

A user-dependent approach to the perception of high-level semantics of music

Micheline Lesaffre¹, Marc Leman¹, Liesbeth De Voogdt¹, Bernard De Baets², Hans De Meyer³ and Jean-Pierre Martens⁴

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 Applied Mathematics and Computer Science, Ghent University.

⁴ Department of Electronics and Information Systems (ELIS), Ghent University.

ABSTRACT

The maturing of music information retrieval (MIR) systems outlines an attractive future for emotion-based retrieval of music. The present paper reports the results of an elaborate study which explores (1) who potential users of MIR systems are, (2) how they perceive affects in music, and (3) what structural descriptions of music best characterize their understanding of music expression. 79 potential users of music information retrieval systems rated sets of adjectives, while they were listening to 160 music excerpts. The stimuli reflect the musical taste of the average participant in a large survey on the demographic and music background of people who are interested in using interactive music systems. The subject group (79) in the annotation experiment was recruited amongst the 774 participants in the survey. The study reveals that perceived qualities of music are affected by the profile of the user. Significant subject dependencies are found for age, music expertise, musicianship, broadness of taste and familiarity with classical music. Interesting relationships are discovered between expressive and structural features. Analyses show that the targeted population most unanimously agrees on loudness and tempo, whilst less unanimity was found for timbre and articulation. A semantic music recommender system is presented that was developed for validating the experimental results in the real world. A test has demonstrated the potential of a user-dependent and emotion-

based retrieval of music.

Keywords

Music Information Retrieval, user profile, semantic description, music recommendation, search by emotion.

1 INTRODUCTION

With the development of musical search and retrieval technology, the interaction between listener and musical content meets new relationships. It is widely accepted that apart from the standard search options ‘title’, ‘album’ or ‘artist’, new query methods should be offered that are more natural (Baumann et al. 2002; Birmingham et al. 2002). Users wish to find music that is similar to other music, music that has a certain emotional content or music in a specified style (Huron & Aarden, 2002). Such very subjective ways of accessing music highlight the need for investigating the relations between semantic description and subjective background. Although indexing and retrieval techniques are being investigated all the time, it has not been thoroughly researched how a given group of users describe high-level semantics according to their perception of emotion and structure in music.

This paper is organized in four sections. The first section regards the context of user studies in the domain of music information retrieval. The second section addresses an elaborate study both on music information retrieval user’s background and on their semantic description of music. A brief summary of the results is given.¹ The validation of these results is discussed in the third section which also includes the presentation of a test tool that explores emotion-based semantic music recommendation. Finally, in the fourth section, conclusions are drawn from the analysis, interpretation and evaluation.

Proceedings of the 9th International Conference on Music Perception & Cognition (ICMPC9). ©2006 The Society for Music Perception & Cognition (SMPC) and European Society for the Cognitive Sciences of Music (ESCOM). Copyright of the content of an individual paper is held by the primary (first-named) author of that paper. All rights reserved. No paper from this proceedings may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information retrieval systems, without permission in writing from the paper’s primary author. No other part of this proceedings may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information retrieval system, with permission in writing from SMPC and ESCOM.

¹ Details of this investigation are reported in Lesaffre (2005), unpublished PhD (available on request) and in Lesaffre et al. (2006).

2 CONTEXT

In music information retrieval research, opinion and attitude questions still need a lot of investigation. To the best of our knowledge, no large-scale studies that aim at monitoring the effects of demographic and musical background on semantic description have been reported yet.

Within the domain of music information retrieval, studies using both a population that is representative for the music information retrieval context and musical stimuli that are representative for the music consumption pattern of this population are inexistent. As a consequence there is a lack of knowledge about the degree to which users can successfully deal with the semantic description of music.

Existing user studies tend to be small-scale and rely on a population whose representativeness is not identified. Lee and Downie (2004) for example, have been reporting results for a population of users that was limited to a university community. Apart from this, only little work has been published on the analysis of the real-world needs and uses of digital music databases. In view of the planning for a music information retrieval test bed, Cunningham (2002) emphasizes the need to ascertain who the potential users of a music collection are.

In pursuit of system evaluations and algorithm testing, studies have been set up that collect user's music ratings (e.g. Yang and lee, 2004). Unfortunately, most of these studies provide scarce reference material in terms of how the ratings were obtained, how representative the population of users was and use a limited musical database.

A number of studies have explored the relationship between descriptions of emotional appraisal of music and musical structure (Juslin and Sloboda 2001, Gabrielson and Juslin 2003). However, these studies are not often related to music information retrieval. Although within the community more input from psychologist and musicologists has been requested (Futrelle and Downie 2002), the relations between semantic description and subjective background still are not well understood.

Leman et al. (2005) investigated the hypothesis that the inter-subjective basis of semantic descriptors can be predicted by a combination of acoustical cues. In this study affect attribution originated from university students while structural content was judged by music experts. The promising results of this study call for large-scale studies involving a more representative population and music that is representative of their taste.

