METHODOLOGICAL CONSIDERATIONS CONCERNING MANUAL ANNOTATION OF MUSICAL AUDIO IN FUNCTION OF ALGORITHM DEVELOPMENT

Micheline Lesaffre[1], Marc Leman[1], Bernard De Baets[2] and Jean-Pierre Martens[3]

[1] IPEM: Department of Musicology, Ghent University, Blandijnberg 2, 9000-Ghent, Belgium
Micheline.Lesaffre@UGent.be
[2] Department of Applied Mathematics, Biometrics and Process Control, Ghent University
[3] Department of Electronics and Information Systems (ELIS), Ghent University

ABSTRACT

In research on musical audio-mining, annotated music databases are needed which allow the development of computational tools that extract from the musical audio-stream the kind of high-level content that users can deal with in Music Information Retrieval (MIR) contexts. The notion of musical content, and therefore the notion of annotation, is ill-defined, however, both in the syntactic and semantic sense. As a consequence, annotation has been approached from a variety of perspectives (but mainly linguistic-symbolic oriented), and a general methodology is lacking. This paper is a step towards the definition of a general framework for *manual* annotation of musical audio in function of a computational approach to musical audio-mining that is based on algorithms that learn from annotated data.

1. INTRODUCTION

Annotation[1] refers to the act of describing content using appropriate space-time markers and labeling. Annotation generates additional information that may be useful in contexts of information retrieval and data-mining, either as indices for retrieval or as training data for the development of computational tools. In that respect, annotation is a broad field that covers semantic content as well as labeling and segmentation in function of algorithm development. In the music domain, annotation pertains to the description of metadata and musical features that users might find particularly relevant in the context of music information retrieval.

In the present paper, we will mainly focus on the *manual* annotation of musical audio in function of the development, through computational learning, of tools for the automatic generation of similar annotations from the audio.

Up to now, there is a general lack of training data and the methodology for manual annotation of musical audio in function of algorithm development is largely under-estimated and under-developed. This is due to the fact that, unlike speech annotation, music is less determined in terms of its content. Unlike speech sounds, music is not defined by a limited set of lexical entities. Its syntax is typically depending on multiple constraints that allow a great and almost unlimited variety of forms and structures. Moreover, its semantics are non-denotative and more depending on subjective appreciation. Consequently, the process of manual annotation is rather complex because it comprises multiple annotation levels and different possible types of content description. The challenge of seeking common ground in the diverse expressions of music annotation has not been addressed thus far. Manual annotation of musical audio indeed raises questions that point to the nature of musical content processing, the context of MIR, and the relationship between natural and cultural constraints involved in musical engagement.

This paper deals with some methodological considerations concerning the manual annotation of musical audio in function of algorithm development. First, the general background in musical audio annotation is reviewed. Then a general framework for annotation is sketched and examples are given of experiments that aim at building up an appropriate methodology. An ongoing large-scale experiment, called *Music Annotation for MAMI,* is reported and the last section is devoted to a discussion and ongoing work.

[1] Music annotation is an open term that integrates any information (textual, visual or auditory) that can be added to music. In music description distinctive music characteristics are described in a way that is close to the original music.

2. BACKGROUND

2.1. Annotation Forms

In the speech community, a large set of annotated databases has been constructed that proved to be useful for the development of algorithms for speech recognition. Since these annotations are based on speech audio, it is natural to investigate to what extent these tools may be useful for music annotation.

The Linguistic Data Consortium (LDC)[2], for example, provides a list of tools and formats that create and manage linguistic annotations. The Annotation Graph Toolkit [4] is a formal framework that supports the development of linguistic transcription and annotation tools. Handschuh et al. [9] describe a framework for providing semantic annotation to the Semantic Web. Bigbee et al. [3] review capabilities of multi-modal annotation by examining how linguistic and gesture analysis tools integrate video data. They point to the increasing importance of multi-modal corpora and interfaces in future tools.

Tools based on linguistics may be useful for simple annotation tasks, but in general, they are not satisfactory. The main problem is that musical content description may require many new description formats, which often go beyond the commonly used metadata-based indexing. A simple example is the annotation of the beat, where the annotator has to tap along with the music, and thus annotates while listening to the played music.

