TOWARDS DESCRIBING PERCEIVED COMPLEXITY OF SONGS: COMPUTATIONAL METHODS AND IMPLEMENTATION

SEBASTIAN STREICH, PERFECTO HERRERA

Universitat Pompeu Fabra, Barcelona, Spain
{sstreich,pherrera}@iua.upf.es

Providing valuable semantic descriptors of multimedia content is a topic of high interest for music content processing. Such descriptors should merge the two predicates of (1) being useful for different operations such as retrieval, visual representation of collections, classification, etc., and (2) being automatically extractable from the source. In this paper the semantic descriptor concept *music complexity* is introduced, and the advantages of their usage for music retrieval and for automated music recommendation are addressed. The authors provide a critical review of existing related proposals and also prospect new methods for automated music complexity estimation.

INTRODUCTION

Semantics (from the Greek *semantikos* = "significant meaning") refers to the meaning of things. When we use a computer to access information, it is usually only a tool for converting the digital data into something our senses can perceive. The computer is blind for the meaning of the information (see also [1]). Nevertheless, we can attach additional metadata to the digital file, which we call *semantic descriptors*. Semantic descriptors, even if they do not convey meaning to computers, they do indeed to humans using computers, and pave the way for computers to behave like if they could understand much more the content they are processing.

Semantic descriptors for multimedia content become more and more important with constantly growing numbers of existing files. The potential of a large collection can only unfold to full extent if content-based queries are possible. To give an example, a customer looking at a large music file collection in an online music store could be interested in finding songs that are played by an orchestra and are "easy" to listen to. Obviously, such a query is not achievable without certain semantic descriptors being associated with the files in the collection. Providing this kind of data in a reliable manner demands considerable effort when done manually. So there is a clear need for ways to automatically compute semantic descriptors from the file itself.

The goal is to provide descriptors that reflect intrinsic characteristics of musical performances but, at the same time, being relevant to listeners in order to allow useful content-based queries. The idea of automatically computing these is not essentially new and research has been carried out on this field for several years since (as e.g. in the MPEG7 context [2]). Yet, it remains intricate to bridge the gap from low-level signal descriptors like the spectral centroid or onset positions, to high-level semantic descriptors like instrumentation, tempo, or key. The former are very closely linked to the consideration of a song as a signal and are obtained relatively easily. The latter are treating the song as music and their direct utilization in queries to databases would be very straightforward provided reliable extractors could be implemented.

In this sense, *musical complexity* appears to be even harder to compute, because it relies – at least partly – on such high-level semantic attributes of music trying to capture characteristics of their temporal evolution and stochastic properties. This is not the only difficulty with music complexity though. The individual estimation of the musical complexity is likely to be highly subjective (see e.g. [3], [4], [5]), because the experiences, the abilities, and the preferences in active music listening can vary significantly from one person to the other.

Nevertheless, we consider music complexity a valuable supplement to existing high-level descriptors of musical content. Temperley in chapter 11.5 of [6] assumes a connection between the complexity of harmonic patterns and individual listener's preferences for pieces of music. In [7] Parry examined the chart performance of rock music titles and associated it with their musical complexity. He found the overall chart performance of the songs to be positively correlated with their melodic and rhythmic complexity. Simonton in [8] reports the results of an extensive study on the relationship between melodic complexity and popularity. For his large sample of 15.618 classical themes he found a clear connection of these two parameters.

Particularly, the subjectivity in sensing complexity can be of advantage when recommending music based on a user profile. In the remainder of this paper we therefore formulate our notion of musical complexity in the given

context, further we will expose ways of utilising the music complexity descriptor concept in music information retrieval scenarios, and we finally will describe possible computation methods for the different facets of music complexity.

1 MUSIC COMPLEXITY DEFINITION

Various definitions of complexity can be found, because different understandings of the term exist in different contexts. For this reason we have to clarify what we are talking about when referring to music complexity in this paper.

