Using fuzzy logic to handle the users' semantic descriptions in a music 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
Marc.Leman@UGent.be

Abstract. This paper provides an investigation of the potential application of fuzzy logic to 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 of music information retrieval systems. In this study analysis was made of the descriptors that best characterize the user's understanding of music. Significant relationships between expressive and structural descriptions of music were found. A straightforward fuzzy logic methodology was then applied to handle the quality ratings associated with the descriptions. Rigorous real-world testing of the semantic music recommendation system 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 semantic descriptors. Rather than having to specify the name of the composer or the title of the song, semantic description would allow one to specify musical content using descriptors such as `happy', 'sad', 'dynamic' and 'harmonious'. Such descriptions focus on high-level properties, whose meaning ranges from structural to kinaesthetic to affective/emotive qualities of the music [1, 2]. However, one of the weaknesses of this approach is that most often, there is a lack of knowledge about the user's background, such as education, gender, familiarity with music. Semantic descriptions are meant to function in a social context and 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. As a result, there is a semantic gap between user and system. Semantic descriptors of music are meant to mediate between the user's verbally described search intention and the audio contained in a music library, yet the

system lacks the tacit knowledge about intentions, background and common cultural context.

Up to now, most solutions to the problem of the semantic gap are based on systems that correlate extracted audio features with semantic descriptors, using techniques based on probabilistic learning methods (e.g. [3]). However, such mappings often assume the homogeneity of the users involved. However, in practice, users may group into categories, 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 a user study that preceded the development of the semantic music recommendation system. In the third part the use of fuzzy logic to flexible querying is explained. Finally, in the fourth part, the testing of the semantic music recommendation system in the real-world is discussed.

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 provided by users of music information retrieval systems.

Usability is a topic of interest in the musical digital library community. Although the importance of interface and system usability is acknowledged [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 what people 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 behavior 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 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 consisted of 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 these users was asked to annotate music by means of semantic descriptions. The study provided a large database that was then used to build a semantic music recommendation system.

3.1 Global setup

The survey resulted in a dataset with information about personal, demographic and musical background of 774 participants. From this group, a sample of 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 (i.e. excerpts of 30 seconds) thus reflected the musical taste of the targeted population. 79 subjects rated the whole set of 160 musical excerpts. In the annotation experiment, a representative population of users, described music using a set of semantic adjectives. Our model distinguished between affective/emotive, structural and kinaesthetic descriptors. 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 and what was their personal judgment. (See Lesaffre [2] for a detailed description).

3.2 Summary of results

Survey

With 774 participants, a representative sample of the targeted population was reached and a 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.

Multiple relationships between the categorical variables gender, age, musical background, and musical taste were found. For example, it was 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.

Annotation experiment

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 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 [18]. 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 based on the results of the foregoing user study. The system incorporates the annotations, that is, the ratings of semantic descriptors made by the participants in the experiment. In the present study we had to deal with vagueness that arose from the quality ratings which used concepts like ‘rather’, ‘moderate’ and ‘very’. In the case of qualitative adjectives there are semantic ambiguities between levels and there exists no definite threshold for which an emotion becomes too ‘passionate’ or ‘carefree’. Rather we have to differentiate between descriptors which are perfectly acceptable for the user. Obviously, the meaning of vague expressions like ‘rather passionate’ is user dependent. In this context, the multi-valued logic of fuzzy logic was considered as a possible option to account for the vague descriptors. The interest of using fuzzy logic for a user is a better representation of the user’s preferences.

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 of music, 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 the particular search task, after having listened to a recommended piece of music. Secondly, the user is asked to evaluate the general usability of content-based querying and of the distinct descriptor sets.

4.2 Fuzzy logic functions

To deal with the problem of interfacing linguistic categories, such as adjectives, with numerical data and for expressing user's preference in a gradual and qualitative way, ideas of fuzzy logics are used. To the best of our knowledge, fuzzy set methods have

not been applied yet to the representation of annotations provided by real users of music information retrieval systems. In what follows, a description is given of the fuzzy logic functions of the semantic music recommender system. This component was built using Visual Basic .NET for Microsoft Access Databases. 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 four different types of users, based on gender (male, female) and musical expertise (expert, novice). As a consequence, for each adjective, four fuzzy functions were calculated. Each function is characterized by three numbers, namely, the 25th, 50th and 75th percentile values of the cumulative rating value. To obtain that function 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 3 values according to the cumulative percentages of 25%, 50% and 75% were calculated. These 3 values each define a fuzzy function score. Then, the cumulative distribution function is built using 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).

