2006－MUS－68 (15)
2006／12／16

“Molto sentito”: 情感溢れるコンピュータ音楽を目指して

五十嵐創,† 岩竹徹,† 古川康一 †

† 慶應義塾大学大学院 政策・メディア研究科
{soh,iwatake,furukawa}@sfc.keio.ac.jp

概要: 本稿では第一著者の近作 “Molto sentito” について，その思想および制作手法を述べる．中心的な問題意識は，コンピュータ音楽において，いかに人間が日々実感する生々しい感情や情感を豊かに表現できるかという点にある．著者は，人間の感情の形式化に関する先駆的研究である Sentics と遺伝的アルゴリズムを組み合わせることにより，特定の性格を備えたサウンドを進化的に作り出す音響生成手法，evolutionary essentic sound を提案した．“Molto sentito” 中で用いられた主な音響素材は本手法によって作られている．本稿ではそのアルゴリズムを説明する．

“Molto sentito”: Toward Computer Music Full of Emotion

Soh Igarashi†, Toru Iwatake†, and Koichi Furukawa†
†Graduate School of Media and Governance, Keio University

Abstract: This paper concerns a first author's recent piece “Molto sentito”. The design concept and the composing technique are described. Our main motivation is emotion communication in computer music. We proposed a new sound synthesis technique named evolutionary essentic sound, which combines Dr. Manfred Clynes' Sentics theory about human emotion with genetic algorithms and evolutionarily generates such sound as expresses an emotion. Most sound materials used in “Molto sentito” are composed with this method. The algorithmic details are given.

1 Introduction

The impact of the computer against music is immeasurable. It is not too much to say that the traditional concept of music has been drastically changed with the appearance of the computer. It means that today we have to reconsider the fundamental issues: what is music, or what is the most essential requirement for being music?

In traditional western music, for example, tonal system has been the dominant backbone for hundreds of years and counterpoint and harmony have been developed as the representative principles of music composition[1].

Now, in the field of computer music in the twenty first century, every composer seeks to develop and refine his/her own music language. The authors' main motivation is emotion expression and emotion communication in computer music and the ultimate goal is to create a new musical composition which fully expresses and communicates human emotion. “Molto sentito” presented at the Intercollege Computer Music Concert 2006 can be regarded as the first step toward the goal.

Most sound materials used in “Molto sentito” are composed with a new sound synthesis technique named evolutionary essentic sound. The technique evolutionary essentic sound combines Dr. Manfred Clynes' Sentics theory about human emotion with genetic algorithms and evolutionarily generates such sound as expresses an emotion gradually rising from some chaotic sound. This paper describes the algorithmic details.

The paper consists of 5 sections. Section 2 introduces Sentics and section 3 describes the technique evolutionary essentic sound in detail. Section 4 gives a discussion and section 5 concludes the paper and gives future work.

2 Sentics

2.1 Essentic Forms

Sentics is the neurophysiological framework approaches to human emotion and was proposed by Dr. Manfred Clynes, a professional pianist and a researcher in the field of neuroscience and engineering[2]. The most important point that distinguishes Sentics from other similar propositions consists in its clear motivation to analyze human emotion by a formal and objective way. According to Sentics, typical human emotions we daily have can be represented by different envelopes called *essentic forms*. Clynes and his colleagues derived this claim from their long years of experimental study. They have developed an original press-button shaped emotion measuring device named *sentograph* and succeeded in capturing emotion expression graphically.

Clynes took 8 kinds of emotion: *love, hate, grief, joy, reverence, anger, sexual desire* and *no emotion*, as an object of study. A subject was first asked to fill his/her mind with a particular emotion and to express the emotion by pushing the button of a *sentograph* like piano performance. A *sentograph* is equipped with a transducer for measuring finger pressure. When a subject presses the button, the horizontal and vertical components of finger pressure are recorded as two separate envelopes. For each emotion thirty to fifty trials were measured and their envelopes were averaged. Then it was discovered that for each of the 8 emotions there is a particular time-varying pattern of finger pressure that is common to many subjects. Clynes has named these envelopes *essentic forms* and insisted that they have an essential relationship with human emotion. Interestingly, the result of further experiments performed over the world showed that *essentic forms* are universally observed without depending upon sex or cultural difference.

2.2 Essentic Sound

If one accepts Clynes' hypothesis that *essentic forms* reflect some essential property of human emotion, it is natural to think that it can be possible to generate such sound as communicates the qualities which *essentic forms* represent by utilizing *essentic forms* as material for sound synthesis. This is a very interesting challenge to anyone who wants to compose expressive music. For this question, Clynes has given an answer himself[3].

