/remix
[40] http://mtg.upf.edu/technologies/essentia
[41] http://www.native-instruments.com/#/en/products/producer/kontakt-5
[42] http://www.motu.com/products/software/machfive
[43] http://www.vsl.co.at

following this approach (such as Steinberg's *LoopmashVST* plugin and iOS app [44]) are also already available. These techniques bring new creative possibilities somewhere in between synthesis control and remixing, and open the path to radically novel control interfaces and interaction modalities for music performance.

*Computer-aided orchestration:* In comparison to other aspects of musical composition (harmony, rhythm, counterpoint), orchestration has a specific status: intended as the art of selecting and mixing individual instrument timbres to produce a given "colour", it relates more closely to the real experience of sound from an orchestra. The same chord can produce a very different timbre depending on the instruments selected for performing it, and, despite existing treatises providing recipes for specific cases, orchestration has generally remained an empirical art based on mostly unelicited rules. An original approach recently developed in the framework of computer-aided composition tools has been to concentrate on the approximation, in terms of sound similarity, of a given sound target from the combination of elementary note samples from a set of selected instruments, using multiobjective optimisation, for managing the combinatorial issue of search into sound sample databases of hundreds of thousands of items. This work has been already used for the composition of numerous contemporary music works and implemented as the *OrchidÃĹe* software [3]. One of its main limitations was however that it only considered the static properties of the source, and the latest advances in related research have been to design a new search algorithm, named *MultiObjective Time Series*, that efficiently computes similarity distances from the coding of the temporal evolution of multiple descriptors of audio samples so that the dynamic properties of the target and of the original sound samples are taken into account in the approximation [5].

## Live performance applications

The applications discussed in the previous parts mainly concern offline processes and composition tools. The process of music information generated in the context of live performance applications imposes specific constraints on the audio analysis algorithms, in terms of causality, latency, and implementation (computing power, distributed processing vs. real-time performance). Applications not only concern live music, but also theatre, dance and multimedia. The computer music community has produced numerous software environments dedicated to the programming and real-time scheduling of algorithms for audio and music information processing, including, among many others, *Max* [45], *Pd* [46], *SuperCollider* [47], and *Chuck* [48].

*Beat syncing* A broad use-case of automatic beat tracking algorithms is live mixing applications for DJs, such as *Native Instrument's Traktor* [49], that enable to manage the transition between tracks in a beat-synchronous way, using time-stretching for managing the tempo evolution between them.

*Improvisation and interaction using symbolic sequence models:* While musaicing or remixing applications mostly rely on low-level signal processing analysis, the following examples focus on musical knowledge and understanding. Assayag et al [2] describe a multi-agent architecture for an improvisation-oriented musician-machine interaction system that learns in real-time from human performers, and establishes improvisatory dialogues with the performers by recycling their own audio material using an Oracle Factor for coding the multiple relationships of music symbol subsequences. Recent applications of this model also include interactive arranging and voicing from a learned musical corpus. The *Wekinator* [6] is a real-time machine learning toolkit that can be used in the processes of music composition and performance, as well as to build new musical interfaces. Pachet is working with Constrained Markov Models (CMM) for studying musical style by analysing musicians, extracting relevant features and modelling them using CMM [12], an approach that allows systems to improvise in a given style or along with any other musicians.

*Score following and interactive accompaniment:* Numerous contemporary music works, named mixed works, rely on the combination of instrumental parts and electronic sounds produced by real-time synthesis or processing of the instrument sounds. Different strategies exist for synchronising those various parts live in concert, the

[44] http://www.steinberg.net/en/products/ios_apps/loopmash.html
[45] http://cycling74.com/products/max
[46] http://crca.ucsd.edu/~msp/Pd_documentation/index.htm
[47] http://supercollider.sourceforge.net
[48] http://chuck.cs.princeton.edu
[49] http://www.native-instruments.com/#/en/products/dj/traktor

most straightforward ones, but least musical, consisting in pre-recording a soundtrack and superimposing the performers to play with it. Conversely, score following aims to automatically synchronise computer actions with real-time analysis of performance and to compare them with an internal model of the performed score. The latest advances of research on this subject, implemented in the *Antescofo* application [50] include a continuous tempo tracking of the performance and the definition of a language for specifying the real-time processes [4]. Another use case of the same algorithms is interactive accompaniment or "music minus one", where a solo performer can train on a pre-recorded accompaniment sound track that follows his (her) tempo evolutions.

*Performance/sound interaction:* The NIME community [51] is very active in the design of new performance/sound interaction systems that extend the traditional notion of musical instruments. The main aspects of the field related to MIR technologies are presented in section 3 and will not be developed here.

### Art installations

Sound has featured extensively in art installations since Luigi Russolo's Futurist manifesto "The Art of Noises" described the sound of the urban industrial landscape and became the source of inspiration for many artists and composers (e.g. Edgard Varèse, John Cage and Pierre Schaefer) [14]. Recent music technologies offer increased opportunities for immersive sound art experiences and physical explorations of sound. Art installations offer novel ways of using these technologies and enable novel experiences particularly through placing the focus on the audience and their context.

