Operating Sound Parameters Using Markov Model and Bayesian Filters in Automated Music Performance

Fumito Hashimoto
Kyushu Institute of Technology
1-1 Sensui-cho Tobata Kitakyushu
Fukuoka, Japan
n350923@tobata.isc.kyutech.ac.jp

Motoki Miura
Kyushu Institute of Technology
1-1 Sensui-cho Tobata Kitakyushu
Fukuoka, Japan
miuramo@mns.kyutech.ac.jp

ABSTRACT
In recent years, there has been an increase in the number of artists who make use of automated music performances in their music and live concerts. Automated music performance is a form of music production using programmed musical notes. Some artists who introduce automated music performance operate parameters of the sound in their performance for production of their music. In this paper, we focus on the music production aspects and describe a method that realizes operation of the sound parameters via computer.

Further, in this study, the probability distribution of the action (i.e., variation of parameters) is obtained within the music, using Bayesian filters. The probability distribution of each piece of music is transformed by passing through a Markov model. After the probability distribution is obtained, sound parameters can be automatically controlled. We have developed a system to reproduce the musical expressions of humans and confirmed the possibilities of our method.

Keywords
Automated music performance, Sound parameters, Markov model, Bayesian filters, PureData, Kraftwerk, Techneus

1. INTRODUCTION
In recent years, concerts in which artists incorporate automated music performances have become popular. Automated music performance is a form of music production that uses automatically programmed musical notes. This approach is typically used to enhance the musical expression rather than substitute humans. A concert that incorporates automated performance is characterized by programmed musical notes interleaved with the sound production process. For example, some artists introduce Arpeggio and sound effects for enhancing their own music. In these examples, the ratio of programmed musical notes is relatively low.

Kraftwerk [10], an electronic music group in Germany, is the originator of this approach. Their sounds were created on the basis of programmed musical notes, and they operate sound parameters in their concerts to maximize the effect of the musical expression, thereby enhancing audience reaction and emotion. Kraftwerk is a pioneer in electronic music that has released many masterpieces in the 1970s (including Autobahn[1]). The group is still active today, and their songs are showcased live in different forms. Further, their music could continue by using production robots when they are no longer alive [1]. If this is realized, theoretically, they can stay active forever.

We consider this long term continuous work to heighten the value of the production. We also respect those songs of Kraftwerk that have gradually changed their shape over several decades. In addition, the electronic music group, led by first author of “Techneus [11]” is also incorporated into the concert production such as Kraftwerk. He consider that performance by computer to enable the long-term activity. However, it is not clear how the operation of the sound parameters that they performed in real time during a concert are to be reproduced when they are no longer here. Accordingly, we aim to automatically reproduce their sounds, using appropriate operating parameters. When the operation of the necessary sound parameters by computer is realized, it is then possible to enrich the experience of the audience, because they can listen to a varied sound that evokes a potentially different feeling each time. Also, from the point of view of the performer, when he/she was troubled at operating sound parameters, the performer can determine the operating of parameters. And there is a possibility that determination principle of operating sound parameters will become clear.

To reproduce the sound, we first need to gather its characteristics because they signify the artists’ features on musical expressions. The characteristics can be extracted by performance of the artists, such as playing a musical instrument and operating sound parameters. In this paper, we propose a method that realizes operation parameters, using a Markov model and Bayesian filters to produce automated performance.

In addition to this introduction, the remainder of the paper is organized as follows: Section 2 describes a method for obtaining operation parameters and an algorithm to operate these sound parameters, Section 3 describes our evaluation and experimental methods, and our conclusions are summarized alongside our plans for future work in Section 4.