In the music domain, there are but a few initiatives that seem to address this problem, such as the SIMAC[3], MAMI and SEMA[4] projects.

In extending the concept of annotation to music analysis in general, it appears that the literature on music annotation is mainly concerned with linguistic and symbolic descriptions. Few studies have investigated methods for the time synchronous annotation of a musical audio stream. The media industry and researchers involved in content-based music analysis are actively discussing the needs for music representation. Currently the Moving Picture Experts Group (MPEG) is starting a new activity aimed at the systematic support of symbolic forms of music representations by integrating Symbolic Music Representation (SMR) into MPEG multimedia applications and formats. The decoding and rendering should allow the user to add annotations to SMR events. Here annotations are considered as audiovisual objects, music representation elements or simple URL-links. The idea is that the annotation format will be normative, as well as the way annotations are issued by the end-user.

Apart from the limited number of annotation tools, most tools focus on music notation[5], like the development of tools for adding specific interpretation symbols to a score such as bowing, fingering, breathes and simple text. Efforts in this context relate to the development of standards for music description, such as MPEG-4, MPEG-7 and MPEG-21.

A few initiatives have been taken that focus on the annotation of musical audio. Acousmograph (GRM)[6], for example, is similar to a sonogram, offering the user the opportunity to select part of a graph and listen to the chosen image. Timeliner [17] is another example of a visualization and annotation tool for a digital music library. It enables users to create their own annotated visualizations of music retrieved from a digital library.

If we then look at the music databases that have been annotated using existing annotation tools, it turns out that the main focus has been on metadata description, and not, or less, on the description of musical content as such. Don Byrd maintains a list[7] as a work-in-progress that surveys candidate MIR collections. Many of these databases, however, already start from symbolic representation of music (scores). The Repertoire International des Sources Musicales (RISM), for example, documents musical sources of manuscripts or printed music, works on music theory and libretti stored in libraries, archives, monasteries, schools and private collections and provides images of musical incipits. The RISM Music Manuscript Database is linked to three other databases providing additional information to specific content: Composer, Library Sigla and Bibliographic Citations. The MELDEX Digital Music Library [16] handles melodic or textual queries and offers twofold access to songs. The results of a query are visualized or presented as an automatically compiled list of metadata, such as titles. At the Center for Computer Assisted Research in the Humanities (CCARH) MuseData[8] has been designed to represent both notational and sound information (MIDI). The Real World Computing (RWC) Music Database [8] is built in view of meeting the need of commonly available databases for research purposes. It consists of 4 databases containing popular music, classical music, jazz music, and royalty free music. Two other component databases were added with musical genre and musical instrument sounds. RWC contains music in both MIDI and audio form and provides lyrics of songs as text files. Standard MIDI files are generated as substitutes for scores, for genres for which no scores are available.

To sum up, most annotation and analysis tools have paid attention to linguistic and symbolic oriented annotation, but the picture of music annotation is rather dispersed. There is no clear methodology, nor is the problem domain very well described. A theory of music annotation is lacking.

2.2. Problem specification

The major task of musical audio-mining is to make a connection between the musical audio stream on the one hand and user-friendly content descriptions on the other hand. The main problem is that audio streams are physical representations, while user-friendly descriptions pertain to high-level human information processing capabilities that involve a complex set of goal-directed cognitive, affective and motor actions. Humans typically process musical information in terms of purposes, goal directed actions, values and meanings.

[2] www.ldc.upenn.edu/annotation/
[3] www.semanticaudio.org
[4] www.ipem.ugent.be
[5] Music notation refers to the representation of music by a system of symbols, marks or characters.
[6] www.ina.fr/grm/outils_dev/acousmographe/
[7] http://www.ismir.net/
[8] www.ccarh.org

They handle a subjective (first person) ontology that is very different from the objective (third person) ontology of physical signals (see [14] for a more detailed account).

A major goal of manual annotation of musical audio is therefore to provide data that allows computational systems to learn the task of annotation and therefore to build bridges between first person descriptions and third person descriptions of music. Modelling based on imitation learning is considered a candidate to cope with the gap between the measurable quantities of an audio signal and the intentionality of subjective qualities.