1.1 Background

The theory of algorithmic information provides a definition of complexity, measuring the amount of information contained in a sequence of numbers. This measure is known as *Kolmogorov complexity* and has already been used in the digital audio domain (see e.g. [9]). Yet, applied directly to a digitized musical recording it captures rather the compressibility than the complexity a human listener would accredit to it. Apart from that, Standish in [10] addresses the flaw of Kolmogorov complexity, that a random sequence of numbers will always yield a maximum complexity although it does not contain any meaningful information. He suggests the use of *equivalence classes* to overcome this. An equivalence class for him is the set of all mutations of a sequence that are equivalent in a given context. So for example random sequences could hardly be distinguished by a human observer and would therefore form a large equivalence class. On the other hand, for a written text only very few mutations exist, that would be judged as equivalent. If the equivalence class is considered in the complexity computation, then the result captures the context dependency and hence is more meaningful.

More specifically related to music, Eerola and North point out in [11], that the traditional information theorist view of complexity does "not address the role of the listener's perceptual system in organising the structural characteristics of music". Therefore they propose an *expectancy-based model* (EBM) to estimate complexity. Their model for melodic complexity is based on tonal, intervallic, and rhythmic features derived from a symbolic representation of the music. Comparing the ability of this model to predict listeners' complexity judgements with an information-theoretic and a transition probability model, they found it to be the most accurate one.

Nevertheless, Pressing in [12] convincingly uses what he calls *information-based complexity* to calculate an estimate of the difficulty musicians would have in producing certain rhythmical patterns. He achieves this by simply applying a processing cost function to the symbolic level (i.e. high-level) attribute syncopation on quarter-note and eight-note level. Pressing also mentions two other slants of complexity in his publication, which he names *hierarchical complexity* and *dynamic complexity*. Referring to music the former would be focussing on the structure of a song, and the latter on the time behaviour and change in a musical performance.

Shmulevich and Povel in [13] propose a measure (refered to as *PS-measure*) for rhythmic complexity. It is also based on the amount of information coded in the rhythmic patterns, but at the same time it takes into account perceptual issues that have been reported by Povel and Essens in [14]. When it is applied to rhythm patterns in symbolic form (i.e. quantized to a grid), the *PS-measure* outperforms the *T-measure* [15] and the *LZ-measure* [16] in predicting human judgements of rhythmic complexity. This is of little surprise as the two latter measures are neglecting perceptual information.

On the other hand, in [17] Scheirer directly utilizes the statistical properties of five psychoacoustic (low-level) features of short musical excerpts to model perceived complexity. These features are the *coherence of spectral assignment to auditory streams, the variance of number of auditory streams, the loudness of the loudest moment, the most-likely tempo,* and *the variance of time between beats* (see [18] Chapters 4-6 for details). He reports that, by using linear regression techniques on these, they are strongly significant in predicting the mean complexity ratings of a group of 30 human listeners.

1.2 Towards computing facets of music complexity

We are looking for a descriptor that gives us a complexity estimate for entire songs. Our complexity measure should reveal the effort the listener has to put into analyzing the music in order to capture what is going on.

In [19] Finnäs states that "unusual harmonies and timbres, irregular tempi and rhythms, unexpected tone sequences and variations in volume" raise the level of perceived complexity. This statement is neither exhaustive nor precise. But combined with the quintessence of the preceding sections we can still formulate the following assumptions:

1. Musical complexity possesses many different aspects.
2. These aspects can be independent from each other.
3. These aspects can be linked, as well, to high-level as to low-level features of the song.
4. The richness of mutations is linked to musical complexity.
5. The rate at which events have to be processed is linked to musical complexity.
6. Expectation and surprise play also a role in complexity perception of music.

Now these assumptions already give us some directions for the design of our complexity descriptor. Assumptions 1 and 2 favour a multidimensional layout. We propose a set of six dimensions of musical complexity to be treated separately. These are *melody, harmony, rhythm, timbre, structure*, and *acoustic properties*. The latter is meant to incorporate aspects of the spatial and dynamical comprehensiveness of the song, which are, strictly speaking, not so much an attribute of the music as of the recording. We will address each individual dimension later in this paper.
Assumption 3 reveals the idea of complexity as a meta-descriptor. It will not be computed directly from the source signal, but rather from high-level and low-level features that have to be extracted first.
Assumptions 4 and 5 reflect statistical and temporal aspects we have to account for. We will consider that the listener is processing a stream of events during his/her listening to music.
Finally, assumption 6 causes the most difficulties. We already stated that the temporal evolution of extracted features will be addressed. But modelling arising expectations during the process of listening to music is not a trivial task. For the moment, we leave this as it is and get back to it later when discussing actual methods of computation.