The 3 values v1, v2 and v3 define the following fuzzy function score:

IF $x \leq v1$ THEN score(x) = 0
IF $v1 < x \leq v2$ THEN score(x)=0,5[(x – v1)/(v2 – v1)] (a number between 0 and 0,5)*
IF $v2 < x \leq v3$
THEN score(x)=0,5+0,5[(x – v2)/(v3 –v2)] (a number between 0,5 and 1)*

v1, v2 and v3 are calculated as follows:

IF "rating x" = 0 : x / frequencyRatings(0)
ELSE "rating x"+(x - frequencyRatings (0.. "rating value x" - 1))/frequencyRatings ("rating value x") to which frequencyRatings (0..y) = the number of the rating values $\leq y$

In what follows an example of the calculation of v1, v2 and v3 for the profile 'male expert' and the descriptor 'cheerful' is given. Figure 1 shows a plot of the fuzzy function for this example.

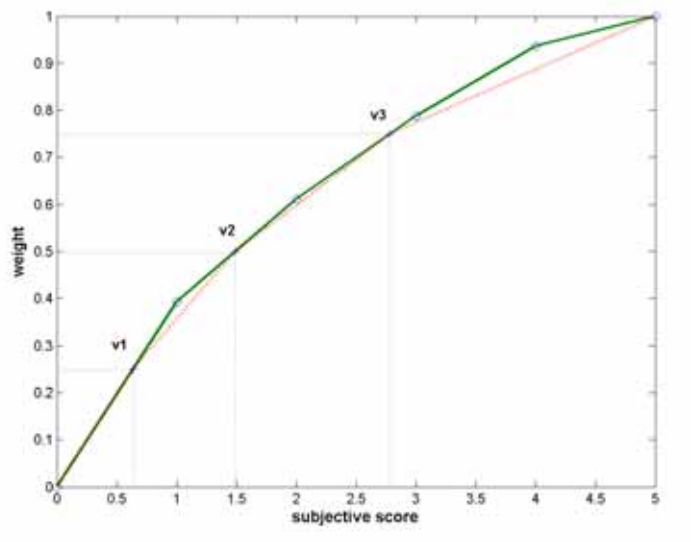

Fig. 1: Cumulative function and fuzzy function

The cumulative distribution function (see Fig. 1 full line) is built on the number of ratings given by male experts for 'cheerful'.

not *[frequencyRatings(0)] = 1252*;
little *[frequencyRatings(1)] = 698*;
moderate *[frequencyRatings(2)] = 564*;
rather *[frequencyRatings(3)] = 474*;
very *[frequencyRatings(4)] = 202*.

The total number of evaluations = 3190. From this discrete set of five known data points (1252, 698, 564, 474 and 202) a new, fuzzy function is built (see Fig. 1 dotted line) on a set of three data points (v1, v2, and v3).

Calculation of recommendation scores per music excerpt, adjective and profile

In order to determine the recommendation scores per music excerpt, 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.

In what follows an example of the calculation of the score for music excerpt one (J.S. Bach, Kommt, ihr Töchter, helft mir klagen, from Matthäus-Passion, BWV 244), the profile 'male expert' and the descriptor 'cheerful' is given.

not *[frequencyRatings(0)] = 12;*
little *[frequencyRatings(1)] = 6;*
moderate *[frequencyRatings(2)] = 0;*
rather *[frequencyRatings(3)] = 2;*
very *[frequencyRatings(4)] = 0.*

The total number of evaluations =20.
Calculation for 1/2(x = 20 / 2 = 10):

“rating value 10” = 0 (< 12), thus :
median = 10 / frequencyRatings (0) = 10 / 12 = 0,83
*score(median) = score(0,83) = 0.5 * (0,83 – v1) / (v2 – v1),*
then v1 = 0,64 < 0,83 < v2 = 1,491...
*=> score(0,83) = 0.5 * (0,83– 0,64) / (1,49– 0,64) = 0,11*

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.