3. GENERAL FRAMEWORK

3.1. Context dependencies

There are at least three observations to keep in mind when dealing with music annotation, namely (1) the intentional nature of human communication, (2) the requirements of music information retrieval contexts, and (3) the development of mediation technology.

First of all, since annotation aims at making the link between the musical audio stream and levels of content description that allow humans to access the information stream, it is necessary to take into account the highly focused level of human communication. This level is called the *cultural* level because the implied ontology is based on human learning, subjective experiences and symbolization. Due to the fact that this level is characterized by goal-oriented behavior and intentional attitudes (thoughts, beliefs, desires, …) its descriptions are therefore very different from objective or nature-driven descriptions that pertain to physical signals. As a consequence, there are two methodologies involved:

- **Naturalistic approaches** (studied in the natural sciences) aim at developing tools that extract nature-driven descriptions from audio. These tools are objective in the sense that they start from physical "energies" and rely upon "universal" principles of human information processing. The resulting descriptions have an inter-subjective basis and do not involve the subjective goal-directed action ontology on which human communication patterns typically rely. Examples are the extraction of pitch from a sung melody, or the extraction of timbre classes from polyphonic audio.

- **Culturalistic approaches** (studied in the human sciences), in contrast, tend to describe music in terms of its signification, its meaning, value, and role as cultural phenomenon. Thus far, culture-determined content description has been strongly linguistic-symbolic oriented, based on textual and visual descriptors. Reference is often made to subjective experience and historical and social–cultural interactions.

Some culturalist musicologists from the postmodern school tend to claim that links between physical descriptions and subjective descriptions of music are impossible. Yet, there is no strong proof of evidence for such statement. The main argument draws on the idea that signification (attribution of meaning to music) is an arbitrary activity that is depending on the cultural environment, the history and the personal taste. Association and signification can be wild indeed, but we do believe that there are at least certain aspects of descriptions, including descriptions at the high semantic (first person) levels, that are not completely arbitrary, and that, given a proper analysis of the goals, can be very functional in MIR contexts. When aiming at semantic description of music, this hypothesis is rather fundamental, and skepticism can only be refuted when the proof has been given of a working system.

A second observation, and closely connected to the first point, is that music descriptions serve a goal that is largely determined by the MIR *context*. Out of a myriad of possible natural objective descriptions, and perhaps also subjective descriptions, we should select those that serve the particular goals of the particular context. Hence, research in audio-mining is not purely a matter of bottom-up signal processing and making the link between third person and first person descriptions. At a certain moment, a thorough analysis has to be made of the retrieval context, the economical value, the ethical value, the purpose etc… and decisions may have to be taken about the context-based bias of the whole enterprise, including the work on manual annotation. Reference can be made to the DEKKMMA-project [9] where researchers are confronted with a large audio database of Central African music, and where little experience is available of *how* people would tend to search, and *what* they would tend to search, in such a database. It is likely, but analysis has to clarify this, that users who are unfamiliar with the Central African idiom behave very differently from users that know the music.

A third observation is concerned with the technology used for music mediation. Mediation refers here to the ways in which streams of musical information are transmitted and the tools that may be used to specify content and to retrieve that content. The most recent developments seem to go in the direction of networked wireless mobile devices giving access to a large amount of music in databases. Such technologies may imply certain constraints on the possible types of musical content specification (see e.g. [2] for experiments with Pocket PC).

To sum up, given the aims of a MIR system, annotation should take into account at least three different types of context, namely *culture*, *user*, and *mediation*. Given that background, Figure 1 shows the general framework for annotation of musical audio that incorporates the naturalistic and culturalistic approaches with their focus on human information processing and

[9] www.ipem.ugent.be

social-cultural context, respectively (Adapted from [14]).