2 APPLICATION SCENARIOS

After the technical basis has been established we can now focus in some more detail on possible applications for our descriptor. We are only interested here in the interaction with music databases and will not discuss other possible fields of appliacation. When employing a music database, three major tasks can be identified:

1. The retrieval of songs that match the user's desires.
2. The generation of a program (playlist).
3. The visualization of the content in order to allow the user navigating through it.

We will address each one of these tasks in the following three paragraphs.

2.1 Song retrieval

There are different possibilities of querying a song database. The most obvious one is the direct specification of parameters by the user. Since the complexity descriptors consist of only one value per song for each dimension, they can be used very easily in queries. The user can specify only those dimensions he is interested in and narrow down the set of results. This way it is very straightforward to find music that, for example, doesn't change much in loudness level over time, or which has a rhythmic complexity that is interesting enough, but not too difficult to dance to.
A second way of querying is the so called *query-by-example* approach. The user presents one or several songs to the database and wants to find similar ones. It is straightforward in this case to compute the complexity descriptors for the provided example and use them for the actual query. The weighting and/or the tolerance of the different dimensions could be specified by the user directly, taken from a precomputed user profile, or extracted from the example in case it consists of more than one song. A user profile would be established by analysing the user's listening habits (i.e. songs he/she has in his/her collection; songs he/she listens to very frequently, etc.).
Probably the most exciting way of querying is a query without any specification apart from the limiting factor that the user should like the retrieved song or at least find it interesting. This is usually referred to as *music recommendation*. Why could descriptors of musical complexity be useful for this?
As pointed out in the introduction, there is good reason to believe that the level of perceived complexity of a piece of music can be associated with the preference for it. We have already cited several studies supporting this assumption for communities of people (i.e. in a more macrosociological sense) [7] [8]. In the application we are interested in here the circumstances are slightly different, since there is a single person interacting with the database. Nevertheless, the complexity descriptor can be useful in this context too, as we will show in the following.

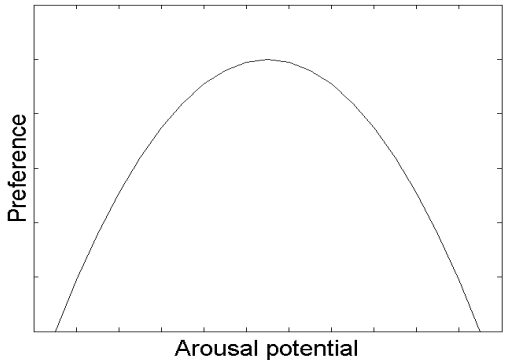

Figure 1: Relationship between preference for music and its arousal potential (after Berlyne).

Back in the 1970s Daniel Berlyne (as cited by [4]) established a theory stating that an individual's preference for a certain piece of music is related to the amount of activity it produces in the listener's brain, to which he refers as the *arousal potential*. According to this theory there is an optimal arousal potential that causes the maximum liking, while a too low as well as a too high arousal potential result in a decrease of liking (see Fig. 1). Berlyne identifies three different categories of variables affecting arousal (see [20] for details). As the most significant he regards the *collative variables*, containing among others *complexity* and *novelty/familiarity* of the stimulus. Since we are

modeling exactly these aspects of music with our descriptor, it is supposed to be very well suited for reflecting the potential liking of a certain piece of music.
Hargreaves and North point to two potential problems with respect to Berlyne's theory: the influence of the listener's mood and intention when selecting music, and the dependence on the appropriateness of the music for the listening situation [4]. As they report, both show a measurable effect on a subject's musical preference in a certain context. However, we believe these effects can, at least to some extend, be absorbed by an automated music recommendation system. For example different user profiles could be established for one person, depending on mood or listening situation.

