The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Phrase and music search engine for musical data

Kentaro Minowa and Eiji Hayashi
Department of Mechanical Information Science and Technology
Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology
680-4, Kawazu, Iizuka-City, Fukuoka Prefecture, Japan

***Abstract**:* The internet has significantly changed most people's environments. It is possible to search for and obtain information about nearly any subject, and it is possible to purchase many things on the internet. Music files can be purchased, and lyrics can be searched via web services, which is convenient for consumers. However, even if the singer, a part of the musical phrase and the release date are known, a song cannot be accessed if one does not remember the name of the song. If one could enter a sample of the voice, do a search using that voice's length and pitch, find the rhythm and a similarly pitched phrase, it may be possible to determine the name of the song. Therefore, we developed a phrase and music search engine. It searches for a tune from among various tunes, using one phrase of voice data. For the purpose of studying the relationship between tempo and pitch changes related to the frequency of the input voice data, we developed a method using dynamic programming (DP) matching as a way to search for similar phrases. This system converts notes into character strings, runs DP matching using the strings, and calculates the degree of disagreement between these strings. We then use these calculations as an index to determine whether the strings resemble each other.

***Keywords*:** automatic piano, knowledge database, computer music, DP matching

I. INTRODUCTION

The internet has significantly changed most people's environments. It is possible to search for and obtain information about nearly any subject, and it is possible to purchase many things on the internet.

Music purchases are no exception. Previously, CDs were purchased or rented from stores, but now they can be purchased or rented from home. This is convenient for consumers. However, even if the singer, a part of the musical phrase and the release date are known, a song cannot be accessed if one does not remember the name of the song.

On the other hand, related to music performance, we have developed pedal-powered equipment and typing equipment. The problem of fast repetitive typing with the same key and complaints about performance have been solved, and a system in which piano playing takes place automatically has been developed. That system can make use of data when you play piano with MIDI functions and edited data of loudness, the length of the sound, tempo, etc [1]. In addition, studies were conducted two years ago at time of the development of an interactive musical editing system. That system made it possible to perform efficient editing of large music data files. The automatic piano performs the music of the large musical data files. Finding the desired song requires a time-consuming search for the correct musical data. So, we thought it would be useful to have a music information search engine. The system makes an efficient search for the desired song names and musical data, which can reduce the burden on the user.

In this research we developed a system that uses DP matching to search the entire score to look for a match to a given phrase. This system converts notes into character strings and calculates the degree of disagreement between these strings during searching. We use these calculations as an index to determine whether the strings resemble each other.

In this paper, we describe the results of searching for similar phrases using DP matching.

Figure 1.1: View of the automatic piano

II. Musical Editing Support System

2.1 System Architecture

The structure of the system is shown in Figure 2.1. The user edits music via the user's interface on a

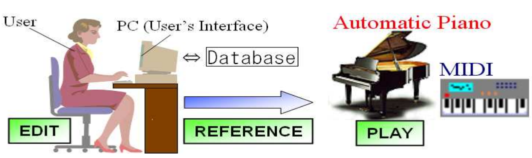


Figure 2.1: Structure of the editing system

computer display. The user can also access a database that has musical grammar, the user's preferences, and so on. As a result, editorial work is reduced and efficient editing becomes possible.

2.2 Performance information

The score is read by the scanner. Scanned data is converted to MusicXML format by the KAWAI Score Maker FX5 program. A database of performance information containing notes, scales, and symbols of music data is created based on the MusicXML data. The created database will be subject to search.

2.2.1 Format of Performance Information

The parameters of performance information are shown in Tables 1 and 2. The automatic piano that we have developed uses a music data structure that is similar to MIDI. We defined performance information, dividing it into two categories: the notes and the pedals. The note information is comprised of the six parameters involved in producing a tone: "Key" (note), "Velo" (velocity), "Gate", "Step", "Bar", and "Time". "Velo" is the dynamics, given by the value of 1–127. "Gate" is the duration of the note in milliseconds. "Step" is the interval of time between notes, and it also exhibits tempo. "Bar" is the vertical line placed on the staff to divide the music into measures.
The pedal information is comprised of four parameters: "Key" (indicating the kind of pedal: "Damper" or "Shifting"), "Velo" (the pedaling quantity), "Time" (the duration for which the pedal is applied)", and "Bar".

