Using fuzzy logic to handle semantic descriptions of music in a content-based retrieval system

Micheline Lesaffre and Marc Leman

IPEM, Institute for Psychoacoustics and Electronic Music, Department of Musicology, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium
Micheline.Lesaffre@UGent.be

Abstract. This paper explores the potential use of fuzzy logic for semantic music recommendation. We show that a set of affective/emotive, structural and kinaesthetic descriptors can be used to formulate a query which allows the retrieval of intended music. A semantic music recommendation system was built, based on an elaborate study of potential users and an analysis of the semantic descriptors that best characterize the user's understanding of music. Significant relationships between expressive and structural semantic descriptions of music were found. Fuzzy logic was then applied to handle the quality ratings associated with the semantic descriptions. A working semantic music recommendation system was tested and evaluated. Real-world testing revealed high user satisfaction.

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

1   Introduction

Research on content-based music information retrieval aims at defining the search and retrieval of music in terms of musical content descriptors. Rather than having to specify the name of the composer or the title of the song, the content-based approach would allow one to specify musical content using semantic descriptors such as `happy', 'sad', 'dynamic' and 'harmonious'. However, one of the weaknesses in content-based music information retrieval research is that most often, there is a lack of knowledge about the user's background, such as education, gender, familiarity with music and so on. Semantic descriptions are meant to function in a social context and such descriptions focus on high-level properties, whose semantics range from structural to kinaesthetic to affective/emotive qualities [1, 2]. In this context, the meaning of semantic descriptors is often determined by tacit knowledge about the user's intentions, the user's background and the common cultural context in which the communication is taking place. In the context of content-based music information retrieval, this tacit knowledge is often absent. Semantic descriptors of music are used to mediate between the user's verbally described search intention and the audio contained in a music library but the system often lacks the tacit knowledge about

intentions, background and common cultural context. As a result, there is a semantic gap between user and system.

Up to now, most solutions are based on systems that correlate extracted audio features with semantic descriptors, using techniques based on probabilistic learning methods (e.g. [3]) However, the success of such mappings often depends on the homogeneity of the users involved. Users may group into categories that apply particular semantic descriptors, or users may use semantic descriptors in a particular way, depending on subjective factors such as education and gender. Therefore, content-based music search and retrieval cannot be fully accomplished when the particularities of users are not taken into account. What is needed is (1) a better definition of the users of such systems, (2) better and more elaborate databases with semantic annotations of music, (3) better tools for handling flexible processing of semantic descriptions and (4) better tools for system evaluation.

This paper consists of four parts. In the first part a brief overview is given of related work on semantic description of music. The second part addresses the user study that preceded the development of the semantic music recommendation system. This study aimed at investigating the relationships between the user's background and semantic description of music. The third part applies fuzzy logic to flexible querying. Finally, in the fourth part, the semantic music recommendation system is evaluated and tested in the context of a large exhibition.

2   Background

During the last decade, the fuzzy logics field has witnessed a tremendous growth in the number of applications across a wide variety of domains that focus on humanlike behavior. It is possible that in the near future, the Semantic Web will be a major field of applications of fuzzy logic [4]. However, to the best of our knowledge, there are no music recommendation systems available that use fuzzy logics to handle semantic descriptions of affective/emotive, structural and kinaesthetic features of music[1].

Usability is a topic of interest in the musical digital library community. Although the importance of interface and system usability is acknowledged (e.g. [5]), it has only recently been suggested that users themselves should be consulted. Previous studies rather focus on trying to find out what people do and would like to do with music. These studies involve, for example, analyzing music queries posted to Usenet News [6] and to the Google ask-an-expert service [7] or watching people's behaviour in CD stores [8]. The usability of existing systems and various methodologies, however, has not been tested with real music information retrieval users. Indeed, the most common method used for studying usability is laboratory-based testing [9].

So far, the use of semantic descriptors for music is based on two approaches. Linking approaches aim at collecting the users' descriptions of music in application

---

[1] Existing commercial music recommender systems such as Amazon, MediaUnbound, MoodLogic, or Songexplorer use semantic descriptors but these systems are often limited, not very refined in terms of semantic descriptors; they often differ a lot in the kind of input that users must provide. In addition, the underlying mechanisms are badly documented.

contexts. Kalbach [10] praises the innovative character of these linking approaches, because they are based on a large population of users dedicated to search and retrieval of music. Yet the semantic description often relies on an ad hoc taxonomy (e.g. MoodLogic, http://www.moodlogic.com/). In contrast, annotation approaches collect the user's description of music in pursuit of system evaluations and algorithm testing (e.g. [11], [12], [13]). Unfortunately, most studies provide scarce reference material in terms of how these ratings were obtained, and how representative the population of users was, despite requests for more input from psychologists and musicologists [14].

A specific field of interest concerns the relationship between different categories of semantic description. In this context, a number of studies have explored the relationship between descriptions of musical structure and descriptions of emotional appraisal (e.g. [15], [16], [17]). The latter form an important sub-category of the category of semantic descriptions. Most studies reveal that semantic/emotive descriptors rely on a number of subjective factors. Yet, these studies are often not related to music information retrieval and therefore they suffer from a lack of representative population and musical excerpts.