2. RELATED WORK
In the field of Music Information Retrieval, there are several researches for identifying/classifying music incorporating machine learning techniques. For genre classification, Hamel and Eck [2] proposed a method to classify the music into 10 genres. They employed Distributed FFT for extracting features. Panagakis [3] utilized LPNTF method to extract music genre. In this study, a robust music genre classification framework has been proposed. This framework resorts to cortical

[1] Autobahn is an album with the theme of a highway; it was a one of Kraftwerk’s hits.

representations for music representation, while sparse representation-based classification has been employed for genre classification. For identifying performers, Stamatatos and Widmer [4] reported a computational approach to the problem of discriminating between music performers playing the same piece of music, and introduced a set of features that capture some aspects of the individual style of each performer. For identifying instruments, Hamel and Eck showed and compare several models for automatic identification of instrument classes in polyphonic and poly-instrument audio [5]. Mantaras and Arcos [6] use AI (specifically: machine learning) techniques in an attempt to express the individuality of music performers (pianists) in machine-interpretable terms by quantifying the main parameters of expressive performance. The results show that the differences between music performers can be quantified. Our aim is to reproduce performers (sound operators) by means of Baysian filters and Markov model in automated music performance.

3. METHOD

In this section, we describe our method to reproduce sounds, using operating parameters. First, we explain our system configuration. Second, we describe how the characteristics of the artists can be extracted from the operation data.

3.1 Obtaining operation data

First, we present a method to obtain operation data of sound parameters. Automated music performance is produced by performance data such as a MIDI controller and tone generator. To read performance data, we utilize a digital audio workstation (DAW) called Cubase7 [12]; further, we employ SYNTH-WERK [13], a synthesizer tone generator. We also use MIDI controller BCF2000 [14], i.e., Behringer's equipped with 32 types of encoders and eight faders, to reduce the burden of controlling the sound parameters. The MIDI controller has eight knobs and eight sliders (hereafter, called encoders) and four buttons to switch the functions of the encoders. Figure 1 shows the mapping of the encoders on the MIDI controller with the parameters of SYNTH-WERK. Generally, the parameters are related to each component of a song (e.g., bass, percussion, cantus firmus, and so on). Using the MIDI controller, artists can intuitively change the sound effects produced.

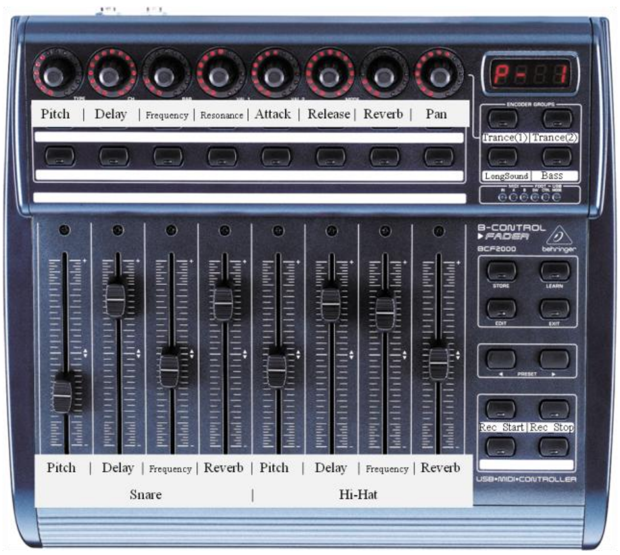

Figure 1. Controller mapping of encoders with parameters of SYNTH-WERK

As shown in figure 2, we have developed a system that records values of the encoders in real time, using Java and PureData (Pd) [15], a visual programming language for multimedia and desktop music creation that produces Pd patches. When Pd receives changes to the values of the encoders, it transmits the new values to the Java program via UDP. The Java program stores the changes of the values along with timing data. Since the MIDI controller is also connected with SYNTH-WERK, the movement of the encoders directly affects the sound generated by the tone generator.

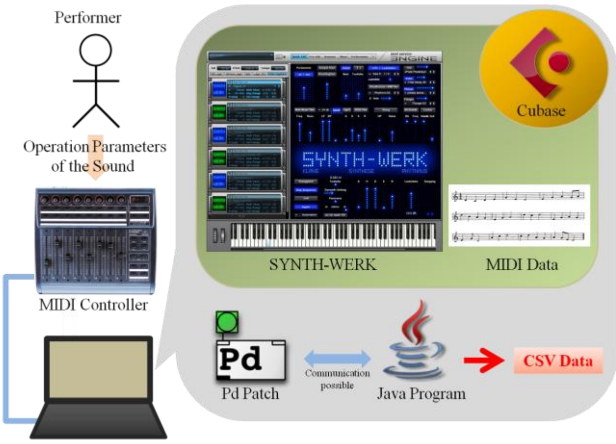

Figure 2. Schematic showing how the system obtains operation data of sound parameters

3.2 Operation of sound parameters by computer

In this section, we present a method of operation of the sound parameters by computer from the obtained data. The obtained data consists of timing data and IDs and values of the encoders when they were moved. The timing data shows a millisecond from beginning of the song.