Table 2.1: Format of the note information

Parameters	Unit (numerical value)	Resolution	Setting
Key	21~108	Stage 88	Pitch
Velo	1~127	Stage 127	Volume
Gate	ms	1ms	Length
Step	ms	1ms	Interval
Time	ms	1ms	Pronunciation time
Bar	-	-	Bar numbers

Table 2.2: Format of the pedal information

Parameters	Unit (numerical value)	Resolution	Setting
Key	119 or 110	-	119=Damper,110=Shift
Velo	0~127	Stage 128	Position of the pedal
Time	ms	1ms	Pedal operation start time
Bar	-	-	Bar numbers

III. Phrase and music search engine

3.1 System summary

We developed a phrase and music search engine. It searches for one tune from among various tunes using one phrase of voice data. The voice is first inputted; then the interval between the sounds of the inputted voice data is converted back to notes according to the tempo of the music. The system then searches for sequence similarity to the phrase rhythm and notes.

In addition, the pitch of the voice input data was analyzed and pitch changes were calculated. Pitch changes are indicated by the ratio of the pitch of a sound to the pitch of the previous one. By using the change in pitch, we hoped to improve the retrieval accuracy of our system.

3.3 Processing that inputs voice

Voice is input in the shape of waves, as shown in Figure 3.1. The vertical axis of Figure 3.1 shows loudness, and the horizontal axis shows time. By setting a threshold on the vertical axis, the presence of voice can be recognized. The interval of this sound is measured as the time. In other words, as shown in Figure 3.1, the interval of sound is the period between the initial iterative peak and the next sound iterative peak.

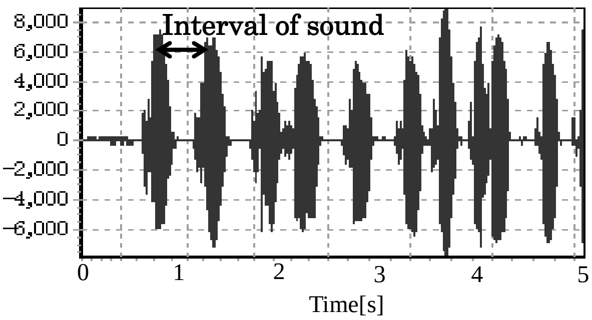


Figure 3.1: Result of sound input

3.4 Converting notes

Measurements of the interval of sound are converted to notes based on the velocity symbol, which appears at the top of the score and represents the number of beats per minute. For example, the indication in Figure 3.2 means 120 quarter note beats per minute. In other words, the length of a quarter note would be 0.5 seconds. The interval of sound is converted into note by using the length of a quarter note.

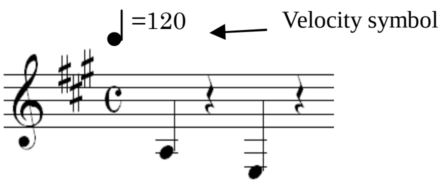


Figure 3.2: Velocity symbol of score

3.5 Frequency analysis of voice data

Pitch is determined by the frequency of the sound, and pitch changes can be determined by the transition of frequency. The voice data frequency analysis is performed using the FFT (fast Fourier transform) algorithm.

Frequency analysis of the voice waveform is conducted for each fixed intervals. And, in one interval, the highest power spectrum frequency is recorded as the main frequency.

IV. Searching for Similar Phrases

As a result of the analysis, it was found that phrases of the same pattern existing in the same tune are performed in a similar expression. In the present study, we used DP matching to search for similar phrases.

4.1 DP matching

DP matching is a technique used widely in the field of speech recognition, bioinformatics and so on. It can calculate the similarity between two words that are different from each other in a number of characters.

In Figure 4.1, the route of minimum cost in each point is taken, and the route with the lowest cost is assumed finally to be the optimal path. The cost at that time is defined as the distance between patterns. In this system, this distance is handled as a threshold to judge whether the phrases are similar to each other.