2.2 Playlist generation

A playlist is a list of titles to be played like a musical program. A user interacting with a database might ask for the automated generation of such a list. As Pachet, Roy and Cazaly point out in [21] the creation of such a list has to be taken serious, since "The craft of music programming is precisely to build coherent sequences, rather than just select individual titles."
A first step towards coherence is to set certain criteria the songs have to fulfill in order to be grouped into one playlist. For example for doing housework the restrictions could be *rather fast tempo* and *intermediate complexity* in each dimension. Also transitional properties could make sense, as for example increasing rhythmic complexity while melodic and harmonic complexity stay fixed.
Pachet, Roy and Cazaly go further and look at an even more advanced way of playlist generation capturing the two contradictory aspects of *repetition* and *surprise*. Listeners have a desire for both, as they state, since constant repetition of already known songs will cause boredom, but permanent surprise by unknown songs will probably cause stress.
In their experiments Pachet, Roy and Cazaly used a hand edited database containing, among others, attributes like *type of melody* or *music setup*. We can see a correspondence here to our melodic and timbral complexity, that encourages the utilization of our complexity descriptors for playlist generation.

2.3 Collection visualization

A user might want to navigate through a digital music collection by other means than artist and title. To allow for this, a suggestive graphical visualization of relevant musical features has to be provided. One example for such visualization is the *Islands of Music* application, developed by Pampalk [22]. This application uses the metaphor of islands and sea to display similarities of songs in a collection. The application uses features that are motivated from psychoacoustic insights, and processes them through a self-organizing map (SOM). In order to compute similarity between songs the sequence of feature values extracted from each song has to be shrinked to one number. Pampalk does this by taking the median. He reports satisfying results, but at the same time states that the median is not a good representation for songs with changing properties (e.g. bimodal feature distribution).
Our complexity descriptor is designed to consist of only one value for each dimension that captures the properties of the whole song. Hence, the problems of reducing a time sequence to one single value that is still representative for the whole sequence doesn't arise. Furthermore, each single dimension reflects specific characteristics of the music that are potentially of direct relevance for the listener. The descriptor is therefore very well suited to facilitate the visualization of musical properties the user might want to explore.

3 METHODS OF COMPUTATION

The whole discussion about our descriptor concept remains simply academic as long as there are no algorithms available that can actually perform the extraction in a reliable manner. In this section we therefore focus on the different dimensions of complexity we defined in 1.2 and report the state-of-the-art of their computability. It should be pointed out, that we content ourselves with the extractors working on music of the western cultures and traditions.

3.1 Melody

Back in 1990 Eugene Narmour proposed a model for melodic complexity. This *Expectation-Realization model* as he calls it is extensively described in [23]. The model uses a set of different interval patterns, raising certain expectations on the listener's side. Frequent realization of these expectations reveals a low level of complexity; frequent disappointment reveals a high level of complexity. The model has been used widely and successfully to estimate melodic complexity in different experiments [7],[24]. It has been extended to capture also aspects of rhythm and tonality [11], since a melody can't be isolated from these parameters.
Lately, experiments were conducted that showed, how the accuracy of the model can be further improved by taking a larger melodic context into account [25]. Since the focus of the original model is limited to two notes at a time only, it neglects the impact of the longer-term melodic evolution (e.g. repetition of motives) on the listeners' predictions of continuation.
It must be stated, that all these models work with a symbolic description of the melody as an input. Usually, digital music files won't have this symbolic description attached to themselves. The key problem thus remains in the automatic extraction of the melody from the audio stream. Many approaches to this kind of automated

transcription have been made (see e.g. [26]), but to date did not lead to a universal and reliable solution.

3.2 Harmony

Although the theory of harmony in music has a long tradition, the authors didn't find one dominant model addressing its complexity. Research has been done on the expectations evoked in listeners by harmonic progressions especially on the field of classical music [27]. It turned out, that listeners usually predict a chord that results from a transition considered as common in the given musical context. Yet, to our knowledge no tests have been carried out that correlated the perceived harmonic complexity with the fulfilment or disappointment of these expectations.

Temperley supposes that his preference rule system [6] could reveal an estimate for the interestingness of music (p. 307). The mapping of achieved scores would go from *incomprehensible* (breaking all rules) over *tense* to *calm,* and finally to *boring* (all rules obeyed). Although he doesn't use the term complexity, this basically reflects what we are looking for. He names four different aspects of this harmonic complexity:

1. The rate at which harmonies change.
2. The amount of harmonic changes on weak beats.
3. The amount of dissonant notes.
4. The distance of consecutive harmonies in a music theoretical sense.