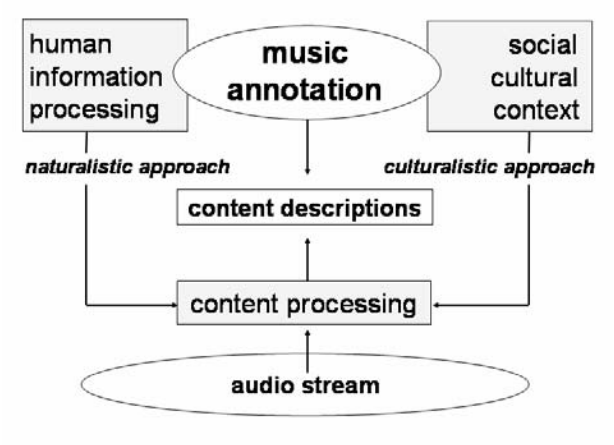

Figure 1. General framework for annotation of musical audio.

3.2. Computer Modeling Approach

Apart from the general framework in which annotation has to be carried out, there is another framework that needs careful analysis, namely that of computational modeling. In the past, the problem of manual annotation of musical content has often been considered from the viewpoint of a *Cartesian* modeling strategy. The strategy consists in the specification of a set of pre-defined feature extractors (the clear and distinct ideas of Descartes) with limited scope to clear meaning in a restricted context (most often) of stimulus-response experimentation. It is then hoped that through combination of a selected set of weighted pre-defined features, high-level semantic knowledge can be predicted. However, it turns out that this strategy has a number of limitations [12] such as a complicated semantic interpretation when features are summarized or combined (linearly as well as non-linearly). If no significant meaning can be given to these features it may be better to give up the idea of working with many local descriptors and look for alternative methods. Pachet & Zils [18] explore an alternative method using an automated processing operators composition method in the framework of genetic programming. In general, however, we believe that straightforward imitation learning based on an appropriate level of manual annotation may be of help.

Taking into account the multiple ways in which users can engage with music, annotation should extend the possibilities of linguistic-symbolic descriptions with non-symbolic and non-linguistic forms of description. This draws on the understanding that the interaction between subjective experience and objective description is a dynamic process constrained by both natural and cultural determinants and that, somehow, levels of annotation in between what is considered to be natural and cultural processing should be chosen. The ecological approach indeed regards any response to music as the result of a complex interaction of the subject in its social-cultural and physical environment. Levels of annotation can be addressed that indeed lie on the borderline of objective/subjective descriptions and that form the connection points with first person and third person descriptions of music.

This calls for an investigation towards new forms of annotation based on mimetic and gesture capabilities of human communication. Performing a manual annotation in the form of motoric action subsumes the interconnectedness between culture-based and nature-based computational music research approaches.

3.3. Representation levels

Annotation comprises diverse types of description, depending on the purpose of the description and the level of annotation. The various types of music annotation are related to syntactic, semantic, structural and articulation elements. Figure 2 shows the distinguished representation levels and associated annotation methods.

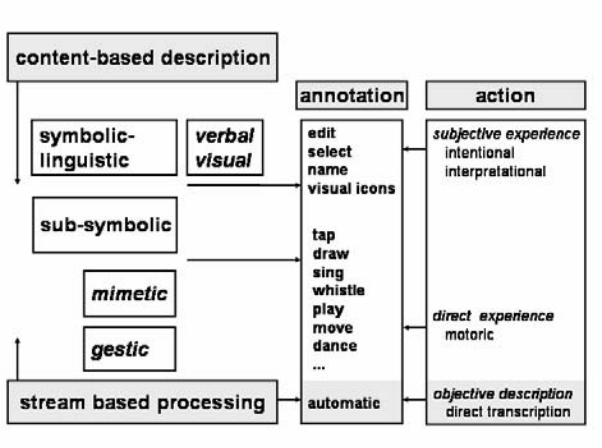

Figure 2. Representation levels and associated annotation methods.

- The **symbolic/linguistic-based annotation** has mainly a focus on the description of structural and semantic units. The user's interaction to symbolic representation relates to verbal and visual descriptors. This could include a score, or conventional music notation machine code (i.e. MIDI, SASL). The main problem of this approach is that symbols are deprived of any semantic connotation and that they require human interpretation based on the inter-subjective semantics. Annotation thus relies on subjective experience of the creator or user who performs an interpretational and intentional action.