*Environmental sound installations:* Art installations have increasingly been using data from field recordings or sounds generated through real-time location-based interaction in order to trigger various behaviours (e.g. *Sound Mapping London Tea Houses*[52]). Artists have explored the integration of unfamiliar sounds into new physical environments (e.g. Bill Fontana's *White Sound: An Urban Seascape*, 2011 [53]). Generative music has been used in response to the environment. For instance *Variable 4* [54] is an environmental algorithmic weather machine which generates a unique musical composition which reflects the changing atmosphere of that particular environment. *Radioactive Orchestra* [55] aims to produce a musical sequence from the radioactivity of nuclear isotopes.

*Collaborative sound art:* Collaborative music making has been expressed through art installations such as *Play.Orchestra* [56] which blurs the borders between audience and player, amateur and professional. Atau Tanaka's *Global String* [57] metaphorically wraps a musical string around the world and through user engagement creates a collaborative instrument between world art galleries exploring the idea of communication via non-linguistic musical interaction and collaboration.

*Body generative sound art:* Using the human body as an instigator in the generation of sound and music is a growing research area. Atau Tanaka's *Sensorband* [58] includes performers wearing a combination of MIDIconductor machines that send and receive ultrasound signals measuring the hands' rotational positions and relative distance; gestural interaction with invisible infrared beams; and the *BioMuse*, a system that tracks neural signals (EMG), translating electrical signals from the body into digital data. Since around 2006 Daito Manabe has been working on the project *Electric Stimulus* [59] which literally plays the body as a sensory network for outputting sound and expression through the application of shock waves to particular nerve centres. The *Serendiptichord Dance* [60] focuses on the act of performing physically with a circular instrument unifying performance and sound.

[50] http://forumnet.ircam.fr/product/antescofo
[51] http://www.nime.org
[52] Sound Mapping London Tea Houses was an installation at the Victoria and Albert Museum in 2011, by the G-Hack group from Queen Mary, University of London.
[53] http://www.resoundings.org
[54] http://www.variable4.org.uk/about/intro
[55] http://www.nuclear.kth.se/radioactiveorchestra
[56] http://www.milkandtales.com/playorchestra.htm
[57] http://www.ataut.net/site/Global-String
[58] http://www.ataut.net/site/Sensorband
[59] http://www.creativeapplications.net/maxmsp/electric-stimulus-maxmsp
[60] http://www.youtube.com/watch?v=lUuCQTgz4Tc

*Public art using MIR:* Few examples of public art installations have used MIR to enable physical presence to interact with music information. Since September 2011 Barcelona's City Council has installed an automatic water and lights choreographies generator for the *Magic Fountain* [61] of *Montjuic* (one of the main tourist attractions of the city), based on MIR techniques (more concretely on the *Essentia* engine [62]). This system allows the person in charge of creating a choreography for the fountain to pick up a musical mp3 track, decide among several high-level parameters' tendencies (such as the average intensity, contrast, speed of change, the amount of repetition, or the main colour tonalities of the desired choreography), and the system generates automatic, music-controlled choreographies at the push of a button. Another example, *decibel 151* [9] installed at SIGGRAPH 2009, turns users into "walking playlists" and encourages physical explorations of music. MIR systems can therefore offer novel art installation experiences and have a profound impact on the way we as human beings understand space, time and our own bodies. The arts are also uniquely placed, with a degree of freedom from the commercial sector, to offer test grounds for MIR research into gestural and environmental applications of music data.

## Commercial end-user applications

As Mark Mulligan states in his 2011 report "digital and social tools have already transformed the artist-fan relationship, but even greater change is coming. . . the scene is set for the Mass Customisation of music, heralding in the era of Agile Music" [11]. *Agile Music* is a framework for understanding how artist creativity, industry business models and music products must all undergo a programme of radical, transformational change. In this context MIR offers new opportunities for creative commercial installations, applications and environments (e.g. creative marketing tools, mobile apps, gaming, commercial installations, environmental installations, indoor and outdoor events).

*Social music applications:* The increased choice of music available to the users is currently being explored through creative applications engaging with social media (e.g. Coke Music *24 hr challenge* [63]). As one of the current market leaders in social recommendation, Spotify has enjoyed an economic growth of 1 million paying users in March 2011 to 3 million paying users by January 2012 [64], partly thanks to its integration with Facebook. The application *Serendip* [65] creates a real time social music radio allowing users the opportunity to independently choose 'DJs' from their followers and share songs across a range of social media via a seamless Twitter integration.

*Mobile music applications:* Application developers are using the range of sensory information available on mobile devices to create more immersive sonic experiences and music generators (e.g. the *Musicity* project [66]; *RjDj* [67]). There are many apps which allow smart devices, for example the iPhone, to be transformed into portable musical instruments which engage with the body and allow for spontaneous performances (e.g. *Reactable App* [68]; *Bloom* [69]). Together with the advent of the Internet-of-Things (IoT), the communication society is witnessing the generalisation of ubiquitous communication, the diversification of media (radio, TV, social media, etc.), the diversification of devices and respective software platforms, and APIs for communities of developers (e.g., iPhone, Android, PDAs, but also Arduinos, Open Hardware, sensors and electronic tags) and the multiplicity of modalities of interaction. This imposes a challenge of facilitating interoperability between devices and facilitating combinations between diverse modalities.