The present research expands on earlier studies carried out by Leman et al. [1], [18]. In these studies, descriptions of emotional and affect appraisal of music were collected from a group of university students, while descriptions of musical structural were collected from a group of musicologists. These studies have been expanded by recruiting and involving a large set of users that are potentially interested in content-based music information retrieval.

3   Foregoing users study

A large-scale study has been set up, which contained two parts. In the first part, a survey of the demographic and the musical background of potential users of music information retrieval systems was carried out. In the second part, a representative set of users was asked to annotate music using semantic descriptions. The study provided a large database that was then used to build a semantic music recommendation system in which the demographic and musical background of the user was taken into account.

3.1   Global setup

The survey was performed using a self administering web-based questionnaire. This survey resulted in a dataset with information about personal demographic and musical background of 774 participants. From this group, a sample of 92 subjects was recruited for the annotation experiment. This provided an annotation dataset with semantic descriptions (i.e. quality ratings) of 160 music excerpts. The latter were selected from 3021 titles of the favorite music of the participants in the survey. The music stimuli, having a duration of 30 seconds, thus reflected the musical taste of the targeted population. 79 out of 92 subjects rated the whole set of 160 musical excerpts (see Lesaffre, 2005 for a more detailed description of the experiment).

3.2 User survey

3.2.1. Global user profile

The 774 participants of the survey, a representative sample of the targeted population was reached and global profile of the envisaged users of content-based music information retrieval systems could be defined. The average music information retrieval system users: are younger than 35 (74%); use the Internet regularly (93%); spend 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 and 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 was for example found 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

3.3.1 Description model

In the annotation experiment, a representative population of 92 users, described music using a set of semantic adjectives. Our model (see Table 1) 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).

3.3.2 Results

There was a significant influence of demographic and musical background such as gender, age, musical expertise, broadness of taste, familiarity with classical music and active musicianship on the use of semantic descriptors. For example, 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

Table 1. Model for the 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		

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. The above results led to a categorization of users in four different groups, based on education (musical and non-musical) and gender (male and female).

Factor analysis revealed that several affective/emotive descriptors were correlated and that three dimensions may account for a large proportion of the variance, namely *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). In a similar way, the structural descriptors also revealed three dimensions. With regard to unanimity among the descriptors 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 system

A semantic music recommendation system was built using the results of the user study. The system incorporates the annotations, that is, the quality ratings of semantic descriptors, of the participants in the experiment. In this context, fuzzy logic was considered as a possible option to account for vague descriptors. In the present study vagueness arose from the quality ratings of semantic descriptors, which used concepts

like "rather", "moderate" and "very". Using these terms, a musical excerpt could be characterized as being 'rather sad' or 'very sad'.

4.1 Design and procedure

An interface of the semantic music recommender tool was designed for use at exhibitions and other testing environments that address different user populations. The tool basically consists of four parts: (1) definition of the user profile (gender and musical interest); (2) specification of the search query using semantic descriptors; (3) recommendation of music, using the music database and (4) evaluation tasks.

The search screen presents four categories of semantic descriptors, allowing any combination of choices between (1) five genre categories (classical, pop/rock, folk/country, jazz and world/ethnic), (2) eight emotion labels (cheerful, sad, tender, passionate, anxious, aggressive, restrained and carefree), (3) four adjective pairs referring to sonic properties of music (soft-hard, clear-dull, rough-harmonious and void-compact) and (4) 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.

In addition to the recommendation system, two assessment tasks are included (see below, Real-world testing). First, the user is requested to assign a degree of satisfaction in using the system for this particular search task, after having listened to a recommended piece of music. Secondly, the user is request to evaluate of the general usability of content-based querying.

4.2 Fuzzy logic functions

Fuzzy logic was applied in three steps. Firstly, fuzzy functions, which account for the vagueness of the semantic descriptors, were calculated per semantic descriptor and per user profile. Secondly, scores were calculated per music excerpt, semantic descriptor and user profile. Thirdly, combined scores were calculated.

Fuzzy functions per semantic descriptor and user profile.
The semantic music recommendation takes into account two different types of users' background, namely gender (male, female) and musical expertise (expert, novice). As a consequence, for each adjective, four fuzzy functions were calculated, according to four different user categories These functions were characterized by three numbers, namely, the 25th, 50th and 75th percentile values, representing the cumulative rating value of the user category for a semantic descriptor. To obtain that function, for each adjective, the rating values attributed by all the subjects who fit a specific profile (i.e. female novice, female expert, male novice and male expert) were sorted in ascending order. After that, the values according to the cumulative percentages of 25%, 50% and

75% were calculated. These three values each define a fuzzy function score. Then, the cumulative distribution function is built on the number of ratings given by a user group for a semantic descriptor for five data points (i.e. not, little, moderate, rather, very). From this discrete set of data points a new fuzzy function is built on a set of three data points (i.e. the three fuzzy function scores).

