

A User-Oriented Approach to Music Information Retrieval.

Micheline Lesaffre¹, Marc Leman¹, Jean-Pierre Martens²,

¹ IPEM, Institute for Psychoacoustics and Electronic Music, Department of Musicology,
Ghent University, Blandijnberg 2, 9000 Ghent, Belgium

Micheline.Lesaffre@UGent.be

²ELIS, Department of Electronics and Information Systems, Ghent University, St.
Pietersnieuwstraat 41, 9000 Ghent, Belgium

Abstract. Search and retrieval of specific musical content such as emotive or sonic features has become an important aspect of Music Information Retrieval system development, but only little research is user-oriented. We summarize results of an elaborate user-study that explores who the users of music information retrieval systems are and what structural descriptions of music best characterize their understanding of music expression. Our study reveals that perceived qualities of music are affected by the context of the user. Subject dependencies are found for age, music expertise, musicianship, taste and familiarity with the music. Furthermore, interesting relationships are discovered between expressive and structural features. These findings are validated by means of a Semantic Music Recommender System prototype. The demonstration system recommends music from a database containing the quality ratings provided by the participants in a music annotation experiment. A test in the real world revealed high user satisfaction which illustrates the potential of querying a music database by semantic descriptors for affect, structure and motion.

Keywords: semantic description, music information retrieval, user profile, music recommendation, query by emotion

1 Introduction

Music researchers who are being challenged by developing content-based music information retrieval (MIR) systems need understanding of the relationships between user dependencies, descriptions of perceived qualities of music and musical content extracted from the audio. One of the weaknesses in music information retrieval research is that there is shortage of information on user-dependencies, especially with respect to the importance of high-level features of music. The success of music information technology, however, primarily depends on its users, that is to say on assessing and meeting the variation among user groups. Thus far no research has been investigating who are the potential users of music information retrieval systems, how they would describe music qualities and how we can define the higher-order understanding of music features that the average users share.

In the present paper we summarize results of an elaborate study that was set up to investigate meaningful relationships between the user's context and their perception of qualities of music by means of ratings of semantic content. This paper consists of four sections. In the first section we introduce the dichotomy in music content that dominates the diversity of approaches to music information retrieval. Second, a global picture is given of a theoretical framework for a user-oriented approach to music information retrieval. The set up and results of an elaborate user study are summarized in section three. Finally, in section four, we describe a semantic music recommender system demo that was developed to validate the outcome of the study.

1.1 Music content dichotomy

The core task of content-based music information retrieval systems is to allow users to search musical pieces using music qualities as a search key. Such high level content will be based on the user's description of musical experiences. Content-based music analysis thus relates to the transformation of sound energy into semantic variables associated with a piece of music.

Many difficulties encountered in content-based music information retrieval system development stem from a music content dichotomy that is defined by a mismatch between two processes. On the one hand there is the process of content extraction by the system (i.e. low level) and on the other hand there is the process of content addition by the user (i.e. high level). Processes that deal with content extraction are bottom-up and approach music content from the angle of physics and computer science. Processes of content addition are top-down and pertain to the domain of human perception and cognition. They deal with aspects of user behavior and experience that are high-level semantics.

From the perspective of computer science, music content consists of data that is stored and used by a computer program. In this sense, content is quantified and does not necessarily entail meaning. Contrarily, from the perspective of music psychology meaning or quality content is very relevant.

Much of the music information retrieval research still focuses on bottom-up technology. In order to make a music information retrieval system appealing and useful to the envisaged user, more effort should be spent on user oriented approaches. Such approaches bear close similarity to music perception, which is an area that is often underestimated in music information retrieval system development.

1.2 User-oriented approaches

Because the social and psychological functions of music are very important it can be expected that the most useful retrieval systems will be those that facilitate searching according to these functions. Typically such indexes will focus on stylistics, mood and similarity information provided by the system users.

Search behavior depends on highly developed abilities to perceive and interpret musical information. A user must call to mind a great deal of analogies, metaphors and memories in order to make coherent sense out of the music content.