*Gaming music applications:* The musical interaction team at IRCAM has been working with motion sensors embedded within a ball to explore some of the concepts integrating fun, gaming and musical experience in the *Urban Musical Game* [70]. *Joust* [71] is a spatial musical gaming system using motion rhythm and pace as the instigator

[61] http://w3.bcn.es/V01/Serveis/Noticies/V01NoticiesLlistatNoticiesCtl/0,2138,1653_1802_2_1589325361,00.html
[62] http://mtg.upf.edu/technologies/essentia
[63] http://www.nexusinteractivearts.com/work/hellicar-lewis/coke-music-24hr-session-with-maroon-5
[64] http://www.guardian.co.uk/media/2012/jan/29/spotify-facebook-partnership-apps
[65] http://serendip.me/
[66] http://musicity.info/home/
[67] http://rjdj.me
[68] http://www.reactable.com/products/mobile
[69] http://www.generativemusic.com
[70] http://www.youtube.com/watch?v=jXGlvmrGBgY
[71] http://gutefabrik.com/joust.html

of the action (Innovation Award: Game Developers Choice Award 2012). Interactive and immersive musical environments have also been used as a way of making commercial products memorable and fun. By gamifying their services and producing interactive experiences, innovative companies are working to increase their products' core values (e.g. Volkswagen's *Fun Theory* [72] and Wrigleys *Augmented Reality Music Mixer* [73]). The *Echo Temple* at Virgin Mobile FreeFest [74] created a shared experience of making music through the use of motion tracking cameras and fans branded with special symbols. The use of gaming is another developing application for MIR with various research and commercial possibilities.

### 5.2.2 Specific Challenges

- **Develop methodologies to take advantage of MIR for artistic applications in close collaboration with creators.** New possibilities for music content manipulation resulting from MIR research have the power to transform music creation. The development of tools for artistic applications can only be done with the involvement of the creators in the whole research and development process.
- **Develop tools for sound processing based on high-level concepts**. New approaches in MIR-related research should provide musicians and sound designers with a high-level content-based processing of sound and music related data. This entails furthering the integration of relevant cognitive models and representations in the creative tools, and also enabling users to implement their own categories by providing them with a direct access to machine learning and automatic classification features.
- **Enable tools for direct manipulation of sound and musical content.** Significant enhancements are required in polyphonic audio analysis methods (e.g. audio-to-score, blind source separation) in order to build applications allowing content-based manipulation of sound. This is expected to have a major impact on professional and end-user markets.
- **Develop new computer languages for managing temporal processes.** This will provide more adapted creative tools, not only for music composition and performance, but more generally for temporal media and interactive multimedia.
- **Improve integration of audio database management systems in standard audio production tools.** These should combine online and offline access to audio materials, and feature content-based search.
- **Develop real-time MIR tools for performance.** Research real-time issues beyond the "faster search engines" in the use of MIR technologies for music performance, addressing the design of specific algorithms and of potential applications in their entirety, in collaboration with the NIME community.
- **Performance modeling and spatial dimensions.** The management of sound and music information in creative tools shall not be limited to basic music categories (such as pitch, intensity and timbre) and must integrate in particular the dimensions of performance modelling and sound space. Beyond direct sound/gesture mapping, the design of new electronic instruments requires a better understanding of the specific structures underlying gesture and performance and their relation to the sound content. As for the spatial dimension of sound, new research advances are expected in the automatic extraction of spatial features for mono- and multichannel recordings and the simulation of virtual acoustic scenes from high-level spatial descriptors, with applications in music production, audiovisual post-production, games and multimedia.
- **Develop MIR methods for soundscaping.** Immersive music environments and virtual soundscaping are growth areas in the creative industries, particularly in relation to physical spaces. Research may involve knowledge gained from collaborations with specialists in building acoustics, architects, and installation artists.

[72] http://www.thefuntheory.com
[73] http://5gum.fr
[74] http://great-ads.blogspot.co.uk/2011/09/interactive-sound-installation-for.html

- **Use artistic sound installation environments as MIR research test grounds.** Immersive discovery experiences and physical explorations of music presented as art installations can contribute to a better understanding of the user's Quality of Experience (QoE); the potential of using sound as an aid to narrative creation and as a non-linguistic means for communication; the use of the the human body as an instigator of the generation of sound; and active engagement of listeners with their environment.
- **Develop creative tools which include data useful to commerce.** Research areas uncovered by consulting commercial and industry practices may include e.g. sonic branding, personalisation, interactive media environments, social platforms, and marketing tools between artists and fans.
- **Improve data interoperability between devices** An effort is required towards the standardisation of data protocols for a pan-European exchange of music software and hardware modalities. This is especially relevant for music, which is a paradigmatic example of multimodal media, with active communities of developers, working with a rich diversity of devices.
- **Develop automatic playlist generation and automatic mixing tools for commercial environments.** Systems which deliver the appropriate atmosphere for purchase or entertainment require music information research in conjunction with consumer psychology. For example, high level descriptors may be developed to include relationships between music and certain types of product, and music psychology may include field work in commercial environments.

## References


[1] C. Agon, J. Bresson, and M. Stroppa. OMChroma: Compositional control of sound synthesis. *Computer Music Journal*, 35(2), 2011. Springer Verlag.

