Composing Affective Music with a Generate and Sense Approach

Sunjung Kim and Elisabeth André

Multimedia Concepts and Applications
Institute for Applied Informatics, Augsburg University
Eichleitnerstr. 30, D-86135 Augsburg, Germany
{andre,skim}@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.

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 behaviour, 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.

There is a high application potential for affective music players. Consider, for example, physical training. Various studies have shown that music has a significant impact on the performance of athletes. However, the selection of appropriate music constitutes a problem for many people since the music does not necessarily match their individual motion rhythm. A personalized coach could sense and collect physiological data in order to monitor the user's physical exercise and to keep him or her in a good mood by playing appropriate music.

In-car entertainment is another promising sector for adaptive music players. Nobody questions that a driver's affective state has an important impact on his or her driving style. For instance, anger often results into an impulsive and reckless behavior. The private disk jockey in the car might realize the driver's emotional state and play soothing music to make him or her feel more relaxed.

On the other hand, driving on a monotonous road may lead to reduced arousal and sleepiness. In such situation, soft music may even enhance this effect. Here, the personalized disc jockey might help the driver stay alert by playing energizing music.

Last but not least, an adaptive music player could be employed for the presentation of background music in computer games. Unfortunately, music in games usually relies on pre-stored audio samples that are played again and again without considering the dramaturgy of the game and the player's affective state. A personalized music player might increase a player's engagement in the game by playing music which intensifies his or her emotions.

To implement a music player that accommodates 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.

Measuring the Emotional Impact of Music

The most direct way to measure the emotional impact of music is to present users with various music pieces and asking them for their impression. This method requires, however, intense user interaction which increases the user's cognitive load and may seriously affect his or her perception of the music. In addition, asking users about their emotional state means an interruption of the experience. In the worst case, the user might no longer remember what he or she originally felt when listening to the music. Furthermore, inaccuracies might occur due to the user's inability or missing willingness to report on his or her true sensations.

Another approach is to exploit expressive cue profiles to identify the emotion a certain piece of music is supposed to

convey (Bresin and Friberg 2000). For instance, to express fear, many musicians employ an irregular tempo and a low sound level. While this approach offers an objective measurement, it does not account for the fact that different users might respond completely differently to music. Also, expressive cue profiles rather characterize the expressive properties of music and are less suitable to describe what a user actually feels.

Previous research has shown that physiological signals may be good indicators for the affective impact of music, see (Scherer and Zentner 2001) for an overview. The recognition of emotions from physiological signals bears a number of advantages. First of all, they help us to circumvent the artifact of social masking. While external channels of communication, such as facial expressions and voice intonation, can be controlled to a certain extent, physiological signals are usually constantly and unconsciously transmitted. A great advantage over self-reports is the fact that they may be recorded and analyzed while the experience is being made and the user's actual task does not have to be interrupted to input an evaluation. Nevertheless, there are also a number of limitations. First of all, it is hard to find a unique correlation between emotion state and bio signals. By their very nature, sensor data are heavily affected by noise and very sensitive to motion artefacts. In addition, physiological patterns may widely vary from person to person and from situation to situation.

In our work, we rely both on self-reports and physiological measurements. Self-reports are employed for new users with the aim to derive typical physiological patterns for certain emotional states by simultaneously recording their physiological data. If the system gets to know users, they are no longer required to explicitly indicate what they feel. Instead the system tries to infer the emotional state based on their physiological feedback.

A Music Composition Approach Based on Emotion Dimensions

The question arises of how the user should specify his or her emotional state. Essentially, this depends on the underlying emotion model.

Two approaches to the representation of emotions may be distinguished: a categorical approach (Ekman 1999) which models emotions as distinct categories, such as joy, anger, surprise, fear or sadness, and a dimensional approach (Lang 1995) which characterizes emotions in terms of several continuous dimensions, such as arousal or valence.

Arousal refers to the intensity of an emotional response. Valence determines whether an emotion is positive or negative and to what degree. Emotion dimensions can be seen as a simplified representation of the essential properties of emotions. For instance, stimulating music could be described by high valence and high arousal while boring music is rather characterized by low valence and low arousal (see Fig. 1).

In our work, we follow a dimensional approach and examine how music attributes that correspond to characteristic positions in the emotion space are reflected by physiological data which seems to be easier than mapping physiological patterns onto distinct emotion categories, such as surprise.