A different approach could be based on the application of rewriting rules as proposed by Pachet in [28]. He addresses the effect of harmonic surprise in Jazz music. By learning chord progressions and applying rewriting rules he tries to model the predictability of a certain chord sequence. A high predictability would then yield a low complexity rating and vice versa.

Both ideas have two drawbacks. As for the melodic complexity models already we here again need a transcription of the chords first before we can start to analyse the complexity. And in further accordance we can find many approaches (see e.g. [29], [30]), but so far no satisfying solution. The second drawback lies in the fact, that the harmonic rules to be used in the model might not be truly universal. Even when we restrict ourselves to western tonal music, it could be difficult to formulate rules fitting all different styles and genres.

On the other hand it can certainly be doubted, that the listener performs an exact harmonic analysis while enjoying music. In this respect, also the perceived harmonic complexity should not need to rely on an exact transcription. The authors therefore want to explore more immediate ways to harmonic complexity in future research. A possibility would be the application of the pitch class profile [31], that is strongly related with the harmonic content of the music [32]. It could be mapped into the spiral array proposed by Chew that defines a three-dimensional space of harmonic instances [33]. The spiral array has the property that the spatial proximity reflects also musical proximity of harmonies to some extend. As the song evolves, the path through this space could be recorded and then analysed. Frequent changes and long distances would both increase the assigned level of complexity.

3.3 Rhythm

In 1.1 we already referred to a publication by Shmulevich and Povel [13] introducing the *PS-measure* for rhythmic complexity. They state, with reference to [14], that a listener tries to establish an internal clock, when hearing rhythmic music. According to this clock the listener then segments the rhythm pattern and tries to code the segments. The PS-Measure utilizes the induction strength of the clock, and the coding efficiency of the rhythm.

Once more, the input data in their experiment was derived from a symbolic representation of the music. Nevertheless, in this case extractors exist, that can compute onsets and accents from the audio stream in an adequate manner (see e.g. [34], [35]). In human performances of music the timing is likely to vary significantly more than in a computer edited symbolic representation. Since we don't want this to affect our complexity measure, it is necessary to apply a reasonable quantization.

The authors implemented another algorithm that is related with rhythmic complexity, the detrended fluctuation analysis (DFA) of intensity. Originating from time series analysis in the medical domain, it was proposed in [36] by Jennings et al. as a feature for genre classification. They state that the strong periodic trends in dance music (as Techno or Brazilian Forró) make it easily distinguishable from high art music by using this feature. "Jazz, Rock and Roll, and Brazilian popular music may occupy an intermediary position between high art music and dance music: complex enough to listen to, but periodic enough to dance to," they speculate. Hence, we can think of this feature as a rating of "danceability". A first informal test we conducted revealed that even a small group of people disagrees about the "danceability" of a song in the majority of cases, giving even oppositional judgments. This could be due to confusion about what was actually asked for, and also to the personal liking and disliking of certain types of music. Further experiments have to be carried out to evaluate the suitability of this feature.

3.4 Timbre

There is no clear and precise definition of timbre that could be regarded as a common agreement on the music analysis field. By the American Standards Association ([37] p. 45) the following statement was released: "[Timbre is] that attribute of auditory sensation in terms

of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar."

For our purpose we think of timbre as the entity that is the most tightly knotted with sound production (i.e. the source of the sound and the way this source is exited). We then can derive several specifications of the general attributes 4 and 5 of complexity we itemised in section 1.2. This gives us features like the *number of instruments playing*, the *rate at which the leading instrument changes*, or *the amount of modulation of the sound sources*.

As reported in [38], source separation and instrument recognition systems for arbitrary polyphonic music signals are not yet available. Nevertheless, we are planning to conduct experiments with machine-learning techniques for example in the way Aucouturier and Sandler used HMMs for music segmentation [39]. Our application has the advantag, that we need only a rough estimate for the number of instruments, but no exact classification of each one of them.

3.5 Structure

Musical structure forms one of the highest levels of abstraction in content analysis. It is unique compared to the other dimensions in the sense that all of them are potentially relevant for its computation.