[2] Gérard Assayag, Georges Bloch, Marc Chemillier, Arshia Cont, and Shlomo Dub. OMax brothers: a dynamic topology of agents for improvization learning. In *ACM Multimedia/ 1st Workshop on Audio and Music Computing for Multimedia*, pages 125–132, Santa Barbara, California, 2006.

[3] G. Carpentier and J. Bresson. Interacting with symbolic, sound and feature spaces in orchidée, a computer-aided orchestration environment. *Computer Music Journal*, 34(1):10–27, 2010.

[4] A. Cont. A coupled duration-focused architecture for realtime music to score alignment. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 32(6), June 2010.

[5] P. Esling and C. Agon. Intelligent sound samples database with multi objective time series matching. *IEEE Transactions on Speech Audio and Language Processing*, 2013. To appear.

[6] R. Fiebrink, D. Trueman, and P.R. Cook. A meta-instrument for interactive, on-the-fly machine learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Pittsburgh, June 2006.

[7] J. Janer and M. De Boer. Extending voice-driven synthesis to audio mosaicing. In *5th Sound and Music Computing Conference, Berlin*, volume 4, 2008.

[8] E. Maestre, R. Ramírez, S. Kersten, and X. Serra. Expressive concatenative synthesis by reusing samples from real performance recordings. *Computer Music Journal*, 33(4):23–42, 2009.

[9] M. Magas, R. Stewart, and B. Fields. decibel 151: Collaborative spatial audio interactive environment. In *ACM SIGGRAPH*, 2009.

[10] P. Molina, M. Haro, and S. Jordà. Beatjockey: A new tool for enhancing DJ skills. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 288–291, Oslo, Norway, 2011.

[11] M. Mulligan. Music formats and artist creativity in the age of media mass communication. a music industry blog report. Accessed at: http://musicindustryblog.wordpress.com/free-reports/.

[12] F. Pachet and P. Roy. Markov constraints: steerable generation of markov sequences. *Constraints*, 16(2):148–172, 2009.

[13] A. Roebel. A new approach to transient processing in the phase vocoder. In *Proc. International Conference on Digital Audio Effects (DAFx)*, pages 2344–349, 2003.

[14] L. Russolo. l'Arte dei Rumori, 1913.

[15] D. Schwarz. Current research in concatenative sound synthesis. In *Proceedings of the International Computer Music Conference (ICMC)*, 2005.

[16] H. Vinet, P. Herrera, and F. Pachet. The CUIDADO project. In *Proc. International Conference on Music Information Retrieval (ISMIR)*, IRCAM, Paris, 2002.

[17] R. Wöhrmann and G. Ballet. Design and architecture of distributed sound processing and database systems for web-based computer music applications. *Computer Music Journal*, 23(3):73–84, 1999.

[18] E. Wold, T. Blum, D. Keislar, and J. Wheaten. Content-based classification, search, and retrieval of audio. *MultiMedia, IEEE*, 3(3):27–36, 1996.

[19] A. Zils and F. Pach. Musical Mosaic. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DaFx-01)*, pages 39–44, University of Limerick, 2001.

## 5.3 OTHER EXPLOITATION AREAS

MIR can be used in settings outside of music distribution and creation, for example in musicology, digital libraries, education and eHealth. In computational musicology, MIR tools have become standard "tools of the trade" for a new generation of empirical musicologists. Likewise, MIR technology is used for content navigation, visualisation, and retrieval in digital music libraries. MIR also shows promise for educational applications, including music appreciation, instrument learning, theory and ear training, although many current applications are still at an experimental stage. eHealth (healthcare practice supported by electronic processes) is also starting to benefit from MIR. Thus, the challenge is to better exploit MIR technologies in order to produce useful applications for other fields of research and practice. For this, current practices and needs from the related communities should be carefully studied. The stakeholders include music professionals, musicologists, music students, music teachers, digital librarians, medical doctors and medical doctors and patients who can benefit from music therapy.

### 5.3.1 State of the art

We review here the already existing and potential relations between MIR and musicology, digital libraries, education and eHealth, which we identified as particularly relevant for our field of research.

#### Applications in musicology

The use of technology in music research has a long history (e.g. see Goebl [19] for a review of measurement techniques in music performance research). Before MIR tools became available, music analysis was often performed with hardware or software created for other purposes, such as audio editors or speech analysis tools. For example, Repp used software to display the time-domain audio signal, and he read the onset times from this display, using audio playback of short segments to resolve uncertainties [27]. This methodology required a large amount of human intervention in order to obtain sufficiently accurate data for the study of performance interpretation, limiting the size and number of studies that could be undertaken. For larger scale and quantitative studies, automatic analysis techniques are necessary. An example application of MIR to music analysis is the beat tracking system BeatRoot [15], which has been used in studies of expressive timing [18, 20, 30]. The SALAMI (Structural Analysis of Large Amounts of Music Information [75]) project is another example of facilitation of large-scale computational musicology through MIR-based tools. A general framework for visualisation and annotation of musical recordings is Sonic Visualiser [8], which has an extensible architecture with analysis algorithms supplied by plug-ins. Such audio analysis systems are becoming part of the standard tools employed by empirical musicologists [9, 10, 22], although there are still limitations on the aspects of the music that can be reliably extracted, with details such as tone duration, articulation and the use of the pedals on the piano being considered beyond the scope of current algorithms [24]. Other software such as GRM Acousmographe, IRCAM Audiosculpt [5], Praat [4] and the MIRtoolbox [76], which supports the extraction of high-level descriptors suitable for systematic musicology applications, are also commonly used. For analysing musical scores, the Humdrum toolkit [21] has been used extensively. It is based on the UNIX operating system's model of providing a large set of simple tools which can be combined to produce arbitrarily complex operations. Recently, music21 [11] has provided a more contemporary toolkit, based on the Python programming language.