- The **sub-symbolic-based annotation** is mainly based on multi-dimensional spaces. Some studies focus on representation forms such as sonification and visualization. Toiviainen[10] for example explored

[10] http://www.cc.jyu.fi/~ptoiviai/

the additional value of visual data mining for large music collections. Through exploration of multiple music dimensions he found that some musical features are more natural oriented and other more cultural. Pampalk [19] presents a visualization method that uses self-organizing maps for grouping similar pieces pertaining to different music genres. Manual annotation involved motor action that is placing of the pieces on a map according to personal style.

- **Mimetic annotation** is related to imitative aspects of motor actions. Imitation behavior in general is a topic of growing interest within the cognitive sciences [22]. The process of imitative learning is based on perceiving real world behavior and learning from it through action. Applied to music research, imitation is a means for analyzing the perception of similarity aspects within music through motor responses. The now popular Query by Voice paradigm supports the idea of retrieving music by vocal imitation of representative melodies (see e.g. [5], [15], [20]).

- **Gestural annotation** accounts for representation as the result of multi-modal gesture-based interaction and emotional expressiveness [13]. The distinction with mimetic annotation is that it needs not be learned or rehearsed. Gestural annotation involves motoric action as a physical manifestation of the sonorous, such as body movement or dancing. Modeling gesture annotation takes into account time dependencies through which it mainly applies to the representation of rhythmic features.
  At the gestic level following annotation forms are distinguished:
  -*sound producing action* such as tapping, hitting and stroking
  -*sound accompanying action* such as body movement and dancing

4. EXPERIMENTAL INVESTIGATION

In view of a search and retrieval technology based on computational learning algorithms that draw upon the notion of imitation, a range of annotation possibilities are currently being studied. In what follows some examples of experimental research, mainly conducted at Ghent University, are given. The focus is on the expertise level of the annotator and the role of different manual annotation methods with relation to the development of computational algorithms.

4.1. Annotation Subjects

A global distinction can be made between annotations made by *experts* (musicologists, performers, teachers, and librarians) and *naive* or common users. Expert annotation has the advantage of being precise and consistent, but often it is singular. The first person descriptions of musicologists, performers and composers are not necessarily shared with those of naive listeners. The latter may perceive the same music completely different. In a similar way distinctions can be made between users who know the musical idiom and users who don't know the idiom.

Furthermore, an annotator can make personal annotations or can use annotations that are provided by others or automatically generated, in a bootstrap process. Tzanetakis & Cook [24] describe a tool for semi-automatic audio segmentation annotation. A semi-automatic approach combines both manual and automatic annotation into a flexible user interface.

The background of the annotator is a determining factor in the annotation process. It is in view of the cultural goal of the annotation that decisions can be taken whether expert annotators or naive annotators are most appropriate. Much depends on the goal and the task of the annotation.

4.2. Annotation Experiments

4.2.1. Linguistic / Symbolic description

At the level of semantic labeling, the use of terminology in the form of selecting or naming keywords is a complex issue. Semantic features don't have a clear meaning and are not universally applicable. Beyond metadata such as title, composer, genre and year there are narrative descriptions of music that draw on the appreciation and subjective experience of users. Recent studies show interest in the terminology used by non-music experts indicating features of music. Kim [11] investigates people's perception of music by means of categorizing the words they use. Bainbridge et al. [1] analyze the questions and answers by which users of MIR systems express their needs. It was found that users are uncertain as to the accuracy of their descriptions and experience difficulty in coming up with crisp descriptions of musical categories such as genre or date. Using the power of the Internet, MoodLogic[11] developed a music meta-database with song information available to users. The MoodLogic user community was questioned through surveys on how they feel about songs and artists, and these answers were collected in a massive database containing information on mood, tempo, genre, sub-genre and beat.

Leman et al. [12] present an empirical study on the perceived semantic quality of musical content in which subjects had to judge musical fragments using adjectives describing perceived emotions and affects. Subjects had to evaluate semantic qualities, presented as 15 bipolar adjectives, on a 7-point scale. More recent results reveal that prediction of affective qualities attributed to music may be possible and therefore usable

[11] www.moodlogic.com/

in MIR-contexts, but results could possibly be improved using induction-based learning paradigms.

