Query-by-Dancing: A Dance Music Retrieval System Based on Body-Motion Similarity

No Author Given

No Institute Given

Abstract. This paper presents Query-by-Dancing, which is a dance music retrieval system that enables a user to retrieve music by dance motion. When dancers search for music to play when dancing, they sometimes refer to music in online dance videos in which motions similar to their own dance are used. Previous music retrieval systems, however, cannot support retrieval specialized for dancing because they do not accept dance motions as the query. We therefore developed our Query-by-Dancing system that uses a video of a dancing user as the input query to search a database of dance videos. This query video is recorded using an ordinary RGB camera that does not obtain depth information, such as one attached to a smartphone. The pose and motions in the query are then analyzed and used to retrieve a set of dance videos with similar pose and motions. The system then enables the user to browse the musical pieces attached to those retrieved videos so that the user can find a piece that is appropriate for his/her dancing. An interesting problem here is that simply finding the most similar videos results in getting results not similar in terms of genres of dance motions. We solved this by using a novel measure similar to tf-idf to weight the importance of dance motions when retrieving videos. We conducted the comparative experiments on four dance genres and confirmed that the system gained an average of three points or more evaluation scored for three dance genres, Waack, Pop, Break, and that our proposed method was able to deal with different dance genres.

Keywords: dance · music · video · retrieval system · body-motion

1 Introduction

Dancers (human dance performers) often dance along music playback. They choose a dancing style that can match the genre or style of a musical piece, synchronize their movements with musical beats and downbeats, and change their movements to follow musical changes. When musical pieces are given on a dance performance stage, for example, they just have to dance to match the piece being performed. On the other hand, when dancers can select musical pieces for their dance performances, practices, or personal enjoyment, they spend much time finding musical pieces appropriate to their intended dancing. This is because the selection of pieces is important for achieving successful dance performances or enjoying dancing. In searching for music to play when dancing, dancers sometimes refer to music in online dance videos in which motions similar to their own dance are used. They may also refer to music used by their favorite dancers or dance groups in dance events or showcases. Since there has been no systematic support for finding certain kinds of dance music, such activities were time consuming and not easy. Although a lot of music retrieval systems have been proposed [4, 7, 8], none have focused on retrieving dance music.

We therefore propose a dance music retrieval system called Query-by-Dancing[1] that enables a dancer user to use his/her dance motions to retrieve music. Given an input query of a short video capturing dancing body motions, our system can retrieve a set of dance videos that include motions similar to the query. The musical pieces in the soundtracks of those videos are expected to be appropriate for the user's dancing.

Our Query-by-Dancing system does not need an expensive high-performance motion capture system or a camera that obtains depth information. It needs only a simple RGB camera like those attached to smartphones. We implemented the system that analyzes the input query video and a database of dance videos by using the OpenPose library by Cao et al. [1] to estimate body motions.

2 Related Work

A number of music retrieval and recommendation systems have been proposed, but none have allowed a dancer to search for music by dance motion. Our system, Query-by-Dancing, is equipped with a novel function based on the similarity of the dance motion and enables the user to input his/her own dance video as the retrieval query. We surveyed studies on music retrieval and recommendation systems using various queries.

Ghias et al. [3] proposed a query-by-humming system that uses humming as a query. They claim that an effective and natural way of searching a musical audio database is by humming the song. Chen et al. [2] proposed a system for retrieving songs by rhythm from music databases. They use strings of notes as music information, and the database returns all songs containing patterns similar to the query. Jang et al. [6] proposed a query-by-tapping system. The system allows the user to search a music database by tapping on a microphone to input the duration of the first several notes of the query song. Maezawa et al. [9] proposed a query-by-conducting system. In this system, the interface allows a user to conduct during the playback of a piece, and the interface dynamically switches the playback to a musical piece that is similar to the users conducting.

Some systems retrieve musical pieces by using a music context, such as the artists cultural or political background, collaborative semantic labels, and album cover artwork [11]. Turnbull et al. [12] presented a query-by-text system that given a text-based query can retrieve relevant tracks from a database of unlabeled audio content. This system can also annotate novel audio tracks with meaningful words.

[1] An anonymized demonstration video is available at `https://youtu.be/lWZe-X3VjiY`.

