

A Generate and Sense Approach to Automated Music Composition

Sunjung Kim

Multimedia Concepts and Applications
Faculty for Applied Informatics
Augsburg University, Eichleitnerstr. 30
D-86135 Augsburg, Germany
skim@informatik.uni-augsburg.de

Elisabeth André

Multimedia Concepts and Applications
Faculty for Applied Informatics
Augsburg University, Eichleitnerstr. 30
D-86135 Augsburg, Germany
andre@informatik.uni-augsburg.de

ABSTRACT

Nobody would deny that music may evoke deep and profound emotions. In this paper, we present a perceptual music composition system that aims at the controlled manipulation of a user's emotional state. In contrast to traditional composing techniques, the single components of a composition, such as melody, harmony, rhythm and instrumentation, are selected and combined in a user-specific manner without requiring the user to continuously provide comments on the music employing input devices, such as keyboard or mouse.

Categories and Subject Descriptors

H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing

General Terms

Human Factors

Keywords

Affective User Interfaces, Automated Music Composition

1. INTRODUCTION

It is commonly agreed upon that music may have a strong impact on people's emotions. Think of the anger you experience when being exposed to obtrusive music or your joy when attending an excellent music performance. To exploit the enormous potential of auditory sensations on human perception and behavior, a systematic treatment of people's emotional response to music compositions is of high relevance. In our work, we examine in how far music that elicits certain emotions can be generated automatically.

Previous research has shown that physiological signals may be good indicators for the affective impact of music [5]. The system described here makes use of physiological sensors as well, but distinguishes from earlier work by the fact that music is not selected from a data base, but generated automatically.

Copyright is held by the author/owner(s). This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in:
IUI'04, Jan. 13–16, 2004, Madeira, Funchal, Portugal.
ACM 1-58113-815-6/04/0001.

To enable a more systematic investigation of musical variables, we handle rhythm and melody independently of each other. In addition, we consider a great number of short rhythms as opposed to a few complex music pieces (e.g. ten in the case of [5]).

To implement a music player that adjusts itself to the user's affective state, the following prerequisites must be fulfilled. First of all, we need a method for measuring the emotional impact of music. In this paper, we describe an empirical study to find correlations between a user's self-reported impression and his or her physiological response. These correlations will then serve as a basis for such a measurement. Secondly, we need a collection of music pieces that can be employed to influence the user's affective state in a certain direction. Here, we present a generate-and-sense approach to compose such music automatically. Finally, we need a component that continuously monitors the user's affective state and decides which music to present to him or her. To draw inferences on the user's emotional state, we employ a rule-based system that is based on empirical data and combines the observations gained from various sensors.

2. EXPERIMENTAL SETTING

In our work, we characterize emotions in terms of the two continuous dimensions: arousal and valence [3]. For instance, disquieting music could be described by negative valence and high arousal while relaxing music is rather characterized by positive valence and low arousal.

To derive typical physiological patterns for certain emotional states, we conducted 10 experimental sessions of 1-2 hours duration with subjects recruited from Augsburg University. In the sessions, the subjects had to evaluate 1422 automatically generated rhythms according to pairs of opposite attributes in the emotion space. We decided to start with "disquieting" versus "relaxing" and "pleasant" versus "unpleasant" since these attributes were rather easy to distinguish for the user. The subjects had to indicate whether an adjective or its counterpart was satisfied. In case none of the adjectives applied, the music should be evaluated as neutral. In each session, the subjects had to concentrate just on one attribute. While the subjects listened to the rhythms and inputted their evaluation, four types of physiological signals were taken: Electrocardiogram (ECG), Electromyogram (EMG), Galvanic Skin Response (GSR) and Respiration (RESP) using the Procomp+ sensor equipment.

Following Schandry [6], 17 features were extracted from the ECG, 2 features from the EMG, 10 features from the GSR and 4 features from the RESP signal. The subjects had to listen to a rhythm for at least 20 seconds before they were allowed to evaluate it. This time period corresponds to the duration determined by Vossel and Zimmer [7] in which the skin conduction values may develop their full reaction. After the user has evaluated the music, the tonic measures of the signal values are observed without any music stimuli for a period of 10 seconds. Then, a new generated music sample is played for at least 20 seconds. The recorded data are then used to identify characteristic physiological patterns with a strong correlation to user impressions.