Although a substantial number of research projects have addressed music information retrieval, the user-oriented approaches are still in their infancy. Existing studies tend to be small (e.g. Yang and Lee, 2004) and mainly rely on a university population (e.g. Lee and Downie, 2004). The literature scarcely reports on responses from real users to carefully crafted questionnaires assessing their context (e.g. personal background, spontaneous behaviour, habits, musical skills, perceptual limitations). Several authors within the music information retrieval community (e.g. Futrelle 2002, Uitdenbogerd, 2002) have been commenting on the need for user-centered approaches.

In user-oriented music information retrieval research, distinct levels of user involvement may be considered. These levels depend on the way the user is being treated in mind during the use of a research method (e.g. algorithm testing). User involvement therefore ranges from being passively to being actively involved. Passive user involvement relates to just thinking about the fact that the system is going to have users, whereas active user involvement engages users as participants in user experiments (e.g. annotation of music) designed in view of the system development.

User-oriented studies recently conducted at Ghent University (IPEM)¹ have been set up mainly from the perspective of active user involvement. The intention has been to provide empirical ground in view of linking between bottom-up and top-down approaches to music information retrieval. Such perspective required the development of a theoretical framework for observing the multiple aspects relevant to person-music interactions.

2 Framework

In context of the Musical Audio Mining (MAMI) project, a user-dependent framework (Leman et al 2002, Lesaffre et al 2003) has been developed. This framework was built on multi-leveled and multi-dimensional taxonomies which specify concept categories that can deal with the broad diversity of how users describe music. A global representation of the description levels of the framework for user-oriented music information retrieval research is presented in Figure 1.

¹ <http://www.ipem.ugent.be/>

STRUCTURE		CONCEPT LEVEL		MUSICAL CONTENT FEATURES				
CONTEXTUAL	GLOBAL DESCRIPTORS	HIGH II	EXPRESSIVE	expression				
				affect- experience				
		HIGH I	STRUCTURAL	melody	harmony	rhythm	source	dynamics
				key profile	tonality cadence	patterns tempo	instrument voice	trajectory articulation
		MID	PERCEPTUAL	successive intervallic pattern	simultane intervallic pattern	beat i o i	spectral envelope	dynamic range sound level
				pitch		time	timbre	loudness
NOT CONTEXTUAL	LOCAL DESCRIPTORS	LOW II	SENSORIAL	periodicity pich	note-duration	roughness	peak neural-energy	
				pitch deviations	onset offset	spectral flux	spectral-centroid	
		LOW I	ACOUSTICAL	fundamental frequency	duration	spectrum	intensity	

Fig. 1. Conceptual framework for user-oriented music information retrieval

Musical content features of the multi-leveled framework are distinguished according to acoustical, sensorial, perceptual, structural and expressive concept levels. Constituent music categories include six elementary classes: melody, harmony, rhythm, timbre (i.e. sound source) dynamics and expression.

The structure distinguishes between two types of descriptors of musically relevant auditory phenomena, namely *local* and *global* descriptors. This distinction is based on the internal representational framework of the IPEM Toolbox (Leman et al., 2001) that reckons with the size of the time frame that content formation has to take into account. Local descriptors are derived from music content within time scales shorter than three seconds, whereas global descriptors are derived from musical context dependencies within time scales of about and beyond three seconds. The threshold boundaries between local and global descriptors are defined by the periphery of places in space or time where quantifiable phenomena flow over into subjective phenomena.

Within this framework both empirical observations and algorithm development can be understood as a part of a coherent whole. The study presented here is situated at the structural and expressive level of the framework. It expands on Leman et al. (2004, 2005). Unlike previous research, where subjects were recruited among university students and stimuli were selected which was assumed to be unknown to the subjects, the idea for the present study was having a sample of users of music information retrieval systems who annotate music with a high degree of familiarity. In the next section a brief overview of the user study is given².

² Details of this investigation are reported in Lesaffre (2005), unpublished PhD (available on request) and in Lesaffre et al. (2006).