In [3], Clynes has proposed the transformation of *essentic form* to the sound reflecting the same quality as the original *essentic form* has for each of the 7 emotions except *no emotion*. Each proposed sound comprises frequency and amplitude envelopes. Those frequency and amplitude envelopes are designed by transforming *essentic forms* in various ways (inverting an envelope, for example). Concrete values are given in Hz as to frequency envelopes. A schematic diagram of proposed envelopes for 7 emotions appears in Fig.1. Fig.1 indicates that 7 emotions have different time lengths respectively. Refer to [3] for the original graphs of these envelopes and the details of discussion Clynes made.

We call the 7 kinds of proposed sound "essentic sound" in this work. To accept Clynes' proposal in [3] represents the starting point of this work. We describe how to evolutionarily generate these essentic sound in the next section.

3 Generate Essentic Sound Evolutionarily

3.1 Essentic Envelopes

Evolutionary essentic sound is generated by some evolutionary method (genetic algorithm, for example). Essentic sound is the sound that expresses the same qualities as an essentic form has. This idea is originally introduced by Clynes[3], as stated in the previous section. In Clynes' framework, 7 emotions, namely *anger, grief, hate, joy, love, reverence*, and *sex*, have each essentic sound. Each essentic sound comprises frequency and amplitude envelopes. In this work these envelopes are called "essentic envelopes". Seven emotions have two essentic envelopes (frequency and amplitude) for each, so there are 14 essentic envelopes in all. These 14 essentic envelopes play an important role in this work.

In fact, the essentic envelopes used in this work are reconstructed as digital data by tracing their pictures appeared in [3], which might affect the result of this work to some extent.

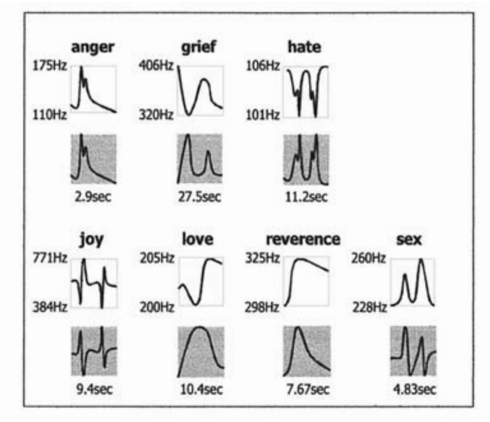

Figure 1: Transformation of *essentic forms* to the sound of the same quality. For each emotion, the upper square represents frequency envelope and the lower shaded square represents amplitude envelope. These 14 envelopes are called "essentic envelopes" in this work.

3.2 Apply Genetic Algorithms

Now we will describe how to apply genetic algorithms to the generation of essentic sound in detail. When using genetic algorithms, what we have to consider can be reduced to the following four points.

- What is an individual?
- What is the fitness function?
- What is the selection function?
- What are the genetic operators?

3.2.1 Individuals

In this work, individuals are sounds. Here, sounds are defined as a pair of envelopes (frequency and amplitude). What should be noted here is that at this stage of the algorithm an envelope is abstracted and only its shape is considered. Every envelope is normalized so that the maximum value is equal to or less than 1 and the minimum is equal to or more than 0. On a computer program every envelope (both frequency and amplitude) is represented as a real-valued array data of the same length (1024) and each element runs from 0.0 through 1.0.

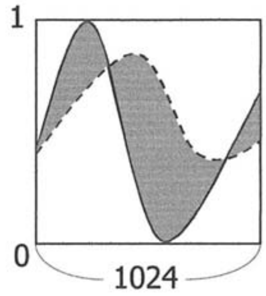

Figure 2: Similarity calculation between two envelopes.

These "abstracted" envelopes, namely sounds, are "concretized" when the sound is played. For how to concretize an individual, refer to a later section.

3.2.2 Fitness Function

The fitness of each individual is calculated according to the following steps.

1. Select one emotion out of seven emotions. The selected emotion is called "target emotion". The essentic envelopes of the target emotion are called "target envelopes".
2. For both frequency and amplitude, calculate the similarity between target envelope and the envelope of a given individual.
3. Calculate the average of the two similarities for frequency and amplitude, and return it as the fitness of the individual.

The similarity between two envelopes is calculated using the difference between their square measures. See Fig.2. Let the dotted line be the envelope of a given individual and the solid line be the essentic envelope. The shaded area indicates the difference between two envelopes, so smaller shaded area results in higher similarity.

Note that essentic envelopes are also abstracted as well as individuals when calculating similarity. This abstraction process (max corresponds to 1, min to 0, and the same data length) makes it possible to compare all envelopes in the same method stated above.