3. INFERRING USER IMPRESSIONS

After an initial training phase, the user is no longer forced to explicitly input his evaluation. Instead we try to infer his or her emotional response from the recorded physiological data. To combine different physiological data, we make use of a Fuzzy Rule Base System (FRBS). The physiological features of each signal are the input variables for the FRBS. These values are first “fuzzified”, i.e. assigned to one of the sets „very low“, „low“, „medium“, „high“ and „very high“. By applying so-called fuzzy rules, we determine information on the expected user impression from the single values. The rules for the FRBS result from the empirically determined correlations between user judgements and physiological values. An example of such a rule is:

If HPAmplitude(GSR) = VeryLow & NumPeaks(GSR) = VeryLow
THEN Relaxing = High

After a “Defuzzification Process”, the user reactions that are represented by fuzzy sets are transformed into numeric values which reflect the intensity of these reactions

4. MUSIC GENERATION WITH GENETIC ALGORITHMS

Composing music is a complicated and highly creative task. There have been a number of attempts to compose music automatically based on techniques, such as context-free grammars, finite state automata or constraints (see [4] for an overview). In our case, we don’t start from a declarative representation of musical knowledge. Rather, our objective is to explore how music emerges and evolves from (active or passive) interaction with the human user. For this kind of problem, genetic algorithms have been proven useful. The basic idea of genetic algorithms is to start with an initial population of solution candidates and to produce increasingly better solutions following evolutionary principles. A genetic algorithm consists of the following components:

1. a representation of the solution candidates called chromosomes
2. mutation and crossing operators to produce new individuals
3. a fitness function that assesses solution candidates
4. a selection method that ensures that fitter solutions get a better chance for reproduction and survival

Genetic Algorithms are applied iteratively on populations of candidate problem solutions. The basic steps are:

1. Randomly generate an initial population of solution candidates
2. Evaluate all chromosomes using the fitness function
3. Select parent solutions according to their fitness and apply mutation and crossing operators to produce new chromosomes
4. Determine which chromosomes should substitute old members of the population using the fitness functions
5. Go to step 3 until a stopping criterion is reached.

As a first step, we concentrate on the automated generation of rhythms. In particular, we try to determine an appropriate combination of percussion instruments (i.e., we combine 4 instruments out of a set of 47) and beat patterns. In our case, each population consists of 20 individuals that correspond to a rhythm to be played by four percussion instruments. Rhythms are represented by four 16-bit strings (one for each of the four selected instruments). A beat event is represented by 1 while 0 refers to a rest event. To create new rhythms, we implemented a number of mutation and crossing operators. For example, we make use of a One Point Crossover Operator that randomly chooses a position out of 16 bits of two rhythms and swaps the components to right of these bit positions to create new rhythms.

We implemented two methods for assessing the fitness of rhythms. The first method relies on explicit user judgments and is used for new users to train the system. After the system gets to know the user, the fitness is computed on the basis of his or her physiological feedback. For example, if our goal is to employ music for relaxation and the FRBS predicts a relaxing effect on the basis of the determined physiological data, the chromosome is assigned a high fitness value.

5. CORRELATING SUBJECTIVE IMPRESSIONS WITH OBJECTIVE MEASUREMENTS

After a first statistical evaluation of the experiment, the GSR-signal (which measures skin secretion) was identified as a useful indicator for the attributes “disquieting” and “relaxing” (see Table 1). In particular, a very low HPAmplitude indicates a relaxing effect while a higher HPAmplitude may be regarded as a sign that the music disquiets the user. Our results are consistent with earlier studies which revealed that arousing music is usually accompanied by a fast increase of GSR (for a review of such studies, we refer to Bartlett’s work [1]).

To discriminate between positive and negative emotional reactions to music, EMG measurements have been proven promising. Earlier studies by Bradley and Lang [2] revealed that facial corrugator EMG activity (eyebrow contraction) were significantly higher for unpleasant sounds as compared to pleasant sounds. Our own experiments with EMG measurements at the subjects’ shoulder led to similar results (see Table 2). Higher activity of this muscle is linked to unpleasant rhythms while lower activity is linked to pleasant rhythms.