#### Applications in digital library

A digital library (DL) is a professionally curated collection of digital resources, which might include audio, video, scores and books, usually accessed remotely via a computer network. Digital libraries provide software services for management and access to their content. Music Digital Librarians were among the instigators of the ISMIR community, and the first ISMIR conference (2000) had a strong DL focus. Likewise the contributions from the MIR community to DL conferences (Joint Conference on Digital Libraries, ACM Conference on Digital Libraries, IEEE-CS Conference on Advances in Digital Libraries) were numerous. This could be due to the fact that at

[75] http://ddmal.music.mcgill.ca/research/salami

[76] https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

the end of 90s, musical libraries moved to digitisation of recordings and to multi-information access (video, score images, and text documents such as biographies and reviews) to create multimedia libraries [17] [16] [25]. In this first trend, the technological aspects of these libraries relied mainly on the server, database, media digitisation, text search, and synchronisation (often manual) between media. Today this trend still exists and is accessible online for a wide audience. Examples of this are the "Live TV" of the Cite de la Musique (large audience) with synchronisation of video concerts with libretto, scores and comments. A second trend, that appeared in the mid-2000s, reverses the relationship between Libraries and Music Information Research and Technology. Research and technology enable content estimation, visualisation, search and synchronisation, which are then used in the context of Digital Libraries to improve the usability and access of the multi-documents in libraries (online or not). Examples of this are: inclusion of automatic audio summaries in the IRCAM Library [26], the Bachotheque to compare automatically synchronised interpretations of a same piece [28], optical score recognition and audio alignment for the Bavarian State Library [12]. Also, thanks to the development of technologies (Flash, HTML5, Java-Script), the de-serialisation of media becomes a major theme, along with improved browsing and access to the temporal aspect of media. New concepts of interfaces to enhance listening have been developed which make use of time-based musical annotations (Ecoute augmentee/Increased-listening [77], or today's SoundCloud [78]). A third trend concerns the aggregation of content and the use of user-generated annotation. The content of dedicated libraries can be aggregated to form meta-libraries (e.g. www.musiquecontemporaine.fr) using the shared protocol OAI-PMH. Content can be distributed over the web or aggregated to a local collection. Using Semantic Web technologies such as Linked Data and ontologies, web content can be re-purposed (e.g. the BBC's use of the Music Ontology). This trend is also found in new forms of music access (such as Spotify) which aggregate content related to the music item (AMG reviews, wikipedia artist biography). Comparing the suggestions of [7] and the observations of [2] a decade later, it is clear that much work is still to be done before MIR technology is fully incorporated into traditional libraries. The European ASSETS project [79], working with the Europeana multi-lingual European cultural collection, aims to improve search and browsing access to the collection, including multimedia objects.

### Applications in education

Dittmar [14] partitions MIR methods that are utilised in music education into three categories: music transcription, solo and accompaniment track creation, and generation of performance instructions. Systems for music education that exploit MIR technology include video games, music education software (e.g. Songs2See [80]), and music-related apps focused on learning (e.g. Rock Prodigy [81]). Regarding instrument learning scenarios, MIR is also seeing an uptake via provision of feedback to learners in the absence of a teacher [29], interactive ear training exercises (e.g. the Karajan iPhone apps), automatic accompaniment [13], page turning [1] and enhanced listening (e.g. iNotes: Orchestral Performance Companion). Research projects focused on music education include IMUTUS (Interactive Music Tuition System), i-Maestro (Interactive Multimedia Environment for Technology Enhanced Music Education and Creative Collaborative Composition and Performance) and M4M (Musicology for the Masses) – for the latter, a web-based tool called Yanno [82], based on automatic chord detection, was proposed for secondary school music classes. Although much work still remains to be done, large-scale experiments have already taken place, such as the IRCAM Music Lab 1 and 2 for the French National Education.

Education is one of the most understudied and yet promising application domains for MIR. While Piaget's constructivism and Papert's constructionism are classics of pedagogy and interaction design relating to children, mashup, remix and recycling of contents might be considered a much more controversial and radical approach, especially for their social, ethical and legal implications. However, it is undeniable that young people are embracing remix en masse, and it is integral to how they make things and express ideas. The cultural practices of mashup and remix brought to school, will force us to rethink the role of teachers as part of this knowledge-building

[77] http://apm.ircam.fr
[78] https://soundcloud.com
[79] http://www.assets4europeana.eu
[80] http://www.songs2see.com
[81] http://www.rockprodigy.com
[82] http://yanno.eecs.qmul.ac.uk

process (Erstad, 2008). The development of learning strategies that support such models of creation represents an ongoing challenge as it defies the current model of schooling, with students taking a more active role in developing knowledge. The introduction of MIR-powered tools for musical education and creation among younger children, combined with recent developments in portable devices, opens a new line of research for suitable novel interfaces and applications.