3.2.3 Selection Function

The probability with which an individual is selected as a parent for reproduction is weighted in accordance with its fitness. An individual of higher fitness value more likely to be selected as a parent. Each mating of two parents produces only one child.

3.2.4 Genetic Operators

Genetic operators for reproduction comprise crossover and mutation. Crossover takes an average of the envelopes of the parent for frequency and amplitude respectively. Mutation takes an average of the target envelope and a newly produced child by crossover. By the definition of crossover here, it can occur that the fitness scores of all individuals converge on the same value and the evolution process stops. The mutation operation plays an important role in achieving an evolution without falling into such premature convergence. The probability of mutation is set to 0.05.

When a genetic algorithm is used for solving an optimization problem, mutation is usually introduced so as not to be stuck in a local optimum. From such a point of view, the above setting of mutation seems almost like "cheating". However, the objective here is to create a new synthesized sounds and the quality of generated sounds seems to suggest the effectiveness of the authors' setting of mutation. Further discussion on a better configuration of mutation is given later.

One generation comprises 100 individuals. The individuals of the first generation are made by random combination of essentic envelopes.

On the experimental settings described above, it has been substantiated that an evolution is successfully achieved for every target emotion.

3.3 Map onto Music Parameters

Having described the algorithmic details so far, we will now explain how to relate the evolutionary framework mentioned above to musical aspects. The aim of this work is to generate essentic sound using evolutionary method and to communicate the essentic quality of an emotion, that is to say, to make a listener clearly feel that emotion. The subtle setting of music parameters described below was eventually determined based on the authors' subjective judgment for that purpose.

This work outputs an audio file every generation. When the algorithm runs to the 7th generation, for example, 7 audio files are generated. One audio file has 100 individuals, namely 100 sounds. For each sound, the following four musical parameters, frequency, amplitude, duration, and timbre, are controlled in accordance with its fitness.

1. Frequency
2. Amplitude
3. Duration
4. Timbre

3.3.1 Frequency

The frequency of a sound gets closer to the value given by Clynes[3], as its fitness becomes higher. Let us suppose that the target emotion is *joy* and the fitness of an individual is 1. For *joy*, the maximum and the minimum frequency is set to 771Hz and 384Hz respectively in [3], so the sound (individual) is played at that range of frequencies. The lower the fitness, the more error is added.

3.3.2 Amplitude

In the case of amplitude, [3] has not stated how loud essentic sounds should be played, whereas the shapes of the essentic amplitude envelopes have been given. In this work the maximum amplitude value of a sound is determined based on its fitness. A sound is played louder as the fitness becomes higher. Concretely speaking, the amplitude of a sound becomes higher in proportion to the fitness.

Since the fitness of a sound is calculated in consideration of the shape of its amplitude envelope by the definition of the fitness function, this assignment method of amplitude seems to reflect Dr. Clynes' original concept.

3.3.3 Duration

As in the case of frequency, the duration of a sound gets closer to the value given by Clynes, as its fitness becomes higher. Suppose that the target emotion is *joy* again. According to [3], the duration of the essentic envelopes for *joy* is approximately set to 9.4 seconds, so an individual whose fitness is 1 is played in 9.4 seconds. Another individual whose fitness is 0.5 is played in half the duration of the essentic envelope, namely in 4.7 seconds in this case.

Though one generation includes 100 sounds of different durations, It is not necessarily the case that all sounds start simultaneously. In the case of an individual whose fitness is less than 1, its start time can vary to the extent that its end time does not exceed the duration of the essentic envelope.

3.3.4 Timbre

In this work, the waveform of a sound is synthesized by harmonic addition of a simple sine wave. As the fitness of a sound increases, namely evolution proceeds, its timbre can be changed gradually from that of a saw-tooth wave to that of a square wave. An individual whose fitness is 0 is played with a full saw-tooth wave which comprises 13 harmonic partials. Increase in the fitness of an individual reduces the even harmonic partials of the sound. Now, let p be the fitness of a sound, the strength of an even harmonic partial of the sound is multiplied by $1 - p^2$. An individual whose fitness is 1 is played with a full square wave. Take a sound whose fitness is 0.5 for example. Its even harmonic partials are played in 0.75 times as loud as its odd harmonic partials.

3.4 Change Target Emotion

When an evolution process to some target emotion is almost finished and the average fitness of individuals is sufficiently high, the algorithm stochastically changes the target emotion and begins a new evolution process to another target emotion.

This operation is exploratively employed in order to avoid the monotony and to acquire the variety of sound. This concept was originally developed by Magnus[6]. The authors have been inspired by the paper.