Calculation of recommendation scores per music excerpt, adjective and profile
In order to determine the recommendation scores per music excerpt, to begin with, the rating values attributed by all subjects who fit a specific user profile, were sorted in ascending order as to semantic descriptor and excerpt number respectively. After that, the cumulative median value was calculated. The score for each adjective, profile and excerpt resulted in the following function value: *score(median)* with score being the fuzzy function corresponding to the adjective and profile concerned.

Calculation of combined recommendation scores per music excerpt
If no adjectives are selected, the combined recommendation score is 1. If one adjective is selected then the combined recommendation score equals the score for the semantic descriptor concerned. If multiple (n) adjectives are selected then the combined recommendation score equals the nth power of the product of the adjective scores.

5   Real-world testing

The semantic music recommendation system was tested in a real-world environment during an exhibition in Ghent, Belgium. The testing aimed at investigating whether another population (i.e. general public), distinct from the one in our study (i.e. MIR users), would agree with the judgments made by users recruited in our study. The test aimed at evaluating the effectiveness of the fuzzy logic approach applied to semantic descriptors and the usability of the system.

The semantic music recommender system was tested by 626 visitors at ACCENTA 2005[2]. These visitors listened to 2993 musical excerpts. All 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.

A qualitative observation of the people that used the system learned that they enjoyed discovering new music by entering semantic-based queries. Quantitative analysis of the satisfaction ratings shows that about three quarter of the users were very satisfied about the recommendation. 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.

---

[2] ACCENTA is Flanders' international annual fair. The semantic music recommendation tool was one of the demonstrations illustrating the research activities at Ghent University (IPEM, department of musicology)

Table 2. Preferred semantic descriptors.

Descriptor	Number of times selected	Descriptor	Number of times selected
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

6   Conclusion

In this paper we described the development and testing of a content-based music information retrieval system that uses fuzzy logics for handling the vagueness of semantic descriptors. Our results show that a fuzzy logics methodology, combined with a user-oriented approach to music information retrieval, may be effective for the development of a music recommendation system. The study reveals that the framework of affective/emotive, structural and kinaesthetic descriptors has an inter-subjective basis whose vagueness can be handled with fuzzy logics.

The system has been tested in a real-world environment. 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. Users confirmed the usability of semantic-based music information retrieval systems. It can be assumed that applying fuzzy logic to semantic descriptors may provide a stable basis for further development of content-based access to music. In the near future tests will be carried out involving other populations and a larger music database.

Acknowledgements

References

1. 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.
2. Lesaffre, M., (2005). Music Information Retrieval. Conceptual framework, Annotation and User Behaviour. Unpublished PhD.
3. Whitman, B., "Learning the meaning of music," Ph.D. dissertation, Department of Architecture, Massachusetts Institute of Technology, 2005.
4. Straccia, U., *A fuzzy description logic for the semantic web*. In Sanchez, E., ed.: Capturing Intelligence: Fuzzy Logic and the Semantic Web. Elsevier (2005)
5. Arms, W. Y. (2000). Digital Libraries. Cambridge, MIT Press
6. Downie, J. S., and Cunningham, S.J. (2002) Toward a theory of music information retrieval queries: system design implications. Proceedings of the International Symposium on Music Information Retrieval (ISMIR02), Paris (France), 299 - 300.
7. Bainbridge, D., Cunningham, S. J., & Downie, J. S. (2003). How people describe their music information needs: a grounded theory analysis of music queries. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR03)*, Baltimore, 221-222.
8. 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.
9. Covey, D. T. (2002). *Usage and usability assessment: library practices and concerns.* Washington, DC: Digital Library Federation and Council on Library and Information Resources. (CLIR Report105). Retrieved October 15, 2003, from http://www.clir.org/pubs/reports/pub105/contents.html
10. Kalbach, J. (2002). Classifying emotion for information retrieval: three websites. *Notes, 59*(2), 408-411.
11. Tzanetakis, G., & Cook, P. (2000). *Experiments in Computer-Assisted Annotation of Audio.* Paper presented at the International Conference on Auditory Display (ICAD), Atlanta, Georgia, USA.
12. 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
13. Lesaffre, M., Leman, M., De Baets, B., & Martens, J.-P. (2004). Methodological Considerations Concerning Manual Annotation of Musical Audio in Function of Algorithm Development. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR04),* Barcelona, 64-71.
14. 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.

15. Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. J. Davidson & H. H. Goldsmith & K. H. Scherer (Eds.), *Handbook of affective sciences*. New York: Oxford University press, 503-534.
16. Gabrielsson, A., & Lindström, S. (2001). The Influence of Musical Expression on Emotion in Music. In P. Juslin & J. Sloboda (Eds.), *Music and Emotion: Theory and Research*. New York: Oxford University Press, 223-248.
17. Juslin and Sloboda Juslin, P. N., & Sloboda, J. A. (2001). *Music and Emotion: Theory and Research*. New York, Oxford: Oxford University Press.
18. 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.