3 MIR users study

A large-scale study was designed that consisted of two successive parts. The first part was a large survey on the demographic and musical background of users of music information retrieval systems. The second part was an experiment that collected manual annotations of music from an extensive number of respondents in the survey.

3.1 Global setup

The survey was performed using a self-administering web-based questionnaire and resulted in a dataset that contains information about the personal and musical background of 774 participants. From this group, 92 subjects took part in the annotation experiment. This provided an annotation dataset that contains semantic descriptions (i.e. quality ratings) of 160 music excerpts (30 seconds). The latter were selected from 3021 titles of the favorite music of the participants in the survey. The music stimuli thus reflect the musical taste of the targeted population. 79 out of 92 subjects rated the whole set of 160 musical excerpts which were presented in four sessions that took part in a computer classroom. The experiment was conducted under guidance in groups of maximum ten participants.

3.2 User Survey

The survey aimed at identifying potential users of music information retrieval systems and investigating relationships between variables (e.g. gender, musical expertise). The use of multiple recruitment strategies such as radio interviews attracted a valid cross-section of users.

3.2.1. Global user profile

With 774 participants in the survey a representative sample of the targeted population was reached. It was found that music plays an active role in their lives which is in agreement with the hypothesis that the targeted population consist of active music consumers. According to the findings in the survey, a global profile of the envisaged users of music information retrieval systems could be outlined. The average music information retrieval system users:

- Are younger than 35 (74%).
- Use the Internet regularly (93%).
- Spent 1/3 of Internet time on music related activities.
- Do not earn their living with music (91%).
- Are actively involved with music.
- Have the broadest musical taste between 12 and 35.
- Have pop, rock and classical as preferred genres.
- Are good at genre description.

- Have difficulties assigning qualities to classical music.
- Assign most variability to classical music.

3.2.2 Relationships

Multiple relationships between the categorical variables gender, age, musical background, and musical taste were found. It is for example likely that:

- Of users who cannot sing, 74% are men.
- Of users who can dance very well, 93% are women.
- Of classical music listeners, 70% are music experts.
- Of musically educated users, 86% play an instrument.
- Of users older than 35 years, 74% listen to classical music.

3.3 Annotation experiment

The experiment on annotation of music qualities aimed at finding out how potential users of music information retrieval systems would describe their search intention using semantic descriptors for affect, structure and motion. The focus was on unveiling relationships that could support linking between musical structure and musical expressiveness.

3.3.1 Description model

The annotation experiment used semantic adjectives to describe music qualities. Our model (see Table 1) for rating high-level music qualities basically distinguished between affective/emotive (I), structural (II) and kinaesthetic descriptors (III). Apart from this, for each of the 160 rated musical excerpts, subjects were also asked to give additional information on how familiar they were with the music they heard (IV) and what was their personal judgment (V).

Table 1. Model for semantic description of music

I. AFFECTIVE/ EMOTIVE	II. STRUCTURAL	III. KINAESTHETIC
I.1 Appraisal	II.1 Sonic	gesture
Cheerful	Soft/hard	imitation
Sad	Clear/dull	
Carefree	Rough/harmonious	IV. MEMORY
Anxious	Void/compact	No recognition
Tender	Slow/quick	Style recognition
Aggressive	Flowing/stuttering	Vaguely known
Passionate	Dynamic/static	Well known
Restrained	II.2 Pattern	

Most typical	Timbre	V. JUDGMENT
I.2 Interest	Rhythm	Beautiful/awful
Annoying	Melody	Difficult/easy
Pleasing	None	
Touching		
Indifferent		

3.3.2 Results

Influence of subject related factors was found for gender, age, musical expertise, broadness of taste, familiarity with classical music and active musicianship. It was found that men rated the musical excerpts more restrained, more harmonious and more static whereas women judged the music more beautiful and more difficult. Subjects older than 35 found the music more passionate and less static than younger listeners did. Lay listeners judged the music as being more cheerful, passionate and dull than experts did. Equal results were found for the influence of musicianship. People with a broad musical taste judged the music to be more pleasing and more beautiful than those with a narrow taste. Familiarity with the music is highly significant for all affective/emotive descriptors.