In what follows an example of the calculation of combined scores for music excerpt one, the profile ‘male expert’ and the descriptors ‘cheerful’, ‘sad’ and ‘passionate’ is given. Cheerful = 0,11; sad = 1; passionate = 0,97

Case 1: ‘cheerful’ and ‘sad’ are selected:
*score = square root(0,115 * 1) = 0,34*

Case 2: ‘cheerful’, ‘sad’ and ‘passionate’ are selected:
*score = 3th power root(0,11 * 1 * 0,97) = 0,48*

5. Real-world testing

The semantic music recommendation system was tested in three different real-world environments, addressing four different types of users. The system was first tested by 626 trade fair visitors (i.e. ACCENTA 2005), then by 20 intellectuals (i.e. ALUMNI 2006), and finally by 34 school children and 119 science fair visitors (i.e. Wetenschapsfeest 2006). The tests aimed at evaluating the effectiveness of the fuzzy logic approach to semantic descriptors. In other words, we investigated whether users would agree with the judgments made by participants in our experimental study. We were interested in their assessment of the usability of the system and the descriptor sets.

Quantitative analysis of the users’ satisfaction ratings showed that on the average three quarter of all users were satisfied with the recommendation system (‘well’ to ‘very well’). There are some minor differences between four user groups. However, children reported that 31% of the recommendations did not or little matches their expectations. This can be explained by the fact that the music in the database is too much ‘middle of the road’ for this population. In their drive to hear the music they like, they might not have taken the satisfaction rating task very seriously.

From the query behavior of the four groups of users we learned that their first preference is for affective/emotive descriptors (50% selected by trade fair visitors up to 57% by school children). Second preferred are sonic descriptors (24% selected by intellectuals to 28% by trade fair visitors). Third preferred are movement descriptors (18% selected by school children to 23% by trade fair visitors).

Although several group dependent differences were found, for all groups the most selected emotion descriptor is ‘cheerful’. School children for example tend to search for music that is ‘aggressive’, ‘restrained’ and ‘anxious’ whereas trade fair visitors search for music that is ‘passionate’, ‘tender’ and ‘carefree’. As similar observation was made for sonic descriptors in that there is an overall agreed on interest in ‘clear’ music. Instead, school children search for ‘hard’ and ’rough’ music whereas trade fair visitors search for ‘harmonious’ and ‘soft’ music. For movement descriptors most agreement is on the descriptor ‘dynamic’. School children however like to find ‘quick’ music whereas trade fair visitors prefer music that is ‘flowing.

Over 90% of the trade fair visitors, intellectuals and science fair visitors responded positively to the overall usability of the system, except from the school children (82%). With regard to the usability of the semantic descriptor sets, affect/emotive descriptors are found most useful, followed by movement descriptors and after everything else sonic descriptors.

6 Conclusion

In this paper we described the development and real-world testing of a music information retrieval system that uses fuzzy logic for handling the vagueness of semantic descriptors used for music annotation. Our results show that a fuzzy logic methodology, combined with a user-oriented approach to music information retrieval, may be effective for the development of a content-based 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.

Users who tested the system in a real-world environment confirmed the usability of semantic-based music information retrieval systems. Even if the tests that were carried out with different user groups showed user dependencies, in mainly 75% of the cases users were satisfied with the music that was recommended.

Our study is suggestive of applying fuzzy logic to a predefined semantic descriptor set. It can be assumed that this methodology may provide a stable basis for further development of content-based access to music.

Acknowledgements

This research has been conducted in the framework of the MAMI (Musical Audio Mining) project (2002-2005), and the DEMCO (New Methods for the description of Musical Content) project (2004-2007). The authors wish to thank Prof. Dr. B. De Baets, Prof Dr. H. De Meyer and MA K. Vermeulen for their assistance with the development of the MAMI research tool.

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". PhD. dissertation, Ghent University.
3. Whitman, B., (2005) "Learning the meaning of music", Ph.D. dissertation, Department of Architecture, Massachusetts Institute of Technology.
4. Straccia, U., (2005) *A fuzzy description logic for the semantic web*. In Sanchez, E., ed.: Capturing Intelligence: Fuzzy Logic and the Semantic Web. Elsevier
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), pp 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, pp 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., pp 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).
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,pp 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, pp 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, pp 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, pp 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), pp 39-67.