3 EXPERIMENTS

The study presented here expands on Leman et al (2004, 2005) both in terms of scale and approach. Unlike previous research which involved university students and a broad range of music excerpts, in current study the subjects were recruited from the real world and it was them who provided the titles of the music stimuli that they annotated. The study consists of two parts. One part is a survey on the demo-

graphic and musical background of the potential users of interactive music systems. The other part is an annotation experiment involving a representative set of participants who took part in the survey.

3.1 Experimental set up

The survey was performed using a self administering web-based questionnaire and resulted in the *main dataset* (see Figure 1). It contains information about the personal and musical background of the 774 participants, including 3021 titles of their favorite music. From this main dataset, 92 subjects and 160 musical excerpts were selected in view of the annotation experiment. The latter provided the *annotation dataset* that contains semantic descriptions (i.e. quality ratings) of the selected music excerpts. The large quantity of data is incorporated in a relational database (RDB). A query builder has been developed as a research tool to provide easy access to the database.

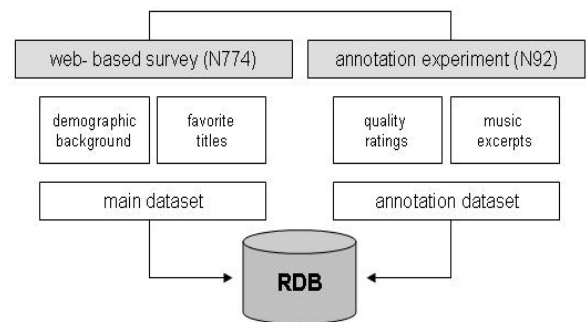


Figure 1. Global set up of the study.

3.2 Demographic and musical background

The first part of the study entailed a self-administering survey that aimed at identifying potential users of content-based search and retrieval of music. Multiple recruitment strategies, such as the launch of a media campaign, were applied in order to attract a valid cross-section of potential end-users. Participants were requested to fill in questionnaires that addressed the following objectives:

- Socio-demographic background.
- Cultural background.
- Familiarity with the internet & use of internet.
- Music education level.
- Skills to play an instrument & instrument type.
- Singing and dancing skills.
- Ways of interacting with music.
- Preferred media for listening.
- Genre preference.
- Evolution of taste.
- Favorite music.

3.3 Semantic description of music

The second part of the study involved an experiment on the annotation of high-level qualities of music. Semantic description of music relates to the verbal description of intrinsic music qualities as perceived by a user. The experiment on the annotation of music qualities aimed at finding out how potential users of music information retrieval systems would describe their search intention using semantic descriptions of music. 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 expressive and structural qualities of music. 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. Semantic description model

I. AFFECTIVE/ EMOTIVE	II. STRUCTURAL
I.1 Appraisal	II.1 Sonic
Cheerful	Soft/hard
Sad	Clear/dull
Carefree	Rough/harmonious
Anxious	Void/compact
Tender	Slow/quick
Aggressive	Flowing/stuttering
Passionate	Dynamic/static
Restrained	II.2 Pattern
Most typical	Timbre
I.2 Interest	Rhythm
Annoying	Melody
Pleasing	None
Touching	
Indifferent	III. KINAESTHETIC
	Gesture
IV. MEMORY	Imitation
No recognition	
Style recognition	V. JUDGMENT
Vaguely known	Beautiful/awful
Well known	Difficult/easy

3.3.2 Stimuli

Out of 3021 titles of favorite music provided by the participants in the survey, 160 titles were selected for the crea-

tion of a music database containing musical excerpts of 30 seconds duration for the annotation experiment. This database includes a wide-ranging selection of genres and styles that are representative of the music consumption pattern of the participants in the survey. The music stimuli thus reflect the musical taste of the targeted population.

3.3.3 Procedure

The annotation experiment took place in four sessions each covering 40 excerpts of music. As each fragment had to be annotated for all the features involved (see Table 1) 40 excerpts per session was considered the maximum. The experiment was conducted in groups of utmost ten participants who performed the test under guidance. The sessions took place in a computer classroom, where the subjects sat in front of a PC, while the music was being played through headphones. The order of the music excerpts was randomized.

3.4 Summary of results

3.4.1 Global profile of MIR users

With 774 participants representing a broad distribution of both music lovers and technology minded people we have reached a sample size that is large enough to permit quantitative and qualitative deductions. In addressing a population of potential users of music information retrieval systems, it was found that music plays an active role in their lives. This 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 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.

3.4.2 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.

3.4.3 Semantic description

The results reported below account for the semantic descriptions by 79 subjects (out of 92) who rated the entire set of 160 music fragments.

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 VALIDATION

For validating the results of the study on users of music information retrieval systems and on the semantic description of music, three strategies were used. First, populations were compared; second consistency tests were carried out and third a research tool was developed. The latter was conceived as a semantic music recommender system for conducting tests in the real world.

4.1 Comparison of populations

Comparison was made between demographic and musical background of the participants in the survey (774) and those who took part in the experiment (92). This analysis confirmed our hypothesis that the subjects in the experiment are a representative sample of the targeted population

of potential users of music information retrieval systems. Only minor differences were found and none were of a kind that they would warrant the questioning of the results.