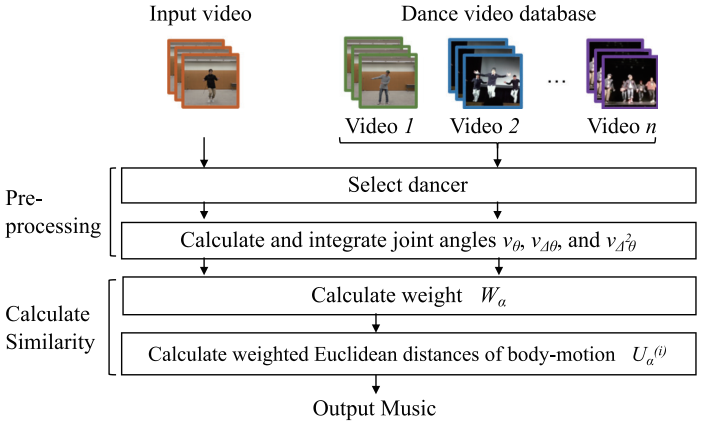
Input video
Dance video database
Video 1
Video 2
…
Video n
Pre-processing
Select dancer
Calculate and integrate joint angles v_θ, $v_{\Delta\theta}$, and $v_{\Delta^2\theta}$
Calculate Similarity
Calculate weight W_α
Calculate weighted Euclidean distances of body-motion $U_\alpha^{(i)}$
Output Music

Fig. 1. System overview.

As described above, several retrieval methods using various queries have been proposed. To the best of our knowledge, however, this is the first study that takes dance motions as a query for music retrieval. Our system focuses on dance motion and acquires candidate musical piece candidates from a dance video database.

3 Dance Music Retrieval System

The system overview is shown in Fig.1. Our system can be divided into two stages, pre-processing stage and similarity calculation stage. These two main stages are described below.

3.1 Pre-processing

Detect a Dancer In this step, the system first estimates the person's skeleton information in all frames using the OpenPose library [1]. Some dance videos have images of multiple people dancing, or the OpenPose library detects the skeleton information incorrectly on a spot that has no person. Therefore the system must select which skeleton information represents a dancer among the detected skeleton information per frame. The difference between the maximum(x_{max}) and minimum(x_{min}) values in the x-axis direction of the obtained skeletal information is defined as the area occupied(A_{o}) by the dancer, which is obtained by multiplying the difference between the maximum(y_{max}) and minimum(y_{min}) values in the y-axis direction. The distance from center($P_{\mathrm{c}}(X_{\mathrm{mean}}, Y_{\mathrm{mean}})$) of the image to the average position of all skeleton information of dancer(P_{d}) is D_{c}, and the dancer is represented by the skeleton information that maximizes $R = \frac{A_{\mathrm{o}}}{D_{\mathrm{c}}}$.

Feature extraction To calculate the dance motion similarity between dance videos, we extract the following feature quantity per frame. We determined that

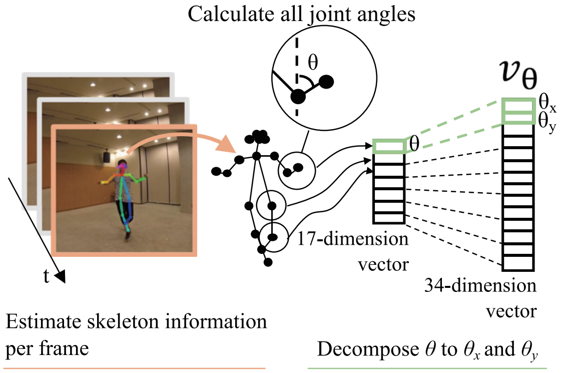
Calculate all joint angles
θ
v_θ
θ_x
θ_y
θ
t
17-dimension vector
34-dimension vector
Estimate skeleton information per frame
Decompose θ to θ_x and θ_y

Fig. 2. All the joint angles are decomposed per frame, and a 34-dimensional vector is created.

the pose and motions are important elements that characterize dancing. Therefore, to consider the pose the system calculates 17 joint angles per frame from the estimated skeleton information. The joint angles are decomposed into two dimensions of θ_x and θ_y by calculating sine and cosine as shown in Fig. 2, and we denote a 34-dimensional feature vector of angles at n-th frame of i-th video by $v_\theta^{(i)}(n)(1 \leq n \leq N^{(i)})(1 \leq i \leq I)$ where N is the number of frames in the i-th video and I is the number of videos in the database. Furthermore, the angle where the bone was not detected is expressed as zero.