For example, if the cost moves up or to the right, then it is increased by 1. If it moves to the uppermost level on the right, then it does not increase. Also, if the characters do not correspond in each point, then the cost is increased by 5.

$$g(i,j) = \min\left\{\begin{matrix} g(i-1,j) \\ g(i,j) \\ g(i,j-1) \end{matrix}\right\} + d(i,j)$$

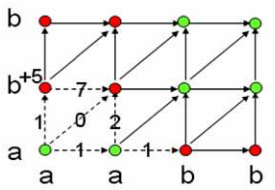


Figure 4.1: DP matching

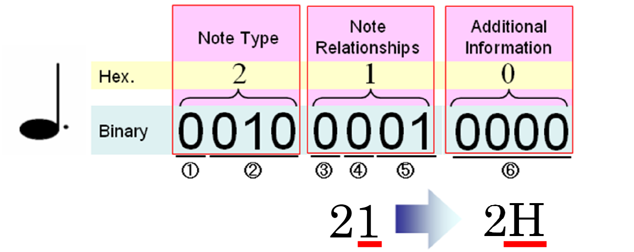
4.2 Searching with DP matching

In this section we describe a method of searching with DP matching. We had to convert a musical score into character strings (a Note Pattern) before searching for similar phrases. This process is explained below.

4.2.1 Patterning

Each note is expressed by three hexadecimal numbers and patterned using only the digits in the first two decimal places, as shown in Figure 4.2 However, the meanings of these two digits vary. Therefore, to avoid confusion, the second digit is distinguished by applying the letter of the alphabet (V from G).

In addition, the sound frequency obtained in Section 3.5 is made into a pattern. First, the system determines the frequency ratio of two successive sounds, and then the ratio of the frequency is patterned by size, as shown in Figure 4.3



① Note or Rest (0: Note, 1: Rest)
② Note Value (000: A whole note, 001: A half note ...etc)
③ Tie (If the note has a tie then this number is 1.)
④ Ornament
(If the note has an ornament then this number is 1.)
⑤ The number of dots
(Exceptionally, if it is "11" then the note is a tuplet.)
⑥ Additional Information

Figure 4.2: Patterning of notes

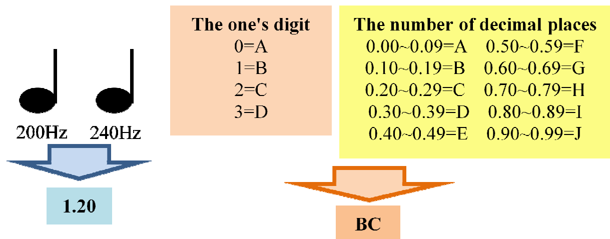


Figure 4.3: Patterning of frequency

4.2.2 The method of searching

A flow chart of a search for a music phrase similar to an input phrase is shown in Figure 4.4.

First, the data of patterned notes column is matched. The phrases with little distance between patterns are searched. Music containing those phrases becomes a search result candidate.

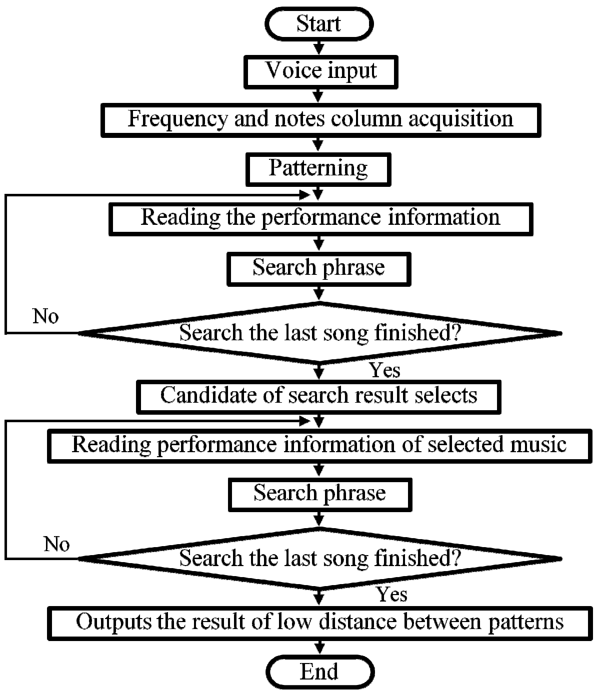


Figure 4.4: The flow of the similar phrase search

Next, the data of patterned sound frequency ratio are matched. Candidates of the search results are narrowed by matching of the frequency ratio.

