In-video Searching for Melody in Piano Lesson Videos

Tatsuya Oshiro, Megumi Wakao, Naoki Morita
School of Information Telecommunication Engineering
Tokai University
Tokyo, Japan
e-mail: {9bjt2136@cc, 9bjt2103@cc, wv062303@tsc}.u-tokai.ac.jp

Chiharu Nakanishi, Chiaki Sawada
Faculty of Music Studies
Kunitachi College of Music
Tokyo, Japan
e-mail: {nakanishi.chiharu,sawada.chiaki}@kunitachi.ac.jp

Kazue Kawai
Miyagi University
Miyagi, Japan
e-mail: kawaik@myu.ac.jp

Kenta Morita
Faculty of Medical Engineering
Suzuka University of Medical Science
Mie, Japan
e-mail: morita@suzuka-u.ac.jp

Abstract— **In learning a musical instrument, such as the piano, it is beneficial for students to review their performances on video. However, it is difficult to search through a video for a part of a melody. This is because there is currently no way to search for a specific melody within a single piece of music. We are working towards the development of an in-video searching system for melodies. As a first step, in this study, we propose a method to detect the time when a particular melody is being played from the audio of a student practicing the piano, and test its feasibility.**

Keywords: in-video searching; spectrogram; piano lesson; key melody.

I. INTRODUCTION

Reviewing oneself on video is effective in acquiring skills [1][2][3], and the same principle applies to piano practice. Students can review their performances objectively if they record them on video. In previous research, several learning methods have been proposed for filming lessons, such as systems that can analyze videos to detect bad habits [4] and methods that involve filming from multiple viewpoints [5].

However, it is difficult to search for a specific melody part in these videos. There are currently several ways to search for music. For example, humming searches, such as Google's hum to Search [6] search for metadata such as the song's title and genre based on the hummed melody. Songle [7] can graphically display the structure of a song, such as its chorus or refrain. Although there are various methods for this type of music retrieval, no method has been proposed for searching for parts of melodies contained within a single song.

Against this background, we are working towards the development of an in-video searching system for melodies that detects scenes in which students are practicing a specific melody part in a video showing them practicing the piano.

More specifically, first, students practice music and record their practicing in a video. After that, the same students perform a short melody that they want to review while watching the video and record it as a 'key melody'. Then, by using the system to detect the parts of the video that match the key melody, the student can immediately find and play back the scene in which they are practicing that same melody.

As a first step, this study proposes a method for detecting sounds that match the key melody from the audio of a video.

The structure of this paper is as follows. Section II describes the specific implementation. Section III verifies and evaluates the effectiveness of the proposed method. Section IV presents the conclusions of the paper.

II. METHOD

This section describes an example of a system of in-video searching for melodies by comparing spectrograms.

(1) The system calculates the audio spectrogram of the captured video using a constant-Q [8] transform. This spectrogram will be described in a "salience representation [9]" that takes overtones into account to enhance the sound of harmonic instruments.

(2) The system stores the spectrogram obtained in (1) as an image. The frequency components with energies higher than the threshold value are drawn in white, and the rest are drawn in black. Figure 1 is an image created by the system from the processing steps (1) and (2) for a video recording of a performance of Twinkle Twinkle Little Star. The horizontal axis is time, and the vertical axis is scale.

(3) The system receives a key melody and generates a spectrogram using the same process described in (1) and (2). Figure 2 is an image generated from the first two bars of a performance of Twinkle Twinkle Little Star.

(4) The system overlaps the spectrogram of the video obtained in (2) with the spectrogram of the key melody obtained in (3) and counts the total number of overlapping white dots as the score. We can say that the higher the score is, the higher the similarity is. The

 ISBN: 978-1-68558-057-5

overlapping position is shifted to the right by 1 px from the left end of the spectrogram of the video until the entire recording has been covered. Figure 3 shows an example of how the system calculates the similarity between Figure 1 and Figure 2. Dots that are common to both images are shown in green, those that are only in Figure 1 are shown in white, and those that are only in Figure 2 are shown in red. The scores in the circles are the total number of green dots in the range of Figure 2. A higher number means a higher similarity to the key melody.

III. EXPERIMENT

We evaluate whether multiple videos and key melodies show higher scores at times that include the melody being searched for.

i. Data used in the experiment

Two recordings of piano practice at a music academy are used as the experimental video. In these videos, students practice their set pieces [11][12] repeatedly according to an instructor's comments. In each video, about two bars of a piece are repeatedly practiced.

As the key melody, the same melody as the one practiced in the video, performed by the same student after practice, is used.

ii. Generating spectrogram

Scores are calculated every 10 milliseconds of the video. The spectrograms of the key melodies searched for in video 1 and video 2 had totals of 916 and 2635 white dots, respectively.

iii. Results and Discussion

Figure 4 and Figure 5 show the changes in scores versus time. The horizontal axis is the number of seconds, and the vertical axis is the score. The gray area represents the time when the melody being searched for is actually being played in the video. The red line represents approximately 75% of the maximum score. Most of the scores were significantly higher at the beginning of the gray area. Thus, it was found that the scores were higher at the time when the melody being searched for was actually being played.

When a score exceeding 75% of the maximum was used as the threshold for similarity, it was found that all melodies being searched for could be extracted.

IV. CONCLUSION

We proposed a melody retrieval method using spectrograms as a method to retrieve specific melodies from audio. Experimental results show that a melody being searched for can be successfully identified and extracted when the threshold is set to about 75% of the maximum score.

As this system uses only the sound of the video to find the time when a melody similar to the key melody is being played, we will develop a search engine in combination with a video viewer and recording functions in the future.