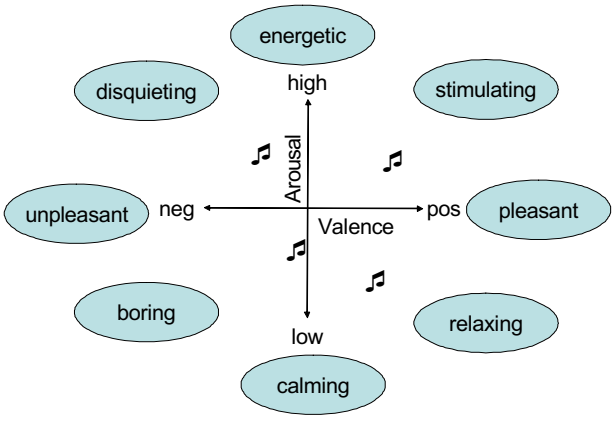

Fig. 1: Emotion Dimensions for Music

To measure the affective impact of music, we confront users with a set of automatically generated music samples and ask them to evaluate them with respect to pairs of attributes that correspond to opposite positions in the emotion space, for example "stimulating" versus "boring" or "energetic" versus "calming". To facilitate a clear distinction, we restrict ourselves to positions for which arousal and valence are either low, neutral or high. While the users are listening to the music and inputting their evaluation, their physiological response is recorded. Based on these data, the system tries to derive typical physiological patterns for the emotion attributes. For instance, the system might learn that energetic music tends to increase skin conductance.

The next step is to produce music that influences the user's arousal and valence in a way that corresponds to the positions of the attributes in Fig. 1. To accomplish this task, the candidates that represent a position best are combined by a genetic optimization process starting from randomly created solution candidates. The objective of this process is to obtain better solutions for each attribute after a number of reproduction cycles.

In a test phase, the affective state of the user is influenced by means of music samples that are selected with respect to their presumed effect on the valence and arousal dimensions. For instance, if the users' arousal is high and should be lowered, a relaxing, boring or calming music sample might be presented to them depending on whether we intend to activate them in a pleasant, unpleasant or neutral manner.

Experimental Setting

For training purposes, 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 users. The subjects had to indicate whether an attribute or its counterpart was satisfied. In case, none of the attributes applied, the music should be evaluated as neutral. In each session, the subjects had to concentrate just on one attribute pair. If subjects have to fill in longer questionnaires, there is the danger that they don't remember the experience any more after some time. While the subjects listened to the rhythms and inputted their evaluation, four types of physiological signals were taken using the Procomp+ sensor equipment:

- *Electrocardiogram (ECG)* to measure the subject's heart rate.
- *Electromyogram (EMG)* to capture the activity of the subjects' shoulder musculature.
- *Galvanic Skin Response (GSR)* to measure sweat secretion at the index and ring finger of the non-dominant hand.
- *Respiration (RESP)* to determine expansion and contraction of the subjects' abdominal breathing.

The ECG signal was taken with a sampling rate of 250 samples per second, the EMG, the GSR and the RESP signal with a sampling rate of 32 samples per seconds. Following (Schandry 1998), 17 features were extracted from the ECG, 2 features from the EMG, 4 features from the RESP and 10 features from the GSR 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 1998) 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. After that, 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.

Music Generation with Genetic Algorithms

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 (Roads 1995) 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 2 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. For users the system knows already, the fitness is computed on the basis of their physiological response. For example, if our goal is to employ music for relaxation and the system predicts a relaxing effect on the basis of the determined physiological data, the chromosome is assigned a high fitness value.

Tables 1 and 2 illustrate the genetic evolution process. The experimental sessions 1-5 in Table 1 served to create populations with individuals that are supposed to disquiet or relax the user. In Session 1, the user was presented with 116 randomly generated rhythms. Five of the rhythms were classified by the user as relaxing, forty as disquieting and seventy-one as neutral, i.e. neither relaxing nor disquieting. The four most relaxing and four most disquieting individuals were chosen for reproduction and survival. As a result, we obtained two new populations each of them consisting of 20 individuals with either relaxing or disquieting ancestors. The same procedure was iteratively applied to each population separately until 20 generations were produced.

Table 1 shows that relaxing rhythms may be found rather quickly. For instance, already after 20 reproduction

cycles most of the individuals were perceived as relaxing. For disquieting rhythms, the evolution process was even faster. Already 10 reproduction cycles led to generations with rhythms that were, for the most part, classified as disquieting. As a reason for this difference we indicate that it was easier to generate rhythms with a negative valence than rhythms with a positive valence.

A similar experiment was conducted to create populations with individuals that correspond to pleasant and unpleasant rhythms (see Table 2). So far, we only produced 10 generations (instead of 20). Nevertheless, Table 2 shows that the algorithm is also able to find pleasant and unpleasant rhythms after a few generations.

Correlating Subjective Measurements with Objective Measurements

As shown in the previous section, the genetic evolution process results into rhythms that match a certain attribute quite well after some re-production cycles. The question arises of whether the subjective impression of users is also reflected by their physiological data.

After a first statistical evaluation of the experiment, the GSR-signal was identified as a useful indicator for the attributes "disquieting" and "relaxing".