4.2.2. Melody imitation

A Query by Voice experiment [15] was conducted that generated an annotated database of 1500 vocal queries, which is freely available on the Internet. For musical imitations in the form of vocal queries user-based and model-based annotation was performed. User-based annotation provided content about the spontaneous behavior of users and model-oriented annotation provided descriptions as a referential framework for testing automatic transcription models. For model-based annotation the PRAAT transcription and annotation tool for speech analysis [6] was used. PRAAT takes in an audio file, allows marking of segments and typing in words. The features investigated for vocal query annotation were segmentation (events), onset time and onset reliability, frequency, pitch stability, query method and sung words or syllables. The results have been used for training the MAMI melody transcriber [7].

Melody imitation is also a useful method for handling the problem of annotation of polyphonic music in that monophonic melodic imitation might serve as reference material. The capability of professional singers for imitating melodies pertaining to different voices has been tested in a pilot study. Each of the eight singers involved had to study the same ten songs. Then participants were asked to imitate as well as possible the main melodies, bass lines and possibly other melodies that they considered important. The files were then transcribed using the MAMI automatic transcription tool [7]. Statistical analysis shows that professional singers can imitate main and other relevant melodies quite well but imitations of the bass lines are less accurate. Another problem imposed by perception issues is the non-consistency in choosing other relevant melodies. [21]

4.2.3. Tonality annotation

In a recent study by Toiviainen and Krumhansl [23] subjects manipulated a virtual slider while listening to a musical fragment (Bach) that contained pulsing probe tones. This method has shown to be interesting because it has the advantage of being based on the theory of analysis of time-series, which is preferable to a scale with a limited amount of steps. At Ghent University a further stage tonality description has been studied which aims at generating manual annotations as natural response [10]. The subjects (26) had to listen to 20 short (60 sec.) fragments of classical and non-classical music (fifty-fifty) and were asked first to sing the best fit and second to express their appreciation. In the first part, it was suggested that participants would sing low, soft and long tones. Appreciation was measured by means of judgment on a seven-point scale of pairs of adjectives related to emotion and tonality features. Graphic annotation was also explored. While they listen to the same music as before, participants had to draw a line which represents the course of the melody. There was a remarkable correspondence among the patterns in the drawings which points to the inter-subjectivity of mental structures when people are acting in the same context and under equivalent conditions.

4.2.4. Rhythm annotation

An ongoing study at our laboratory deals with drum annotations of real music recordings. Aiming at providing ground truth measures for drum detection in raw musical audio the method of annotation by imitation has been tested. A professional drummer imitated, by use of an electronic drum kit, the drums he heard in 4 entire pieces of music. Besides that 5 drum loops were also annotated in the same way. To obtain maximum accuracy the music was given beforehand to the player who was expected to study the drum part. From analyzing the files some problems raised due to cross talk between drums, poor synchronization between the signals (from 20 up to 80 msec.) and insertion of notes generated by the pedal as a result of body movement. The recorded files needed additional manual checking. Manual annotation of polyphonic musical pieces containing drums needs to be done by percussion experts and is very time consuming. Future work might include similar studies in which the player himself also manually corrects the recordings using a sequencer program.

4.2.5. Multiple annotation forms: Music Annotation for MAMI

In context of the MAMI and SEMA project, a new large-scale annotation experiment has been set up. *Music Annotation for MAMI* is a study that investigates the requirements for the development of a system that relies on multiple forms of textual and motoric music annotation. The experiments that are currently conducted aim at collecting a large amount of annotated data that rely on syntax, genre description and appreciation. Focus is on the exploration of the usability of several annotation methodologies that may facilitate handling music databases.

To begin with, an explorative study in the form of an online inquiry has been done to recruit a large group of subjects willing to participate in diverse annotation experiments spread over several months. Until now 717 persons, aged between 15 and 75 year, filled in the inquiry of which 663 (300 male and 363 female) are willing to participate in the experiments. The questionnaire provides individual profiles by collecting following information:

- socio-demographic info
- cultural background
- acquaintance with the Internet
- musical background

- music preferences

In addition people are requested to provide titles of their preferred music together with the composer or performer of the piece. Indication of the genre they think the piece belongs to is asked as well. For each title two sets of 5 bipolar adjectives related to emotion, expression and style are presented to be judged on a 7-point scale.