4.2 Consistency tests

The reliability of the semantic descriptions was checked by means of consistency tests which were carried over time. The outcome of two repeated follow-up experiments was compared with the results of the original experiments. Although the ratings were quite reliable some terms appeared to be ambiguous. Affective/emotive descriptors such as anxious, aggressive and passionate can be indistinct when measurements are repeated. There was also little inter-subjective agreement on structural descriptors that relate to brightness (clear-dull) and roughness (rough-harmonious).

4.3 User satisfaction

There are two reasons why a validation tool in the form of a prototype of a *semantic music recommender system* was developed. The first reason is the objective of investigating whether another population that is distinct from the one in the study can agree with the judgments from the latter. The second reason concerns testing of user-friendliness and usability of a semantic music recommender system.

4.3.1 Approach

The presented semantic music recommendation system uses fuzzy logic in order to account for the subjective character of the semantic descriptions of music qualities. The semantic music recommender system incorporates the annotations (i.e. quality ratings) of the participants in the experiment on semantic description of music. Our study has shown that when attempting to define what quality features constitutes a piece of music, individual perceptions and backgrounds have to be taken into account. From that perspective, the integration of fuzzy logic in the relational database is an interesting option because it bypasses the sharp-edged true-false logic of semantic descriptors. In practice this means that fuzzy logic allows a semantic description to have any value between the numerical value of 1 (true) and the value of 0 (false).

4.3.2 Design and procedure

The interface of the semantic music recommender demonstration was designed for use at exhibitions and for multiple testing possibilities 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 care-free), 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.

4.3.3 Evaluation tasks

Two assessment tasks are included in the demo: (1) assignment of the degree of satisfaction after the user has listened to a recommended piece of music and (2) evaluation of the usability of emotion-based querying and of the semantic descriptor sets.

4.3.4 Results

The system was tested by 626 visitors at ACCENTA 2005². Together they listened to 2993 music recommendations and together they 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

² 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)

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

Each time a user listens to a recommended piece of music he/she is invited to rate the degree of agreement between the query and the music recommendation. The query output (see Figure 2) provides the user with individual scores for each descriptor in the query. These scores reflect the agreement among the participants in the experiment. 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.

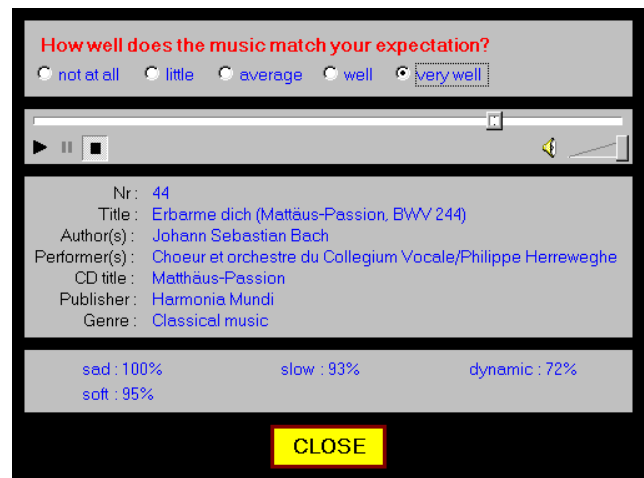


Figure 2. User satisfaction: window with media player, query output and satisfaction rating.

5 CONCLUSION

The present study shows that a user-dependent approach to the perception of high-level features of music provides evidence for the use of semantic descriptors as a means to access music. Users are able to give a cognitive assessment of music in terms of linguistic-based semantic descriptors. The semantic description model that was used distinguished between affective/emotive, structural and kinaesthetic descriptors. The study reveals that the semantic framework has an inter-subjective basis even if demographic and musical background has an impact too. It was found that genders, age, expertise, musicianship, broadness of taste are influential subject-related factors. 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.

In order to track reliability of the semantic description model three validation strategies were used which all of them confirm the results of the study. It can be assumed that semantic description of music may provide a stable basis for further development of content-based access to music. A prototype of a semantic music recommender system that was tested in the real world showed that this semantic framework can easily be used to formulate a search intention. Positive user experience confirmed the usability of music information retrieval systems that are able to deal with queries that are related to music qualities pertaining to emotion, structure and movement.

ACKNOWLEDGMENTS

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

- Baumann, S., Klüter, A., & Noriel, M. (2002). Using natural language input and audio analysis for a human-oriented MIR system. In *Proceedings of the International Conference on Web Delivering of Music (Wedelmusic)*, Darmstadt
- Birmingham, W., Pardo, B., Meek, C., & Shifrin, J. (2002). The MusArt Music-Retrieval System: an overview. *D-Lib Magazine*, 8(2).
- Cunningham, S. J. (2002). User studies: a first step in designing a MIR testbed. In *The MIR/MDL Evaluation Project White Paper Collection*, 2nd ed., 17-19.
- 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.
- Huron, D., & Aarden, B. (2002). *Cognitive issues and approaches in music information retrieval*. In S Downie and D Byrd, editors, Music Information Retrieval.
- 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.
- Juslin, P. N., & Sloboda, J. A. (2001). *Music and Emotion: Theory and Research*. New York, Oxford: Oxford University Press.
- 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)