First, to eliminate differences in beats per minute of the song, we normalize the timing data to match the number of beats recorded from the beginning of the song to its end. Second, movement events of the encoders are separated by their encoder IDs. We call these events "actions."

Next, we relate the actions with a corresponding music scene. The music scene represents a combination of parts that compose the song. In our research, we utilized songs with seven parts. Therefore, the music scene is defined by a value using seven bits, as shown in Figure 3. We have prepared a music scene for each beat by considering the performance data.

Thus organizing the data, the probability distribution of the action that will occur is obtained by Bayes' theorem. In the process for computing the probability distribution, we divided value changes of the parameters into eleven levels and generated histograms of actions for each music scene and encoder ID.

From the actions of the encoder ID, we observed that each action that occurs was highly related to a previous action. For example, an action to lower a value often appears after an action to increase the same value; i.e., we assume that the actions of each encoder follow the Markov process. Thus, to determine the probability distribution of state transitions in the encoder, we generated the first-order Markov model [7]. The number of states in the model was set to eleven, similar to the level of the probability distribution. We applied the probability distribution to the Markov model to transform the probability of action that occurs in each music scene. For example, an action which changes encoder ID 2 under the music scene ID was 23. Then, the distribution which represents width of parameter value appears, as shown in Figure 4 top. In order to transform the distribution, we apply the encoder ID 2 of the Markov model as shown in the Figure 4 middle. When the previous action increased the value of encoder ID 2 +100, the distribution of the corresponding row in the Markov model is added to the distribution of the action as shown in the Figure 4 bottom. In this way, distributions of actions are transformed.

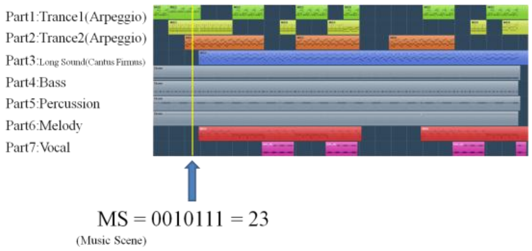

Figure 3. An example of music scene encoded using seven bits

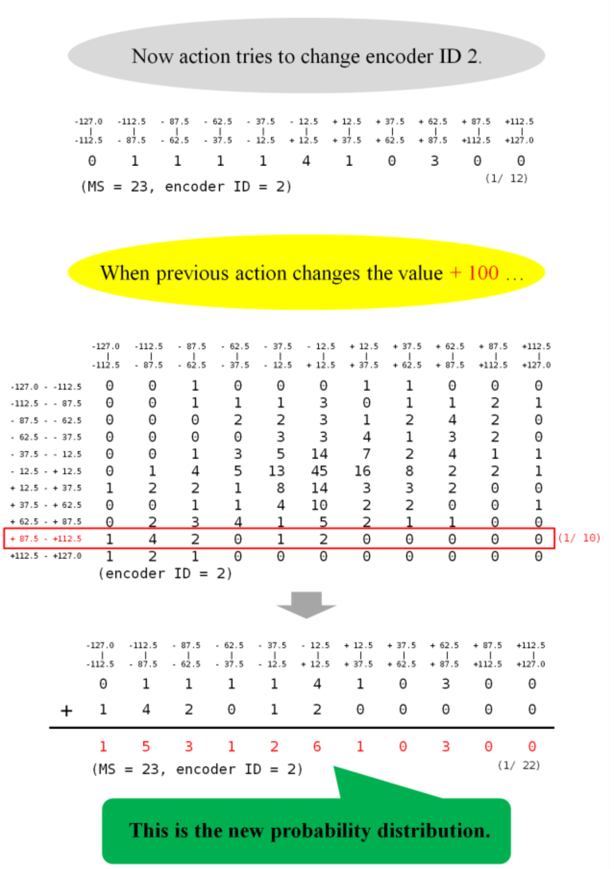

Figure 4. Transforming of probability distribution by Markov model.