Statistical analysis will lead to the distinction of specific user groups. The main selection criterion is the formation of equally distributed groups according to age, gender, cultural background, music education, music experience and genre preferences.

People's favorite music is the starting point for the creation of a large database containing a large number of music fragments of 30 seconds. For different experimental issues various subsets of this database are used. Ongoing experiments focus on mimetic (rhythm and melody imitation) and symbolic-linguistic (perceived semantic qualities) annotation.

Table 1 summarizes music characteristics and annotation methods involved in ongoing and future experiments. They relate to the conceptual framework of a taxonomy worked out within the context of audio mining [15].

CATEGORY	SUBCATEGORY	ANNOTATION
STAND. INFO		textual, list selection
MELODY & HARMONY	melody	notated melody, recording, musical excerpt
	chord progression	notated chord progression, recording, musical excerpt
	tonality	notated tonality, recording, musical excerpt
TIMING & RHYTHM	tempo	BPM, tempo tapping, drawing, recording, mus. excerpt
	timing/structure	textual, list selection
	drum patterns	notated drum pattern, recording, musical excerpt
LOUDNESS		list selection, drawing
TIMBRE		list selection
SUBJ. QUAL.		selection of qualities, moving a slider

Table 1. Music characteristics and annotation methods involved in the 'Annotation for MAMI' experiment.

5. DISCUSSION AND ONGOING WORK

Access to large music databases requires interoperable data representations and methods for handling new description formats for querying and delivering musical data. Nowadays audio databases are still in a stage of simple annotation methodology. Existing systems only provide tools that allow metadata-based indexing in view of searching by name or format. Most audio is represented in compressed formats such as MP3, RealAudio, QuickTime etc., and only little research concentrates on real audio. Annotations for musical files usually include information about performer, title, year etc. and are not satisfactory in more sophisticated search and retrieval. For this purpose new descriptors associated with audio signals have to be developed.

A scan of the literature on annotated databases that would allow the training of computational systems however reveals that probably only a few of such databases are available and moreover that they are limited in scope. The current state-of-the-art suffers from a lack of a well-defined conceptual frame that supports (learning) the mapping of the interconnectedness of diverse conceptual levels.

It has been argued that new music annotation methodology is likely to have strong influence on the improvement of information search and retrieval processes and on the most efficient system's usability possible. The development of an annotation system that deals with the interconnectedness between culturalistic and naturalistic approaches might benefit from elaborated exploration of manual annotation and new description methods. An attempt is made to define a general framework for modeling based on imitation learning. It is estimated to facilitate easy computerized handling of large music databases. Previous annotation studies have proven that the use of more advanced methods based on mimetic and gesture skills lead to promising results. These studies are the first steps towards modeling of the relationships between high-level content descriptions and stream-based musical audio. However, the value of this paradigm is only estimable when a large amount of annotated musical data from different user groups is available. The currently conducted Music Annotation for MAMI experiments deal with this issue. It involves multiple manual annotation of a music database incorporating linguistic-symbolic and sub-symbolic descriptions.

Acknowledgements

This research has been conducted at IPEM, Department of musicology at Ghent University in the framework of the MAMI project for audio recognition. We are grateful for the financial support given by The Flemish Institute for the Promotion of Scientific and Technical Research in Industry.

6. REFERENCES

[1] Bainbridge, D., Cunningham S. J. and Downie J. S. "Analysis of queries to a Wizard-of-Oz MIR system: Challenging assumptions about what people really want", *Proceedings of the 4th International Conference on Music Information Retrieval,* Baltimore, 2003.

[2] Baumann, S., and Halloran J. "An ecological approach to multimodal subjective music similarity perception*", Proceedings of the Conference in Interdisciplinary Musicology (CIM)*, Graz, 2004.

[3] Bigbee, T., Loehr D., and Harper L., "Emerging Requirements for Multi-Modal Annotation and Analysis Tools", *Proceedings, Eurospeech 2001 Special Event: Existing and Future Corpora -- Acoustic, Linguistic, and Multi-modal Requirements,* Aalborg, Denmark, 2001.