Table 3 provides a comparison of the GSR for "disquieting" and "relaxing" rhythms. In particular, a very low GSR indicates a relaxing effect while a higher GSR 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 1996).

To discriminate between positive and negative emotional reactions to music, EMG measurements have been proven promising. A study by (Lundquist et al. 2000) detected increased zygomatic EMG (activated during smiling) for subjects that were listening to happy music as opposed to sad music. Earlier studies by (Bradley and Lang 2000) 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. As shown in Table 4, 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. Fig. 2 shows how the amplitude of the GSR increases 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.

Finally, we evaluated whether the improvement of later generations were reflected by the user's physiological data. Our statistical evaluation revealed that this is indeed the case. But, the effect was more obvious for disquieting than for relaxing, pleasant or unpleasant rhythms.

Fig. 3 and Fig. 4 show the GSR curves for randomly generated rhythms before and after the evolution process. It can easily be seen that the curve in Fig. 4 is more characteristic of disquieting rhythms than that in Fig. 3.

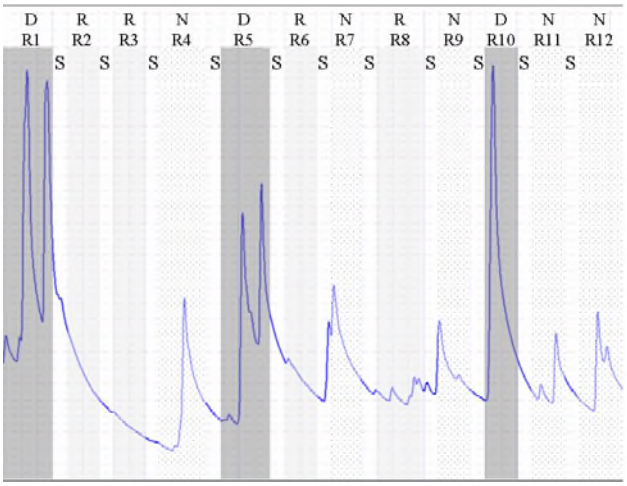

Fig. 2: GSR during the Presentation of Rhythms (R1 … R12) and Periods of Silence (S)

Fig. 3: Randomly Generated Rhythms before Evolution Covering a Time Period of 1:05'

Related Work

Early experiments to derive auditory responses from brainwaves and biofeedback of the performer were conducted by (Rosenboom 1977-1984), a pioneer in the area of experimental music. The main motivation behind his work is, however, to explore new interfaces to musical instruments to create new aesthetic experiences and not to compose music that elicits a certain emotional response.

A first prototype of a mobile music player was developed at MIT Media Lab by (Healey et al. 1998) who illustrated how physiological data could be employed to

direct the retrieval of music from a data base. More recent work at Fraunhofer IGD focuses on the development a music player that adjusts the tempo of music to a runner's speed and body stress (Bieber and Diener 2003). In contrast to the work above, we don't select music from a data base, but generate it automatically using a genetic optimization process.

Fig. 4: Disquieting Rhythms after Evolution Covering a Time Period of 1:10'

For this reason, we are not only able to adapt the music tempo to a user's affective state as in the case of the Fraunhofer IGD player, but also to other musical variables, such as instrumentation. In addition, we consider a great number of short samples (around 1500) as opposed to a few complex music pieces, e.g. ten in the case of (Healey et al. 1998). Therefore, we don't have to cope with the problem that the response to an arousal stimulus decreases because the chance of repetition is very high.

A number of automated music composition systems are based on genetic algorithms as ours. However, they usually rely on explicit user statements (Biles 2002) or music-theoretical knowledge (Wiggins et al. 1999) to assess the chromosomes while our system also considers the user's physiological feedback. Furthermore, the authors of these systems are less interested in generating music that conveys a certain emotion, but rather in finding a collection of music samples that matches the user's idea of what a certain music style should sound like.

In contrast, (Casella and Paiva 2001) as well as (Rutherford and Wiggins 2002) present systems that automatically generate music for virtual environments or films that is supposed to convey certain emotions. Their music composition approach is similar to ours. However, they don't aim at objectively measuring the affective impact of the generated music using physiological data.

Conclusions

In this paper, we presented a perceptual interface to an automated music composition system which adapts itself by means of genetic optimization methods to the preferences of a user. In contrast to earlier work to 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 an automated 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 artefacts. 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 programme 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 features.

References

Bartlett, D.L. 1996. Physiological Responses to Music and Sound Stimuli. In D.A. Hodges ed. *Handbook of Music Psychology*, pp. 343-385.