Through this procedure, we generated a model from training data. This model represents the characteristics of artists, and the accuracy of the model can be increased by increasing the size of the training data.

As shown in Figure 4, after the model is generated, we can obtain actions by inputting music scenes of other songs. These actions also contain timing data and IDs and values of the encoder. When the play button on the MIDI controller is pressed, the song starts via Cubase7 and the event is sent to our Java program. It transmits the actions to PureData, which in turn controls the corresponding parameters by moving the MIDI controller. When the encoder of the MIDI controller is moved, the sound is affected by the effect, thereby operating sound parameters to the computer.

4. EXPERIMENTATION

We describe our preliminary experimentation below, as well as our subjective assessment of the system.

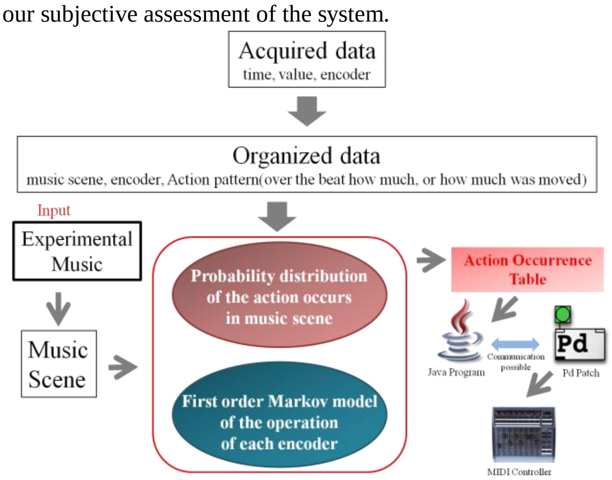

Figure 4. Schematic of operation of sound parameters by computer

4.1 Preliminary experimentation

Figure 5 shows the experimental equipment of our system, as described in Section 2. As experimental songs, we utilized works by Techneus [11], an electronic music group influenced by Kraftwerk. Also note that the first author of this paper is the prime member of the group.

The other member of the group was a participant of the experiment. The participant operated the MIDI controller while watching the Music Scene Monitor, the computer screen shown in Figure 5. We prepared three songs for our experiment. We asked the participant to produce sounds three times for each song; therefore, we obtained nine operation datasets. We generated a model that represents the participant's characteristics from this data. Finally, we obtained three automatically generated songs by applying music scene transitions of the three original songs.

4.2 Subjective assessment

From our experimentation, we found that our system could successfully reproduce songs with controlling parameters by the generated actions. However, when we listened to the songs generated, the movement of the parameters was chaotic, and the effect of the sound sometimes showed unnatural behavior. We considered the following reasons for these phenomena:

(1) The feeling of effect by changing parameters was not the same for qualities of tone, even if the parameters were same.
(2) Some parameters caused drastic changes in the feeling of sounds; however, the generated actions sometimes exceeded the proper range (e.g., pitch and cutoff). For such parameters, we should set restrictions on parameter values.
(3) Since the actions do not consider the current value of the parameter, some actions tried to increase (decrease) the corresponding value even when the value was already maximized (minimized).

(4) When the action changed pitch, it was required to follow the scale of the song. Otherwise, this caused the produced sounds to be out of scale.

We consider (1) and (2) solvable by determining the range of parameters in the qualities of tone settings. Regarding (3) and (4), special handling of parameters is required. For example, changing the pitch should be maintained within scaling rules.

In addition to the above issues, we observed that the size of the operation data also affects the results of the musical expression. We plan to continue our experimentation by adding more songs as well as new participants.