[4] Bird, S. and Liberman, M. "A formal framework for linguistic annotation", *Speech Communication*, 33(1,2), 2001.

[5] Birmingham W., "MUSART: Music Retrieval Via Aural Queries", *Proceedings of the 2nd International Conference on Music Information Retrieval,* Bloomington, 2001.

[6] Boersma, P., and Weenink, D. (1996). *Praat. A system for doing phonetics by computer*. Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam. Retrieved July 31, 2003, from http://www.praat.org.

[7] De Mulder, T., Martens J.P., Lesaffre M., Leman M. and B. De Baets "An auditory model based transriber of vocal queries", *Proceedings of the 4th International Conference on Music Information Retrieval,* Baltimore, 2003.

[8] Goto, M., Hashiguchi H., Nishimura T., and Oka R. "RWC Music Database: Music Genre Database and Musical Instrument Sound Database". *Proceedings of the 4th International Conference on Music Information Retrieval,* Baltimore, 2003.

[9] Handschuh, S., Staab S. and Volz R., "On Deep Annotation". *Proceedings of the 12th International World Wide Web Conference, WWW 2003,* Budapest, Hungary, 2003.

[10] Heylen, E. et al. Paper in progress

[11] Kim, J.-Y. and Belkin, N. J. "Categories of Music Description and Search Terms and Phrases Used by Non-Music Experts", *Proceedings of the 3th International Conference on Music Information Retrieval,* Paris, 2002.

[12] Leman, M., Vermeulen V., De Voogdt L., Taelman J., Moelants D. and Lesaffre M., "Correlation of gestural musical audio cues and perceived expressive qualities", in A. Camurri and G. Volpe (Eds.) *Gesture-based communication in human-computer interaction.* Berlin, Heidelberg, Springer-Verlag, 2003, 40-54.

[13] Leman M., and Camurri A., "Musical content processing for Interactive Multimedia", *proceedings of the Conference on Interdisciplinary Musicology,* Graz, 2004.

[14] Leman, M. *From Music Description to its Intentional use*. Manuscript, 2004.

[15] Lesaffre M., Tanghe K., Martens G., Moelants D., Leman M., De Baets B., De Meyer H. and Martens J.-P. "The MAMI Query-By-Voice Experiment: Collecting and annotating vocal queries for music information retrieval", *Proceedings of the 4th International Conference on Music Information Retrieval,* Baltimore, 2003.

[16] McNab, R. J., Smith L. A., Bainbridge D. and Witten I. H. "The New Zealand Digital Library MELody inDEX", *D-Lib Magazine*, 1997.

[17] Notess, M., and Swan, M., "Timeliner: Building a Learning Tool into a Digital Music Library", *Accepted to the 2004 ED-MEDIA World Conference on Educational Multimedia, Hypermedia & Telecommunications to be held* Lugano, Switzerland, 2004.

[18] Pachet, F. and Zils, "A. Evolving automatically high-level music descriptors from acoustic signals", in *Computer Music Modeling and Retrieval: International Symposium, CMMR 2003*, Berlin, Heidelberg Springer-Verlag LNCS, 2771, 2003, 42-53.

[19] Pampalk, E., Dixon S. and Widmer G. "Exploring music collections by browsing different views", *Proceedings of the 4th International Conference on Music Information Retrieval,* Baltimore, 2003.

[20] Pauws S., "CubyHum: a fully operational "query by humming" system", *Proceedings of the 3th International Conference on Music Information Retrieval,* Paris, 2002.

[21] Stijns, F. et al. Paper in progress

[22] Tomasello M. *The Cultural Origins of Human Cognition*. Harvard University Press.

[23] Toiviainen, P. and Krumhansl, C.L. "Measuring and modeling real-time responses to music: The dynamics of tonality induction", *Perception,* 32, 6, 2003.

[24] Tzanetakis, G. and Cook P. "Experiments in computer-assisted annotation of audio", *Proceedings International Conference Auditory Display (ICAD),* Atlanta, Georgia, 2000.