Since we are interested in a controlled manipulation of the user’s emotional state, we also investigated how the user’s physiological reactions changed over time in dependency of the presented rhythms. Figure 1 shows how the amplitude of the GSR increases

Table 1. GSR Table with Emotions Disquieting (D) vs. Relaxing (R) and Neutral (N)
HPAmp = maximal amplitude within the time window corresponding to a stimulus

GSR-Signal		High Peak Amplitude			
Statistical Analysis		Pattern of HPAmp	Mean (HPAmp)	Two Group t-Test	
				Groups	correlated significantly?/result
Emotion	Disquieting	High/Very High	3.0480131	Group 1: D, Group 2: N and R	Yes/t(914)=-25.399; p<0.001
	Relaxing	Very Low	0.0402349	Group 1: R, Group 2: D and N	Yes/t(914)=-21.505; p<0.001

Table 2. EMG Table with Emotions Pleasant (P) vs. Unpleasant (U) and Neutral (N)
NumPeaks = the number of peaks within the time window referring to a stimulus /the time window referring to a stimulus

EMG-Signal		Number of Peaks			
Statistical Analysis		Pattern of NumPeaks	Mean (NumPeaks)	Two Group t-Test	
				Groups	correlated significantly?/result
Emotion	Pleasant	Medium	1.5078111	Group 1: P, Group 2: N and U	Yes/t(504)=-23.422; p<0.001
	Unpleasant	High	2.0283725	Group 1: U, Group 2: N and P	Yes/t(504)=8.151; p<0.001

during the presentation of music rated as “disquieting” (D) and decreases again for music evaluated as “neutral” (N) or “relaxing” (R). Note that this effect is stronger for relaxing than for neutral rhythms. The different duration of the activation phases results from the fact that the music continues while the users input their rating.

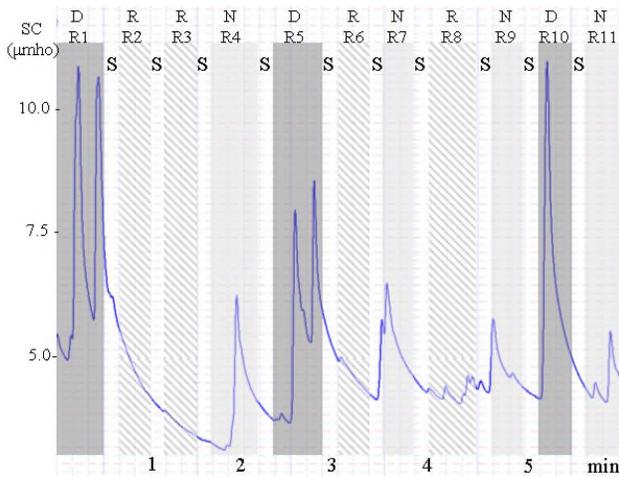


Figure 1. GSR during the Presentation of Rhythms (R1 ... R11) and Periods of Silence (S)

6. CONCLUSIONS AND OUTLOOK

In this paper, we presented a perceptual interface to an automated music composition system that adapts automatically by means of genetic optimization methods to the preferences of a user. In contrast to earlier work on automated music composition, our system is based on empirically validated physiological training data. First experiments have shown that there are indeed representative physiological patterns for a user’s attitude towards music which can be exploited in a music composition system.

Despite of first promising results, there are still many problems associated with physiological measurements. Well-known

pitfalls are uncontrollable events that might lead to artifacts. For example, we can never be sure whether the user’s physiological reactions actually result from the presented music or are caused by thoughts to something that excites him or her.

Another limitation is the great amount of data needed to train such a system. We recruited 10 subjects from Augsburg University for testing specific aspects of the generated rhythms, e.g. their GSR to disquieting rhythms. However, so far, only one subject underwent the full training program which took about 12 hours and is necessary to achieve a good adjustment of the system to a specific user. Our future work will concentrate on experiments with a greater number of subjects and the statistical evaluation of further physiological data.

More information on the music composition system along with various audio samples can be found under: http://mm-werkstatt.informatik.uni-augsburg.de/project_details.php?id=14

7. REFERENCES

- [1] Bartlett, D.L. Physiological Responses to Music and Sound Stimuli. In D.A. Hodges (ed.) Handbook of Music Psychology, MMB Music, 343-385, 1996.
- [2] Bradley, M.M., and Lang, P.J. Affective Reactions to Acoustic Stimuli. Psychophysiology, 37, 204-215, 2000
- [3] Lang, P. The emotion probe: Studies of motivation and attention. American Psychologist 50(5):372–385, 1995
- [4] Roads, C. The Computer Music Tutorial. Cambridge, MA: MIT Press, 1995.
- [5] Healey, J., Picard, R., Dabek, F. A New Affect-Perceiving Interface and Its Application to Personalized Music Selection, Proc. of the Perceptual User Interfaces Workshop, San Francisco, CA, 4-6, 1998
- [6] Schandry, R. Lehrbuch Psychophysiologie, Weinheim: Beltz PsychologieVerlagsUnion, 1998.
- [7] Vossel, G., Zimmer, H. Psychophysiologie, Urban Taschenbücher Bd. 553, Kohlhammer, 56-75, 1998.