#### Applications in eHealth (healthcare practice supported by electronic processes)

Use of music information research for eHealth is still in its infancy. Its main use to date has been in music therapy, where it has been employed for quantitative analysis of therapy sessions and selection of musical material appropriate to the user's ability and taste. Music information technologies have traditionally been used to characterise one's musical preferences for applications such as music retrieval or recommendation (see for example the Musical Avatar of [6]). Moreover, there has been much research on technologies for affective analysis of music, e.g. on music emotion characterisation. These technologies have a great potential for contributing to music therapy, e.g. providing personalised music tools. For instance, according to E. Bigand, advances in cognitive neurosciences of music have revealed the potential importance of music for brain and cognitive stimulation [3]. At this ISMIR 2012 keynote speech, he referred to some examples of the relationship between music technologies and cognitive stimulation (e.g. "Happy Neuron" project [83]). Systems making use of MIR for music therapy have already been proposed inside the MIR community, e.g. the work by the team led by Ye Wang at the National University of Singapore [84] [23, 31]. In [31], an MIR system is used to automatically recommend music for users according to their sleep quality in the goal of improving their sleep. In [23] an MIR system that incorporates tempo, cultural, and beat strength features is proposed to help music therapists to provide appropriate music for gait training for Parkinson's patients. The Mogat system of [32] is used to help cochlear implant recipients, especially pre-lingually deafened children. In this system, three musical games on mobile devices are used to train their pitch perception and intonation skills, and a cloud-based web service allows music therapists to monitor and design individual training for the children.

### 5.3.2 Specific Challenge

- **Produce descriptive content analysis tools based on concepts used by musicologists.** Current MIR tools do not fit many of the needs of musicologists, partly due to their limited scope, and partly due to their limited accuracy. To fill the acknowledged gap between the relatively low-level concepts used in MIR and the concepts of higher levels of abstraction central to music theory and musicology, will call for, on the one hand, the development of better algorithms for estimating high level concepts from the signal, and on the other hand, the proper handling of errors and confidence in such estimation.
- **Overcome barriers to uptake of technology in music pedagogy.** Generic tutoring applications do not engage the user, because they ignore the essential fact that users have widely varying musical tastes and interests, and that the drawing power of music is related to this personal experience. User modelling or personalisation of MIR systems is an open challenge not just for tutoring but for all MIR applications. Another issue is that MIR technology is currently not mature or efficient enough for many educational applications, such as those involving real-time processing of multi-instrument polyphonic music. Further research is required in topics such as polyphonic transcription, instrument identification and source separation, and in the integration of these techniques, in order to develop more elaborate music education tools than currently exist.
- **Provide diagnosis, analysis and assessment of music performance at any level of expertise.** A further barrier to uptake is that current music education tools have shallow models of music making (e.g. focusing only on playing the correct notes), and fail to give meaningful feedback to learners or assist in the development of real musical skills. More advanced tools will need to aid learners in areas such as phrasing, expressive timing, dynamics, articulation and tone.

[83] http://www.happy-neuron.com
[84] http://www.comp.nus.edu.sg/~wangye

- **Develop visualisation tools for music appreciation.** The listening experience can be enhanced via visualisations, but it is a challenge to provide meaningful visualisations, for example those which elucidate structure, expression and harmony, which inform and stimulate the listener to engage with the music.
- **Facilitate seamless access to distributed music data.** In order to satisfy information needs and promote the discovery of hidden content in digital music libraries, it is necessary to provide better integration of distributed data (content and meta-data, regardless of location and format) through the use of standards facilitating interoperability, unified portals for data access, and better inter-connections between institutional meta-data repositories, public and private archive collections, and other content. Open source tools for indexing, linking and aggregation of data will be particularly important in achieving this goal.
- **Expand the scope of MIR applications in eHealth.** Some preliminary work has demonstrated the value of MIR technologies in eHealth, for example to assist health professionals in selecting appropriate music for therapy. However, the full use of MIR in medicine still needs to be deeply explored, and its scope expanded within and beyond music therapy.

## References

[1] A. Arzt, G. Widmer, and S. Dixon. Automatic page turning for musicians via real-time machine listening. In *18th European Conference on Artificial Intelligence*, pages 241–245, Patras, Greece, July 2008.

[2] M. Barthet and S. Dixon. Ethnographic observations of musicologists at the British Library: Implications for music information retrieval. In *12th International Society for Music Information Retrieval Conference*, pages 353–358, Miami, Florida, USA, 2011.

[3] E. Bigand. Cognitive stimulation with music and new technologies. In *Keynote speech, ISMIR 2012*, Porto, Portugal, 2012.

[4] P. Boersma and D. Weenink. Praat: Doing phonetics by computer, 2006.

[5] N. Bogaards, C. Yeh, and J. Burred. Introducing ASAnnotation: a tool for sound analysis and annotation. In *International Computer Music Conference*, Belfast, Northern Ireland, August 2008.

[6] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambó, E. Gómez, and P. Herrera. Semantic audio content-based music recommendation and visualization based on user preference examples. *Information Processing & Management*, 49:13–33, 01/2013 2013.