Factor analysis revealed that several affective/emotive descriptors are correlated. For affective /emotive adjectives the 12 dimensional description model was reduced to three dimensions which are described as *high intense experience*, *diffuse affective state* and *physical involvement*. These factors are closely related to the dimensions *Interest*, *Valence* and *Activity* uncovered in previous research (Leman et al., 2005).

Variable reduction of the structural descriptors also revealed three dimensions. With regard to unanimity among semantic descriptors, adjectives were tested that relate to loudness, timbre, tempo and articulation. Subjects agreed most on loudness and tempo, whilst less on timbre and articulation.

Interesting relationships were found between affective/emotive and structural descriptors. There is a strong correlation between the appraisal descriptor (tender-aggressive) and the structural descriptor loudness (soft-hard). This result is suggestive of the possibility to decompose semantic descriptors in terms of structural descriptors, which mediate the connection with acoustical descriptors.

4 Semantic music recommendation tool

For validating the results of the study on users of music information retrieval systems and on the semantic description of music a research tool has been developed. The latter is conceived as a semantic music recommender system for conducting tests in the real world.

There are two reasons why a validation tool in the form of a prototype of a semantic music recommender system was designed. The first reason was the objective of investigating whether another population which is distinct from the one in the study

could agree with the judgments from the latter. The second reason concerned testing of user-friendliness and usability of a semantic music recommender system based on affective/emotive, structural and kinaesthetic descriptors.

4.1 Design

The design of the semantic music recommendation system is based on the idea of using fuzzy logic. The integration of fuzzy logic is an interesting option because the subjective character of vague concepts is taken into account. The system incorporates the annotations (i.e. quality ratings) of the participants in the experiment on semantic description of music.

The interface of the semantic music recommender demonstration was designed for multiple testing possibilities (e.g. use at exhibitions) which address different populations. The validation tool basically consists of four parts: (1) definition of the user profile; (2) presentation of the input options; (3) recommendations of music and (4) evaluation tasks.

The interaction paradigm is the following: a user provides input (i.e. profile and query) and the system processes that information to generate a ranked list of music recommendations. Profile specification relates to subject dependencies such as gender and musical interest. Our study has shown that these factors explain differences in the perception of high-level features.

In the search screen four selection fields are presented that allow any combination of choices between five genre categories (classical, pop/rock, folk/country, jazz and world/ethnic), eight emotion labels (cheerful, sad, tender, passionate, anxious, aggressive, restrained and carefree), four adjective pairs referring to sonic properties of music (soft-hard, clear-dull, rough-harmonious and void-compact) and three adjective pairs reflecting movement (slow-quick, flowing-stuttering and dynamic-static).

The output is a hierarchically ordered list with music titles. The user can browse the list and listen to the music. Each time a user listens to a recommended piece of music a popup window provides the user with individual scores for each descriptor in the query. These scores reflect the agreement among the participants in the experiment.

Two assessment tasks are included in the demo. First the user is requested to assign a degree of satisfaction after having listened to a recommended piece of music. The second task involves evaluation of the usability of emotion-based querying and of the semantic descriptor sets (i.e. expression, structure, motion).

4.2 User test

The semantic music recommender system was tested by 626 visitors at ACCENTA 2005³. Together they listened to 2993 music recommendations and together they

³ ACCENTA is Flanders' international annual fair in Ghent that celebrated its 60th anniversary in 2005 (September 17-25). The prototype on music and emotion was one of the demonstrations illustrating the research activities at the department of musicology (IPEM)

selected 18415 adjectives. In Table 2 semantic descriptors are sorted by the number of responses. Affective/emotive, structural and kinaesthetic descriptors as well get high ranking.