We want to refer to structure on a rather macroscopic level (i.e. in terms of intro, verse, and chorus rather than motive or theme). Along with our remarks in section 1.2 we can identify attributes of structural complexity such as the *number of distinguishable parts*, or the *level of periodicity of their appearance*. It would also be desirable to *estimate the dissimilarity of consecutive parts*. Very contrasting parts following each other would be very surprising and thus probably enhance the perceived complexity.

Once more, we have to face the fact, that before we can perform any structural complexity processing, first the structure itself has to be extracted. Various approaches to this problem have been taken and are still explored (see e.g. [40], [41]). The general purpose solution has yet to be found.

3.6 Acoustics

As we mentioned in 1.2 acoustic complexity is not completely intrinsic to the music, but rather to the recording (or the performance). We want to distinguish between two aspects of acoustical effects here, forming subdimensions of our descriptor: dynamics and space.

Dynamic complexity could be referred to in terms of abruptness and frequency of changes in dynamic level. There are different options for defining the time scope. By keeping the frame size small one would find the distinction between dynamically compressed and uncompressed material. With longer windows one could detect fades and dynamic changes between larger segments. The regularity of dynamic changes has to be observed as well, since an uncompressed drumloop will have many abrupt changes in short-term dynamic level, but because of its periodicity these will not be found very complex by a listener.

Calculating a very accurate estimate of the perceived loudness of the complex sounds that form a musical performance is a very complicated task. Several aspects of psychoacoustics have to be considered [42]. Finally, since there are also subjective components, and the final playback level can not be known, the loudness can only be approximated.

In [43] Vickers proposes a simplified algorithm to calculate long-term loudness and dynamic spread of whole audio files. He proposes the mean absolute deviation of the per-frame loudness as a definition for dynamic spread. Scheirer in contrary defines the dynamic range as "the greatest of the local differences in total loudness within short windows throughout the signal" ([18] p. 166). He uses windows of only 200ms to compute this feature.

To compute spatial complexity we consider only stereo recordings and no advanced multi-channel formats in this paper. So far, they form by far the majority of items in digital music file databases. A straightforward example for spatial complexity thus could be the disparity of the stereo channels. A quasi mono situation with similar channels would reveal less complexity than a recording that has only little correlation between the two channels. But also more advanced aspects could be considered, such as the movement of the acoustical center of effect within the stereo image. Yet, this is not trivial from a computational point of view.

There is also another aspect of spatial complexity which originates from either natural or artificial sound effects. Namely these are all types of delay and reverberation. Filter, flanger or chorus effects we would rather group under timbral complexity (3.4).

The measurement of reverberation has a solid tradition in room acoustics, where several different measures exist. Griesinger gives an overview over several measures of spaciousness in [44]. Usually, these measures take the room impulse response as their input and are thus not suited for a continuous signal. An exception is the InterAural Difference (IAD) introduced by Griesinger, which, as he states, can also be found as a continuous function of music signals.

$$IAD = 10 \cdot \log_{10}\left(\frac{eq(L(t)-R(t))^2}{L(t)^2 + R(t)^2}\right) \quad (1)$$

where the equalization *eq* consists of a low frequency enhancement of 6dB per octave below 300Hz.

At the time of writing of this paper our experiments regarding acoustic complexity are still ongoing. First results seem to indicate, that for extreme cases the

features coincide with human perception, whereas a sensible continuous resolution, also for intermediate values has not been achieved yet.

4 STRATEGIES FOR EVALUATION

In this section we want to give a very brief overview of possible ways to evaluate the complexity estimation algorithms.

A straightforward approach for the evaluation of content analysis algorithms is the correlation of computed results with manually edited ones. This is the way for example the melodic complexity models mentioned in section 3.1 have been evaluated. It must be stated though, that these tests were performed on isolated melodies and not on real recordings of songs. For the "danceability" judgements mentioned in section 3.3 we saw already, that subjects' ratings are not necessarily consistent. It can not be counted out, that at least some of the complexity dimensions are perceived in an unconscious manner by some listeners.

Alternatively, the subjects could be asked to rank a given set of items according to one complexity dimension. The results could be clustered and matched against the automatically extracted values. Yet, this task could be even more difficult for untrained listeners, especially when the items are very distinct in their genre, instrumentation, etc.