[7] A. Bonardi. IR for contemporary music: What the musicologist needs. In *International Symposium on Music Information Retrieval*, Plymouth, Massachusetts, USA, 2000.

[8] C. Cannam, C. Landone, and M. Sandler. Sonic Visualiser: an open source application for viewing, analysing, and annotating music audio files. In *ACM Multimedia 2010 International Conference*, pages 1467–1468, Florence, Italy, October 2010.

[9] N. Cook. Computational and comparative musicology. In E. Clarke and N. Cook, editors, *Empirical Musicology: Aims, Methods, and Prospects*, pages 103–126. Oxford University Press, New York, 2004.

[10] N. Cook. Performance analysis and Chopin's mazurkas. *Musicae Scientae*, 11(2):183–205, 2007.

[11] M.S. Cuthbert and C. Ariza. music21: A toolkit for computer-aided musicology and symbolic music data. In *9th International Society for Music Information Retrieval Conference*, pages 637–642, Utrecht, Netherlands, August 2010.

[12] D. Damm, C. Fremerey, V. Thomas, and M. Clausen. A demonstration of the Probado music system. In *Late-breaking session of the 10th International Society for Music Information Retrieval Conference*, Miami, Florida, USA, 2011.

[13] R.B. Dannenberg and C. Raphael. Music score alignment and computer accompaniment. *Communications of the ACM*, 49(8):38–43, 2006.

[14] C. Dittmar, E. Cano, J. Abesser, and S. Grollmisch. Music information retrieval meets music education. In M. Müller, M. Goto, and M. Schedl, editors, *Multimodal Music Processing*, volume 3 of *Dagstuhl Follow-Ups*, pages 95–120. Dagstuhl Publishing, 2012.

[15] S. Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58, 2001.

[16] J. W. Dunn and E. J. Isaacson. Indiana University digital music library project. In *Joint Conference on Digital Libraries*, Roanoke, VA, USA, June 2001.

[17] M. Fingerhut. The IRCAM multimedia library: a digital music library. In *IEEE Forum on Research and Technology Advances in Digital Libraries*, Baltimore, MD, USA, May 1999.

[18] S. Flossmann, W. Goebl, and G. Widmer. Maintaining skill across the life span: Magaloff's entire Chopin at age 77. In *International Symposium on Performance Science*, pages 119–124, Auckland, New Zealand, December 2009.

[19] W. Goebl, S. Dixon, G. De Poli, A. Friberg, R. Bresin, and G. Widmer. Sense in expressive music performance: data acquisition, computational studies, and models. In P. Polotti and D. Rocchesso, editors, *Sound to Sense - Sense to Sound: A State of the Art in Sound and Music Computing*, pages 195–242. Logos Verlag, Berlin, 2008.

[20] M. Grachten, W. Goebl, S. Flossmann, and G. Widmer. Phase-plane representation and visualization of gestural structure in expressive timing. *Journal of New Music Research*, 38(2):183–195, 2009.

[21] D. Huron. *Music Research Using Humdrum: A User's Guide*. Center for Computer Assisted Research in the Humanities. Stanford, California, 1999.

[22] D. Leech-Wilkinson. *The Changing Sound of Music: Approaches to Studying Recorded Musical Performance*. CHARM, London, 2009.

[23] Z. Li, Q. Xiang, J. Hockman, J. Yang, Y. Yi, I. Fujinaga, and Y. Wang. A music search engine for therapeutic gait training. In *Proc. of ACM Multimedia*, pages 627–630, Florence, Italy, 2010. ACM.

[24] S. McAdams, P. Depalle, and E. Clarke. Analyzing musical sound. In E. Clarke and N. Cook, editors, *Empirical Musicology: Aims, Methods, and Prospects*, pages 157–196. Oxford University Press, New York, 2004.

[25] J.R. McPherson and D. Bainbridge. Usage of the MELDEX digital music library. In *2nd International Symposium on Music Information Retrieval*, Bloomington, Indiana, USA, October 2001.

[26] F. Mislin, M. Fingerhut, and G. Peeters. Automatisation de la production et de la mise en ligne de resumes sonores. Master's thesis, Institut des Sciences et Techniques des Yvelines, 2005.

[27] B.H. Repp. Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America*, 95(5):2546–2568, 1992.

[28] F. Soulez, X. Rodet, and D. Schwarz. Improving polyphonic and poly-instrumental music to score alignment. In *4th International Conference on Music Information Retrieval*, pages 143–148, Baltimore, Maryland, USA, 2003.

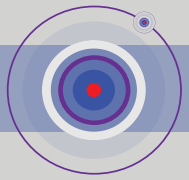
[29] Y. Wang and B. Zhang. Application-specific music transcription for tutoring. *IEEE MultiMedia*, 15(3):70–74, July 2008.

[30] G. Widmer, S. Dixon, W. Goebl, E. Pampalk, and A. Tobudic. In search of the Horowitz factor. *AI Magazine*, 24(3):111–130, 2003.

[31] W. Zhao, X. Wang, and Y. Wang. Automated sleep quality measurement using EEG signal: first step towards a domain specific music recommendation system. In *ACM Multimedia*, pages 1079–1082, Florence, Italy, 2010.