To consider the speed of the body-motion, the system focuses on the change in joint angles between frames. It calculates $v_{\Delta\theta}^{(i)}(n)$ and $v_{\Delta^2\theta}^{(i)}(n)$ as follows:

$$v_{\Delta\theta}^{(i)}(n) = \mathrm{abs}(v_\theta^{(i)}(n) - v_\theta^{(i)}(n-1)) \tag{1}$$

$$v_{\Delta^2\theta}^{(i)}(n) = \mathrm{abs}(v_{\Delta\theta}^{(i)}(n) - v_{\Delta\theta}^{(i)}(n-1)) \tag{2}$$

where $abs(x)$ denotes a vector containing the absolute value of each element of x. We concatenate the above three feature vectors, $v_\theta^{(i)}(n)$, $v_{\Delta\theta}^{(i)}(n)$, $v_{\Delta^2\theta}^{(i)}(n)$, into one 102-dimensional vector $v_\alpha^{(i)}(n)$.

3.2 Similarity Calculation

The system calculates the Euclidean distances $d(v_\alpha^{\mathrm{in}}(n), v_\alpha^{(i)}(m))$ between all frames$(1 \leq n \leq N^{\mathrm{in}})$ of an input video (in) and all frames$(1 \leq m \leq N^{(i)})$ of a video in the video database $(1 \leq i \leq I)$, where $d(x, y)$ denotes an Euclidean distance between x and y, as shown in Fig.3. The system computes these Euclidean distances in all frame combinations and divides them by the total number of combinations $(N^{\mathrm{in}}N^{(i)})$. They are denoted by following formula:

$$R_\alpha^{(i)} = \frac{1}{N^{\mathrm{in}}N^{(i)}} \sum^{N^{\mathrm{in}}} \sum^{N^{(i)}} d(v_\alpha^{\mathrm{in}}(n), v_\alpha^{(i)}(m)). \tag{3}$$

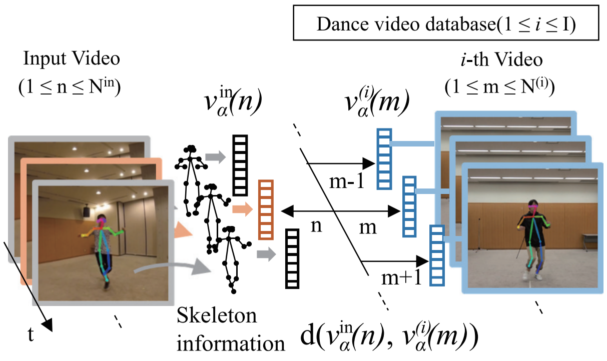
Dance video database(1 ≤ i ≤ I)
Input Video
(1 ≤ n ≤ N^in)
$v_\alpha^{\text{in}}(n)$
$v_\alpha^{(i)}(m)$
i-th Video
(1 ≤ m ≤ N^(i))
m-1
n
m
m+1
t
Skeleton
information
$\mathrm{d}(v_\alpha^{\text{in}}(n), v_\alpha^{(i)}(m))$

Fig. 3. The system computes the Euclidean distances in all frame combinations and divides them by the total number of combinations.

Now the system simply finds the most similar videos by using $R_\alpha^{(i)}$ as the similarity of dance motion per video, which leads to getting results not similar in terms of genres of dance motions. We solve this problem by using a novel measure similar to tf-idf to weight the importance of dance motions when retrieving videos. The weight representing the importance of dance motions is calculated as follows:

$$W_\alpha(n) = \frac{\frac{1}{N^{(i)}} \sum^{N^{(i)}} d(v_\alpha^{\text{in}}(n), v_\alpha^{(i)}(m))}{\max\limits_{i \in I}\{\frac{1}{N^{(i)}} \sum^{N^{(i)}} d(v_\alpha^{\text{in}}(n), v_\alpha^{(i)}(m)\}}. \tag{4}$$

where $max(x)$ denotes a maximum value of x. To sharpen weight gradient, the system calculates $W_\alpha(n)$ to the 30-th power to obtain $W'_\alpha(n)$. The exponent 30 was determined experimentally. Then the system multiplies $W'_\alpha(n)$ by all the Euclidean distances. These distances are given by following formula:

$$U_\alpha^{(i)} = \frac{\sum^{N^{\text{in}}} (W'_\alpha(n) \sum^{N^{(i)}} d(v_\alpha^{\text{in}}(n), v_\alpha^{(i)}(m)))}{N^{\text{in}} N^{(i)}}. \tag{5}$$

As a set of videos containing similar dance motions to those of the input video, the system finds the set of videos with the top k values among $U_\alpha^{(i)}$. Finally, the system presents the candidate musical pieces from the searched dance videos.

4 Evaluation

We conducted two evaluation experiments to investigate whether the retrieval results are easy for dancers to use as dance music. In the first experiment, the system retrieved dance music depending on whether the importance of each video is weighted or not. We adapted the calculated weights to the two methods respectively and retrieved dance music using each method. In the second experiment, the system retrieved dance music by taking the dance videos of four dance genres as the queries. This retrieval was done using the method that gave the best results in the first experiment.