A third way of evaluation could comprise the presentation of a list of items to the subjects and the task to identify the underlying concept of arrangement. The ordering of the list would be done according to the output of the extraction algorithm under test. This involves much more effort in the interpretation phase; the subjects would make statements in verbal form instead of providing simple numbers. The advantage is that untrained listeners might be less confused with this kind of task.

5 CONCLUSIONS

We have presented a musical content descriptor concept to capture aspects of music complexity as they are perceived by listeners. Looking at the experimental results reported by others and cited in this paper, the application of a music complexity descriptor in the field of musical content retrieval and interaction seems very promising. As we have shown, providing a content description in terms of complexity could serve to facilitate and enhance the interaction with digital music databases.

Regarding the computability of the descriptor, we have pointed to several algorithms and approaches that could be suitable for our needs. Further investigation and experiments are planned by the authors to fathom this. For certain dimensions of complexity, like Melody and Harmony, current extraction algorithms seem still very far from our demands. But, as Scheirer points out, the normal human listener does not perceive music in the way a transcription system does. "When human listeners are confronted with musical sounds, they rapidly and automatically orient themselves in the music. Even musically untrained listeners have an exceptional ability to make rapid judgments about music from very short examples, such as determining the music's style, performer, beat, complexity, and emotional impact."([18] Abstract).

In other words, we can perceive one melody or chord sequence as more complex than another one without being able to write down the musical score. Therefore, other ways of complexity estimation could be thought of, that don't rely on a symbolic representation of the music. We want to address this as well in further studies.

6 ACKNOWLEDGEMENTS

This research has been partially funded by the EU-FP6-IST-507142 project SIMAC (Semantic Interaction with Music Audio Contents).

More information will be found at the project website http://www.semanticaudio.org.

REFERENCES

[1] T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web," *Scientific American* vol. 284, no. 5, pp. 34—43 (2001).

[2] International Standards Organization, *MPEG-7: Context and Objectives*, JTC1/SC29/WG11 N2460 (1998).

[3] L. Steck, M. Machotka, "Preference for musical complexity: Effects of context," *Journal of Experimental Psycholog: Human Perception and Performance* vol. 104, no. 2, pp. 170—174 (1975).

[4] A. C. North, D. J. Hargreaves, "Experimental aesthetics and everyday music listening," *The social psychology of music* pp. 84—103, Oxford University Press (1997).

[5] N. Birbaumer et al., "Perception of music and dimensional complexity of brain activity," *International Journal of Bifurcation and Chaos* vol. 6, no. 2, pp. 267—278 (1996).

[6] D. Temperley, *The cognition of basic musical structures*, The MIT Press (2001).

[7] R. M. Parry, "Musical complexity and top 40 chart performance," unpublished, (2002).

[8] D. K. Simonton, "Drawing inferences from

symphonic programs: Musical attributes versus listener attributions," *Music Perception* no. 12, pp. 307—322 (1995).

[9] E. D. Scheirer, "Structured Audio, Kolmogorov Complexity, and Generalized Audio Coding," *IEEE Trans. on Speech and Audio Processing* vol. 9, no. 8, pp. 914—931 (2001).

[10] R. Standish, "On complexity and emergence," *Complexity International* no. 9 (2001).

[11] T. Eerola, A. C. North, "Expectancy-Based Model of Melodic Complexity," *Proc. of the Sixth International Conference on Music Perception and Cognition,* CD-ROM (2000).

[12] J. Pressing, "Cognitive complexity and the structure of musical patterns," *Proc. of the 4th Conference of the Australasian Cognitive Science Society* (1999).

[13] I. Shmulevich, D. J. Povel, "Measures of temporal pattern complexity," *Journal of New Music Research* vol. 29, no. 1 (2000).

[14] D. J. Povel, P. J. Essens, "Perception of temporal patterns, " *Music Perception* no. 2, pp. 411—441 (1985).

[15] A. S. Tanguiane, *Artificial Perception and Music Recognition,* Springer Verlag (1993).

[16] A. Lempel, L. Ziv, "On the complexity of finite sequences," *IEEE Trans. on Information Theory* vol. 22, no. 1, pp. 75—81 (1976).