[32] Y. Zhou, K.C. Sim, P. Tan, and Y. Wang. MOGAT: Mobile games with auditory training for children with cochlear implants. In *ACM Multimedia*, pages 1309–1310, Nara, Japan, October 2012.

# CONCLUSION



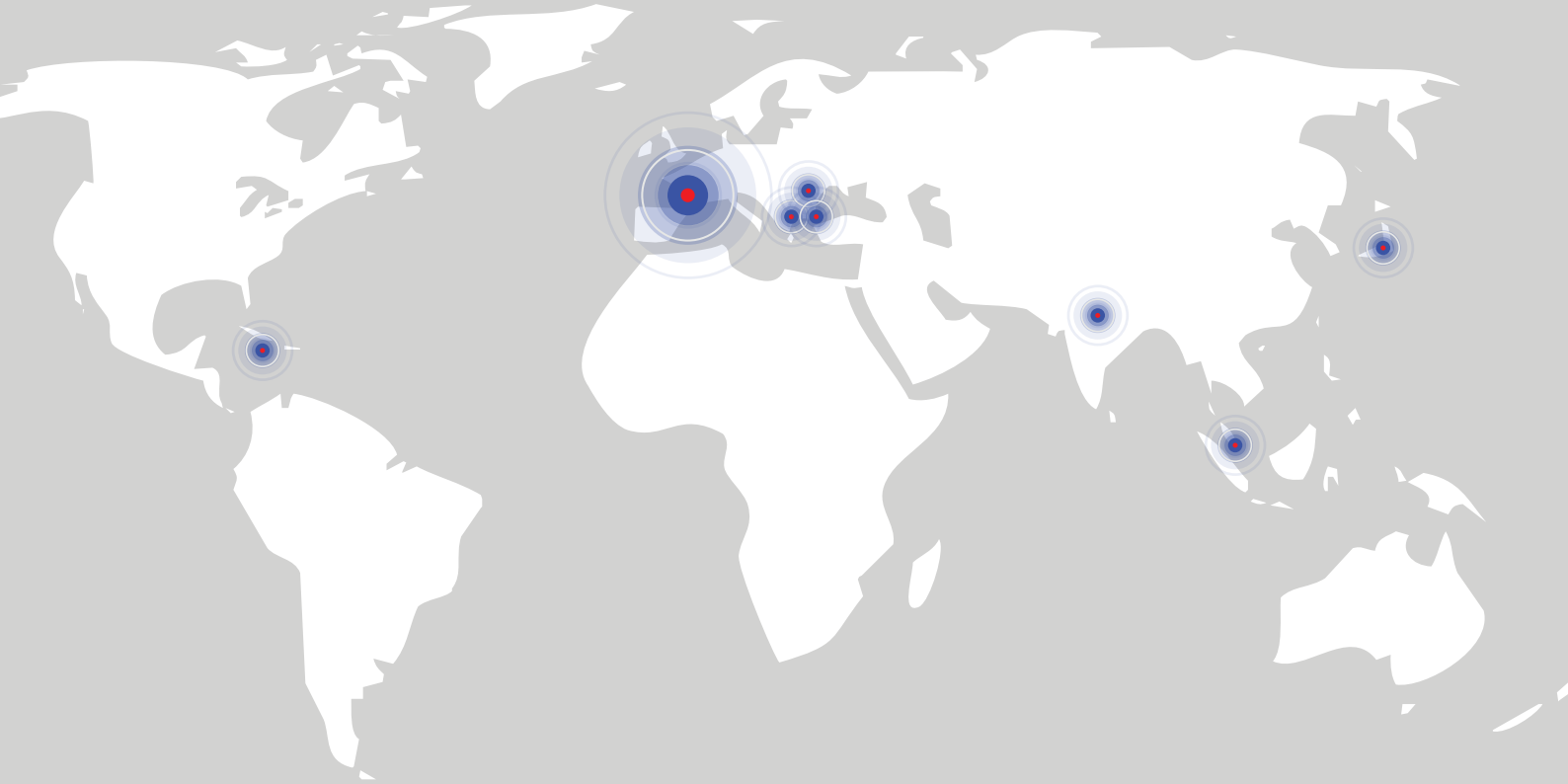
CHAPTER 6

# Conclusions

This document has been conceived in order to identify current opportunities and challenges within the field of Music Information Research. The aim has been to open up the current views that drive the MIR field by taking into account science, industry and society. A review of the state of the art of the field has been conducted and the challenges have been identified by involving a variety of stakeholders. The proposed challenges have great potential for future impact on both academia and industry. In addition to the scientific and engineering points of view, the challenges have focused on social and industrial perspectives, thus aligning the Roadmap with Horizon 2020, the new EU Framework Programme for Research and Innovation [1].

By involving a variety of experts and points of view we hope to have provided a document of interest to both the research community and to policy makers. The open discussions that have been organised in diverse forums have already made a very positive impact upon the MIR community. From here on the success of this initiative will be reflected by the number of students and researchers that read and use this document to make decisions about the direction of their research, especially when deciding which research challenges to address. The proposed research challenges, as well as the knowledge and the network built during the coordination process should also be relevant for policy makers, facilitating future gazing and the establishment of key Music Information Research funding strategies.

[1] http://ec.europa.eu/research/horizon2020



EUROPEAN COMMISSION

SEVENTH FRAMEWORK PROGRAMME