Bieber, G.; and Diener, H. *StepMan und akustische Lupe.* Fraunhofer IGD, Institutsteil Rostock, http://www.rostock.igd.fhg.de/IGD/files/IGD/Abteilungen/AR3/download/pdf/stepman_a3.pdf

Biles, J.A. 2002. GenJam: Evolution of a Jazz Improviser. In: Bentley, P. J.; and Corne, D. W. eds. *Creative Evolutionary Systems*, Academic Press, pp. 165-187.

Bradley, M.M.; and Lang, P.J. 1990. Affective Reactions to Acoustic Stimuli. *Psychophysiology* 37:204-215.

Bresin, R.; and Friberg, A. 2000. Emotional Coloring of Computer-Controlled Music Performances. *Computer Music Journal* 24:4:44-63.

Casella, P.; and Paiva, A. 2001. MAgentA: An Architecture for Real Time Automatic Composition of Background Music. In Proc. of *IVA 01*, Springer: NY.

Ekman, P. 1999. Basic Emotions. In Dalgleish, T. and Power, M. J. eds. *Handbook of Cognition and Emotion*, pp. 301-320. John Wiley, New York.

Healey, J.; Picard, R.; and Dabek, F. 1998. A New Affect-Perceiving Interface and Its Application to Personalized Music Selection, In *Proceedings of the 1998 Workshop on Perceptual User Interfaces*, 4-6, San Fransisco, CA.

Lang, P. 1995. The emotion probe: Studies of motivation and attention. *American Psychologist* 50(5):372–385.

Lundquist, L.G.; Carlsson, F.; and Hilmersson, P. 2000. Facial Electromyography, Autonomic Activity, and Emotional Experience to Happy and Sad Music. In: *Proc. of 27th Interational Congress of Psychology*, Stockholm, Sweden.

Roads, C. 1995. *The Computer Music Tutorial*. MIT Press.

Rosenboom, D. 1977-1984. On Being Invisible: I. The qualities of change (1977), II. On being invisible (1978), III. Steps towards transitional topologies of musical form, (1984). *Musicworks* 28: 10-13. Toronto: Music Gallery.

Rutherford, J.; and Wiggins, G.A. 2002. An Experiment in the Automatic Creation of Music which has Specific Emotional Content. In *7th International Conference on Music Perception and Cognition*, Sydney, Australia.

Schandry, R. 1998. *Lehrbuch Psychophysiologie,* Psychologie Verlags Union, Weinheim Studienausgabe.

Scherer, K.R.; and Zentner, M.R. 2001. Emotional Production Rules. In Juslin, P.N.; and Sloboda, J.A. eds. *Music and Emotion: Theory and Research*, Oxford University Press: Oxford, 361-392.

Vossel, G.; and Zimmer, H. 1998. *Psychophysiologie,* W.Kohlhammer GmbH, 56-75.

Wiggins, G.A.; Papadopoulos, G.; Phon-Amnuaisuk, S.; and Tuson, *A. 1999.* Evolutionary Methods for Musical Composition. *Int. Journal of Computing Anticipatory Systems.*

Experiment 1: Production of Different Populations for Relaxing and Disquieting Rhythms		**Random Phase**	**Evolution of Relaxing Rhythms**		**Evolution of Disquieting Rhythms**	
		Session 1:	Session 2: Relaxing Rhythms	Session 3: Relaxing Rhythms	Session 4: Disquieting Rhythms	Session 5: Disquieting Rhythms
Duration		1:05':49''	1:50':59''	1:51':06''	1:59':18''	1:49':18''
# Evaluated as Relaxing		5	120	172	0	1
# Evaluated as Disquieting		40	40	20	159	199
# Evaluated as Neutral		71	40	8	41	0
Overall Number	Produced Rhythms/Generations	116	200/1-10	200/11-20	200/1-10	200/11-20

Table 1: Illustration of the Evolution Process for Relaxing and Disquieting Rhythms

Experiment 2: Production of Different Populations for Pleasant and Unpleasant Rhythms		**Random Phase**	**Evolution of Pleasant Rhythms**		**Evolution of Unpleasant Rhythms**	
		Session 1:	Session 2: Pleasant Rhythms	Session 3: Pleasant Rhythms	Session 4: Unpleasant Rhythms	Session 5: Unpleasant Rhythms
Duration		59':04'	55':52''	56':20''	56':17''	55':01''
# Evaluated as Pleasant		18	45	31	0	0
# Evaluated as Unpleasant		18	21	3	63	80
#Evaluated as Neutral		70	34	66	37	20
Overall Number	Produced Rhythms/Generations	106	100/1-5	100/6-10	100/1-5	100/6-10

Table 2: Illustration of the Evolution Process for Pleasant and Unpleasant Rhythms

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

Table 4: EMG Table with Emotions Pleasant (P) Versus Unpleasant (U) and Neural (N)
NumPeaks = number of peaks within the time window corresponding to a stimulus