ACKNOWLEDGMENT

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 21K18528.

REFERENCES

[1] M. Kagawa, “Investigating the effects of video feedback using digital content on the acquisition and demonstration of motor skills,” Journal of information education, Naruto University of Education, Vol.8, pp. 1-9, March 2011.

[2] K. Yomoda, K. Matsuda, T. Okimura and K. Saito, “The characteristic of students’ reflections from video feedback on high-jump lessons in PE: Content analysis based on specificity, movement phases, and skill levels,” Japanese journal of sports and health science, Vol. 43, pp. 87-101, 2021.

[3] H. Mihara et al., “Educational Practices of Medical Training via Video Learning and Video Assessment,” Medical education, Vol. 52, No. 3, pp. 187-192, June 2021.

[4] R. Matsui, A. Hasegawa, Y. Takegawa, K. Hirata and Y. Yanagisawa, "Design, Implementation and Assessment of a Support System to Find Bad Fingering Habits for Piano Teachers," Transactions of Information Processing Society of Japan, Vol.61, No.4, pp. 789-797, April, 2020.

[5] R. Matsui, Y. Takegawa and K. Hirata, "Tel-Gerich:Remote Piano Lesson System Considering Joint Attention Camera Switching and Camera Switching," The Transactions of Human Interface Society, Vol. 20, No. 3, pp. 321–332, 2018.

[6] Google Inc. *Song stuck in your head? Just hum to search* [Online]. Available from: https://blog.google/products/search/hum-to-search/

[7] M. Goto, K. Yoshii, and T, Nakano, “Songle: active music appreciation service that uses music understanding technology to estimate the content of songs on the web,” 2013-MUS-100 Vol. 16, pp1-9, August 2013.

[8] Judith C. Brown, "Calculation of a constant Q spectral transform, " The Journal of the Acoustical Society of America 89, 425, 1991.

[9] Nicholas Huang, "Auditory salience using natural soundscapes," The Journal of the Acoustical Society of America 141, 2163 ,2017.

[10] Atelier Music School. *Twinkle Twinkle Little Star* [Online]. Available from: https://atelier-music.com/sheetmusic/twinkle-twinkle-little-star

[11] Prokofiev, *Visions fugitives, Op.22*, Collected Works (Собрание сочинений), Vol.1 (pp.133-62), Moscow: Muzgiz, 1955. Plate M. 23404 Г

[12] Beethoven, *Piano Sonata No.15, Op.28*, Ludwig van Beethovens Werke, Serie 16: Sonaten für das Pianoforte (pp.27-44), Nr.138, Leipzig: Breitkopf und Härtel, n.d.[1862-90]. Plate B.138.

 ISBN: 978-1-68558-057-5

Figure 1. Image corresponding to Twinkle Twinkle Little Star

Figure 2. Image corresponding to key melody

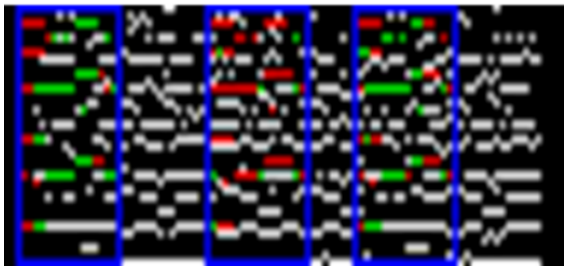

Figure 3. Example of similarity audio of a video and key melody

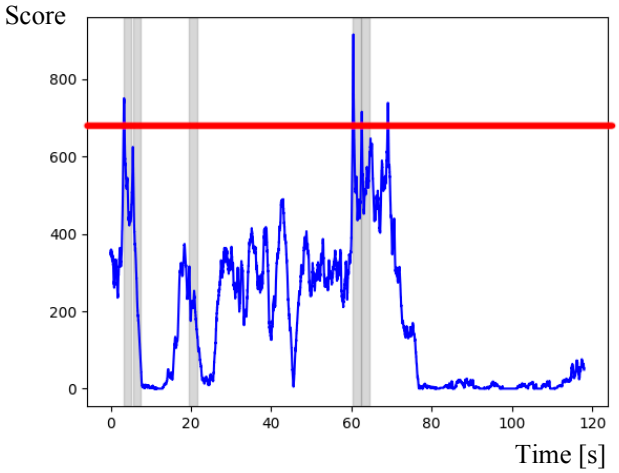

Figure 4. Score versus time of video 1

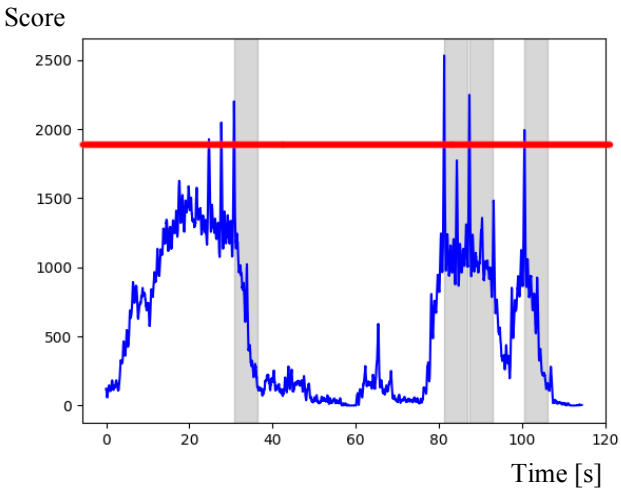

Figure 5. Score versus time of video 2

 ISBN: 978-1-68558-057-5