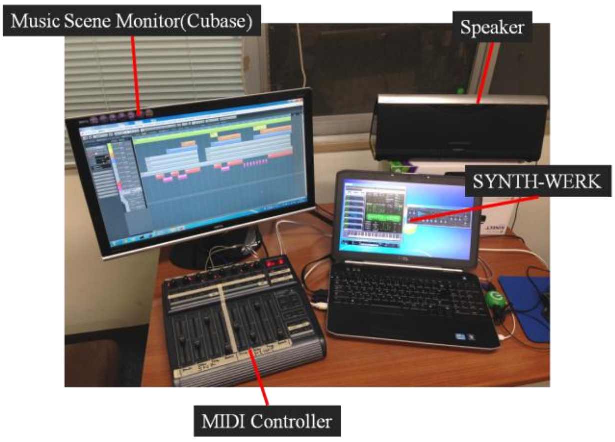

Figure 5. Experimental equipment

5. CONCLUSION

In this paper, we proposed a method to reproduce songs, based on changing parameters of artists. When our method was realized, it was possible to enrich the experience of the audience, because he or she could listen to a varied sound that evokes a different feeling each time.

To reproduce the sounds, we developed a system implemented in Java and PureData. The system captures parameter changes and generates actions that represent the characteristics of the performer, using a Markov model and Bayesian filters.

We performed a preliminary experiment to confirm the reproduction. The system worked successfully, but the generated sound sometimes included irrelevant musical expressions. We plan to resolve these issues and take the following future directions to further enhance reproducibility:

1. Feature value of music: We have considered that the tonality, cord change level, number of beats per minute, maximum beat level, maximum signal level, and average signal level can all augment the music scene. Further, the music atmosphere can be classified by these values [8]. To examine the causal relationship between parameter changes and the atmosphere of the produced music, we aim to incorporate feature values into our model.

2. Mel frequency Cepstral coefficients (MFCC): MFCC is the future value of audio data used in such applications as voice recognition [9]. MFCC determined from audio data will be incorporated into our model to enable us to examine the causal relationship between this and the movement of parameters. However, when considering MFCC as input, the computer needs substantial processing power because calculations must be made in real time.

The actions generated by the model in our experiment result in phenomena that differ from characteristics of the original participant; however, the results produced an attractive and appealing sound. In future work, we will focus on not only reproducibility, but also the quality of reproduction.

6. REFERENCES

[1] Pascal Bussy, Man-Machine and Music, Suiseisha, 1994.
[2] Philippe Hamel, Douglas Eck, Learning Features from Music Audio with Deep Belief Networks, The International Society for Music Information Retrieval, 2010.
[3] Yannis Panagakis, Constantine Kotropoulos, Gonzalo R. Arce, Music Genre Classification Using Locality Preserving Non-Negative Tensor Factorization and Sparse Representations, The International Society for Music Information Retrieval, 2009.
[4] Efstathios Stamatos, Gerhard Widmer, Automatic identification of music performers with learning ensembles, Artificial Intelligence 165 (2005) 37-56.
[5] Philippe Hamel, Douglas Eck, Automatic Identification of Instrument Classes in Polyphonic and Poly-Instrument Audio, The International Society for Music Information Retrieval, 2009.
[6] Ramon Lopez de Mantaras, Josep Lluis Arcos, AI and Music From Composition to Expressive Performance, AI Magazine Vol.23, 2002.
[7] Wei Chai, Barry Vercoe, Folk Music Classification Using Hidden Markov Models, International Joint Conference on Artificial Intelligence, 2001.
[8] Yasuteru Kodama, YasunoriSuzuki, Fumio Matsushita, Satoshi Odagawa, Shinichi Gayama,Takehiko Shioda, Development of “Feeling Play”, technical journal "PIONEEER R & D" Vol.14 No.3.
[9] Lisha Zhong, Jiangzhong Wan, Zhiwei Huang, Gaofei Cao, Bo Xiao, Heart Murmur Recognition Based on Hidden Markov Model, Journal of Signal and Information Processing, 2013.
[10] Kraftwerk http://www.kraftwerk.com/
[11] Techneus http://usiot.hanamizake.com/zbob.htm
[12] Cubase7 http://www.steinberg.net/en/home.html
[13] SYNTH-WERK http://www.bestservice.de/en/synth-werk.html
[14] BCF2000 http://www.behringer.com/EN/Home.aspx
[15] PureData http://puredata.info/

7. Appendices

The music data and the obtained data can be downloaded from the following URL.
http://istlab.mns.kyutech.ac.jp/~hashi/app.htm