[17] E. D. Scheirer, R. B. Watson, B. L. Vercoe, "On the perceived complexity of short musical segments," *Proc. of International Conference on Music Perception and Cognition*, CD-ROM (2000).

[18] E. D. Scheirer, *Music-Listening Systems*, PhD Thesis, MIT Media Lab (2000).

[19] L. Finnäs, "How can musical preference be modified? A research review," *Bulletin of the Council for Research in Music Education* no. 102, pp. 1—58 (1989).

[20] D. E. Berlyne, *Aesthetics and psychobiology,* Appleton-Century-Crofts (1971).

[21] F. Pachet, P. Roy, D. Cazaly, "A combinatorial approach to content-based music selection," *IEEE Multimedia* vol. 7, no. 1, pp. 44—51 (2000).

[22] E. Pampalk, "Islands of Music: Analysis, Organization, and Visualization of Music Archives," Master's thesis, Vienna University Technology(2001). http://www.oefai.at/elias/music/thesis.html

[23] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures,* The University of Chicago Press (1990).

[24] E. G. Schellenberg, "Simplifying the expectation-realization model of melodic expectancy," *Music Perception* no. 14, pp. 295—318 (1997).

[25] T. Eerola, P. Toiviainen, C. L. Krumhansl, "Real-time predictions of melodie: continuous predictability judgements and dynamic models," *Proc. of the 7th International Conference on Music Perception and Cognition* (2002).

[26] E. Gómez, A. Klapuri, B. Meudic, "Melody Description and Extraction in the Context of Music Content Processing," *Journal of New Music Research* vol. 32, no. 1 (2003).

[27] M. Schmuckler, "Expectation in music: Investigation of melodic and harmonic processes," *Music Perception* no. 7, pp. 109—150 (1989).

[28] F. Pachet, "Surprising Harmonies," *International Journal on Computing Anticipatory Systems* no. 4 (1999).

[29] A. Klapuri, *Signal Processing Methods for the AutomaticTranscription of Music*, PhD Thesis, Tampere University of Technology (2003).

[30] J. P. Bello, *Towards the Automated Analysis of simple polyphonic music: A knowledge-based approach,* PhD Thesis, Queen Mary, University of London (2003).

[31] T. Fujishima, "Realtime chord recognition of musical sound: A system using common lisp music," *Proc. of International Computer Music Conference*, pp. 464—467 (1999).

[32] A. Sheh, D. Ellis, "Chord Segmentation and Recognition using EM-Trained Hidden Markov Models," *Proc. of International Symposium on Music Information Retrieval* (2003).

[33] E. Chew, *Towards a Mathematical Model of Tonality,* PhD Thesis, MIT (2000).

[34] C. Duxbury et al.,"Complex domain onset detection for musical signals, " *Proc. of 6th International Conference on Digital Audio Effects* (2003).

[35] A. Klapuri, "Musical meter estimation and music transcription," *presented at Cambridge Music Colloqium* (2003).

[36] H. D. Jennings, "Variance fluctuations in nonstationary time series: a comparative study of music genres," *Condensed Matter* (2003). http://xxx.lanl.gov/abs/cond-mat/0312380

[37] American Standards Association, *American Standard Acoustical Terminology*, Definition 12.9, p. 45 (1960).

[38] P. Herrera, G. Peeters, S. Dubnov, "Automatic classification of musical instrument sounds," *Journal of New Music Research* vol. 32, no. 1 (2003).

[39] J. J. Aucouturier, M. Sandler, "Segmentation of musical signals using hidden markov models," *Proc. of AES 110th Convention* (2001).

[40] W. Chai, B. Vercoe, "Structural analysis of musical signals for indexing and thumbnailing," *Proc. of ACM/IEEE Joint Conference on Digital Libraries* (2003).

[41] D. van Steelant et al., "Discovering Structure and Repetition in Music Audio," *Proc. of Eurofuse Workshop* (2002).

[42] E. Zwicker, H. Fastl, *Psychoacoustics*, Springer Verlag (1999).

[43] E. Vickers, "Automatic long-term loudness and dynamics matching," *Proc. of AES 111th Convention* (2001).

[44] D. Griesinger, "Objective measures of spaciousness and envelopment," *Proc. of AES 16th International Conference on Spatial Sound Reproduction* (1999).