The search is performed song by song through the database. When the search engine has processed one song, it then proceeds to the next song, continuing to look for the phrase. When the song changes, search engine is change the tempo of the song. And, search phrase is patterned again. After pattern matching in each of two, search result of the lowest distance between patterns is outputted.

V. Retrieval experiment and consideration

We searched for music using the phrase and music search engine. The phrase beginning in bar 59 of "SEKAI NI HITOTU DAKE NO HANA" is sung. The music is searched based on the inputted voice data.

First, the data of the patterned notes column is matched. The system output phrases with a distance of 10 or less. Next, the data of the patterned output phrase frequency ratio is matched. Table 5.1 summarizes the results of the matching. Figure 5.1 represents graphically the difference in the frequency ratio.

A glance at the distance between patterns (Note column) of Table 5.1 will reveal that the phrase beginning in bar 32 of "SEKAI NI HITOTU DAKE NO HANA" and the phrase beginning in bar 59 of the same music are the smallest. However, a search cannot determine the correct one because the two phrases have the same value.

On the other hand, a glance at the distance between patterns (Frequency) of Table 5.1 will reveal that the phrase beginning in bar 32 of "SEKAI" is lower than the phrase beginning in bar 59 of the same music. Thus, in terms of frequency ratios, it is revealed that the phrase begiining in bar 32 is similar to the phrase beginning in bar 59 of "SEKAI NI HITOTU DAKE NO HANA".

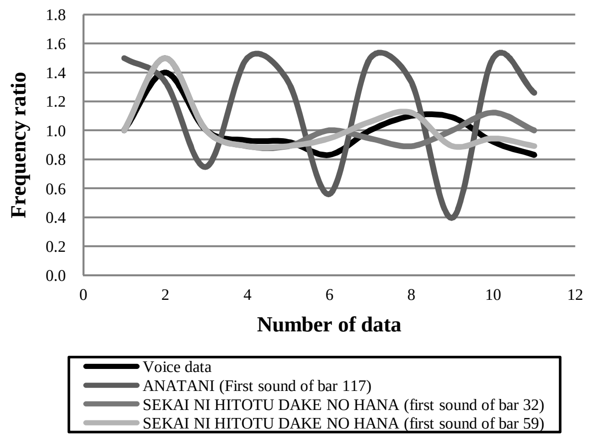


Figure 5.1: Difference in the frequency ratio

Table 5.1: Search result

Name of Music	Bar Number	Note Number	Distance between patterns (Note column)	Distance between patterns (Frequency)
SEKAI NI HITOTU DAKE NO HANA	28	2	10	42
SEKAI NI HITOTU DAKE NO HANA	32	1	5	31
SEKAI NI HITOTU DAKE NO HANA	43	2	10	25
SEKAI NI HITOTU DAKE NO HANA	59	1	5	25
ANATANI	117	1	10	55
ANATANI	118	1	10	55

VI. CONCLUSION

We designed methods of searching for similar phrases using DP matching and combined these functions into a single system.

In the similar phrase search, the system was able to find similar phrases using DP matching in a short time, and it was even possible to find phrases whose resemblance might not be immediately apparent.

In the phrase and music search engine, we developed a system to search for a tune by inputting the data of a voice performing the tune. The interval between the sounds of the inputted voice data is converted to a note according to the tempo of the music. Using this technique, we can search for sequence similarity to a phrase's rhythm and notes. In addition, the search using the patterns of frequency ratio became possible. The search using the patterns of frequency ratio could improve the retrieval accuracy.

In this study, we were able to perform similar phrase searches and searching for a tune by voice. In our future research we will perform evaluations with different pieces of music and will evaluate the existing system.

REFERENCES

[1] Hayashi, E. et al, "Behavior of piano-action in a grand piano.I", Journal of acoustical Society of America, Vol.105, pp.3534-3544, 1999.

[2] Hayashi, E., et al, Interactive musical editing system for supporting human errors and offering personal preferences for an automatic piano, Proc. of the 7th International Symposium on Artificial Life and Robotics, Vol. 2, pp. 513-516, 2002.

[3] Hikisaka, Y., Hayashi, E., et al, Interactive musical editing system for supporting human error and offering personal preferences for an automatic piano –Method of searching for similar phrases with DP matching and inferring performance expression-, Proc. of the 12th International Symposium on Artificial Life and Robotics, GS4-3, 2007.