You can judge whether this function effectively works by listening to the sound examples introduced below. In those examples, this change-target function was set to be able to operate only if the average fitness of individuals is more than or equal to 0.9.

3.5 What Sounds Are Generated?

As stated in 3.3, this work outputs an audio file every generation. Since the first generation comprises individuals of low fitness values, the output sound is usually full of chaotic sound. Sound generated gradually converges on a particular essentic sound over generations. By the nature of the algorithm employed, the average fitness of all individuals increases monotonically. When an evolution process is proceeded to some extent, many individuals of very close frequency sound together.

When the convergence, namely evolution, is achieved to a certain extent, the target emotion may be changed to another one. This is due to the change-target function stated above. If it is done, the sound becomes chaotic again and starts new evolution process to another essentic sound.

Some examples of the synthesized sound can be obtained through the author's web page[5]. Visit there and read instructions. There are 3 examples. Each of them includes 14 to 20 generations and covers several evolution processes to different target emotions. The authors hope that these examples help readers realize what this work has done.

4 Discussion

Let us now evaluate this work and discuss the significance of it. The first point that we should discuss is whether the original goal we set was achieved or not. The initial aim of us is to synthesize essentic sound evolutionarily and to generate such sound material as expresses certain emotion gradually rising from some chaotic sound. The authors think that to some extent the goal is obtained. Listening to synthesized sound, we can surely recognize some convergence to a particular target through a gradual formation of the particular shape of both frequency and amplitude envelopes and through a gradual change of timbre. This fact seems to show the effectiveness of an evolutionary method as a tool for sound synthesis and support the authors' way to apply genetic algorithm to music domain.

On the other hand, it is not very easy to be affirmative as to whether evolutionarily generated essentic sounds certainly rouse some particular emotions. Though the authors admit that they have not found the apparent reason for this point, it seemed to need to study Sentics in more detail and to seek a more effective way to apply the framework of Sentics to sound synthesis and composition.

With these issues in mind, we will now shift the emphasis to the discussion on the significance of this work from the viewpoint of a new methodology of music composition. This work can be categorized as one of the challenges to seek a new effective methodology of composition in computer music and its central point consists in the attention to human emotion. Broadly speaking, human emotion and emotion expression have not been the main research themes in the history of computer music, though they seems to be one of the most interesting topics to many composers and listeners. In particular, it is not an exaggeration to say that there is no established theory to convincingly explain how music arouses emotion in listener's mind. Sentics is a very unique framework in the sense that it directly tackles the issue of human emotion in the musical point of view. Admitting the present status is immature, this work seems to be of significance as a case study applying Sentic framework to sound generation.

5 Conclusion

This paper described a new sound synthesis technique named evolutionary essentic sound, with which "Molto sentito" was composed. We will conclude the paper by discussing the next step should be taken so as to improve this work. Various future research directions are conceivable.

Exploring a more effective setting of genetic and music parameters will be well worth a trial. As regards mutation, for example, it is used entirely to promote evolution in the current setting. In addition to it, introducing a new mutation operation which suppresses evolution may contribute to generate sounds of more variety in character. The authors also think there is room for improvement in the way of determining the duration of a generated sound. It can be worth consideration to lengthen a sound of low fitness as well as to shorten.

From a broader perspective one of the most interesting future steps seems to compose more musical works by using the algorithm described here. The history of music obviously tells that the worth of a newly proposed composing technique is only demonstrated by musical works composed with the proposed technique. The authors think that they can claim the significance of this work with full confidence only if fully attractive and convincing musical works are composed with the method described in this paper. "Molto sentito" is merely a first challenge toward the goal.

Acknowledgments.

The authors thank Dr. Manfred Clynes, who is the inventor of Sentics and made a lot of useful comments on this work.

References

[1] Burkholder, J.P., Grout, D.J., and Palisca, C.V.: A History of Western Music, 7th edition, W.W.Norton & Company, 2006.

[2] Clynes, M.: Sentics, the touch of emotion, Doubleday Anchor, New York, 1977.

[3] Clynes, M. and Nettheim, N.: The Living Quality of Music, Neurobiologic Patterns of Communicating Feeling, Music, Mind, and Brain (The Neuropsychology of Music), Ed. Manfred Clynes, Plenum Publishing Corp., pp.47-82, 1982.

[4] Digital Art Awards 2006, http://daa.sfc.keio.ac.jp/2006/en/index.html

[5] The first author's web page. http://web.sfc.keio.ac.jp/~soh/EvoSent/examples.html

[6] Magnus,C.:Evolutionary Musique Concrète, Applications of Evolutionary Computing, Proceedings of the 4th European Workshop on Evolutionary Music and Art(EvoMUSART), pp.688-695, 2006.