Table 2. Preferred semantic descriptors

Descriptor	Number	Descriptor	Number
cheerful	1764	not sad	551
bright	1271	sad	517
flowing	1247	slow	458
passionate	1233	compact	405
dynamic	1134	restrained	380
soft	1048	stuttering	323
harmonious	893	rough	285
tender	843	anxious	271
hard	837	not carefree	240
quick	829	not tender	234
carefree	649	void	223
not anxious	592	static	168
not restrained	570	not passionate	130
aggressive	554	dull	124
not aggressive	552	not cheerful	90

From observation of the people using the system we learned that they enjoyed discovering new music by entering emotion-based queries.

Analysis of the satisfaction ratings has shown that around three quarter of the users were very satisfied of the fit between their query and the recommendations made by the system. With regard to the usability of the semantic descriptors, affect/emotive and kinaesthetic descriptors are found useful by 79% of the participants whereas structural descriptors by 70% of the participants. Over 90% of the participants responded positively to the overall usability of the system.

5 Conclusion

The present study shows that a user-oriented approach to music information retrieval which focuses on active user involvement provides evidence for the use of semantic descriptors as a means to access music. The study reveals that the framework of linguistic-based semantic descriptors has an inter-subjective basis. Using the profile information collected in a large survey, analysis of the influence of subject related factors revealed subject dependencies for gender, age, expertise, musicianship broadness of taste and familiarity with classical music. Apart from this familiarity with the musical piece showed to have the highest significant effect on all semantic

descriptors. Music search and retrieval systems should therefore distinguish between different categories of users.

Our findings on how users of music information retrieval systems perceive music qualities have been directly confirmed by a test in the real world of a semantic music recommender system that reflects the degree to which users agree about music qualities. Positive user experience has shown that the semantic framework of affective/emotive, structural and kinaesthetic descriptors can easily be used to formulate a search intention.

Acknowledgements

This research has been conducted in the framework of the Musical Audio Mining (MAMI) project, which is funded by The Flemish Institute for the Promotion of Scientific and Technical Research in Industry. The authors wish to thank MA Frank Desmet and MA Kurt Vermeulen for their assistance with the development of the query builder and the demonstration tool.

References

Downie, J. S. (2004). The creation of music query documents: framework and implications of the HUMIRS project. In *Proceedings of the Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (ALLC/ACH)*, Göteborg.

Futrelle, J., & Downie, J. S. (2002). Interdisciplinary communities and research issues in Music Information Retrieval. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR02)*, Paris, 215-221.

Lee, J. H., & Downie, J. S. (2004). Survey of music information needs, uses and seeking behaviours: preliminary findings. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR04)*, Barcelona, 441-448.

Leman, M., Clarisse, L., De Baets, B., De Meyer, H., Lesaffre, M., Martens, G., Martens, J., and Van Steelant, D. (2002) Tendencias, perspectivas, and opportunities of musical audio-mining. In A. Calvo-Manzano, A. Pérez-López, & J. S. Santiago (Eds.), *Forum Acusticum Sevilla 2002*, 16-20 september, 2002. Madrid: Sociedad Española de Acustica - SEA.

Leman, M., Vermeulen, V., De Voogdt, L., Taelman, J., Moelants, D., & Lesaffre, M. (2004). Correlation of gestural musical audio cues and perceived expressive qualities. In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction* (40-54). Berlin Heidelberg: Springer-Verlag.

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesaffre, M. (2005). Prediction of Musical Affect Attribution Using a Combination of Structural Cues

Extracted from Musical Audio. *Journal of New Music Research*, 34(1), 39-67.

Lesaffre, M., (2005). Music Information Retrieval. Conceptual framework, Annotation and User Behaviour. Unpublished PhD.

Lesaffre, M., De Voogdt, L., Leman, M., De Baets, B., De Meyer, H., & Martens J.-P., (2006). How potential users of music search and retrieval systems describe the semantic quality of music. (Submitted)

Lesaffre, M., Leman, M., De Voogdt, L., De Baets, B., De Meyer, H., & Martens J.-P., (2006) A user-dependent approach to the perception of high-level semantics of music in *Proceedings of the International Conference on Music Perception and Cognition (ICMPS)*, Bologna

Yang, D., & Lee, W. (2004). Disambiguating Music Emotion Using Software Agents. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR04)*, Barcelona, 52-58