4.1 Dance Video Database

We used 100 dance videos available on YouTube [13] and Instagram [5]. They were 25 videos of each of four dance genres Break, Hip-hop, Waack, and Pop and their average duration was 82 sec. The soundtracks of all these videos contained music, and the dancers in the videos danced to that music.

4.2 Experiment I: weighted or unweighted

Experiment conditions We recruited 12 participants (four males and eight females) who were students belonging to a street dance club. All had between 1 and 15 years of dance experience (average = 8.5 years).

We compared four retrieval methods: ADD (unweighted), ADD (weighted), DTW (unweighted) and DTW (weighted). The method here called ADD, our proposed retrieval method using the Euclidean distance between frames, calculates one feature vector $v_{\alpha}^{(i)}(n)$ of 102 dimensions by concatenating $v_{\theta}^{(i)}(n)$, $v_{\varDelta\theta}^{(i)}(n)$, and $v_{\varDelta^2\theta}^{(i)}(n)$ per frame. ADD (unweighted) does not use $W'_{\alpha}(n)$ and extracts musical pieces in ascending order of $R_{\alpha}^{(i)}$. ADD (weighted) uses $W'_{\alpha}(n)$ and extracts musical pieces in ascending order of $U_{\alpha}^{(i)}$.

The method here called DTW as a retrieval method using dynamic time warping, which is a sequence matching algorithm that considers longer-term similarity. In this method, while sliding $v_{\theta}^{(i)}(n)$ one frame at a time, we create a sequence $V_{\mathrm{dtw}}^{(i)}(n)(1 \leq n \leq N^{(i)} - 5)(1 \leq i \leq I)$ for every six frames. The system calculates the dynamic time warping $dtw(v_{\mathrm{dtw}}^{\mathrm{in}}(n), V_{\mathrm{dtw}}^{(i)}(m))$ between all sequences($1 \leq n \leq N^{\mathrm{in}} - 5$) of an input video(in) and all sequences($1 \leq m \leq N^{(i)} - 5$) of a video in the video database ($1 \leq i \leq I$), where $dtw(x, y)$ is the Euclidean distance between x and y calculated by FastDTW [10]. Then $R_{\mathrm{dtw}}^{(i)}$ and $W_{\mathrm{dtw}}(n)$ are calculated using an equation in which the $d(v_{\alpha}^{\mathrm{in}}(n), v_{\alpha}^{(i)}(m))$ in equation (3) are replaced with $dtw(v_{\mathrm{dtw}}^{\mathrm{in}}(n), V_{\mathrm{dtw}}^{(i)}(m))$. To sharpen weight gradient, the system calculates $W_{\mathrm{dtw}}(n)$ to the 40-th power to obtain $W'_{\mathrm{dtw}}(n)$. The exponent 40 was determined experimentally. Then the system multiplies $W'_{\mathrm{dtw}}(n)$ by all the Euclidean distances and obtains $U_{\mathrm{dtw}}^{(i)}$. DTW (unweighted) does not use $W'_{\mathrm{dtw}}(n)$ and extracts musical pieces in ascending order of $R_{\mathrm{dtw}}^{(i)}$. DTW (weighted) uses $W'_{\mathrm{dtw}}(n)$ and extracts musical pieces in ascending order of $U_{\mathrm{dtw}}^{(i)}$.

We asked a Waack dancer who had 15 years of dance experience to participate in the experiment of and shot about 11 seconds of her Waack dancing. Using that video as a query, we used each of the four methods to retrieve musical pieces. We denoted the top five music groups in the retrieval results obtained with each of the four methods as MG-A, MG-B, MG-C, and MG-D. Each music group had five musical pieces.

Procedure At the beginning of the session, participants filled out a pre-study questionnaire on their dance experience. We then gave them a brief explanation of the experiment. Then after the watched the Waack dancers 11-second video, which did not contain music and was used as a query, they were asked to listen to the five musical pieces in each music group and evaluate them on a 5-point Likert scale ranging from 1 for not agree to 5 for totally agree. They were given the music groups MG-A, MG-B, MG-C, and MG-D in random order. We gave them the evaluation item below.

Q1: When it is supposed that "someone" dances according to the music with the dance on this video, considering the atmosphere of the dance and the atmosphere of the music of these five musical pieces, this musical piece is easy to dance to.

Finally, they filled out a questionnaire about the dance music retrieval.

We prepared a MacBook Pro (Retina display, 15-inch, Mid 2015) and used QuickTime Player to play the musical pieces and the video. The dance video that was the query was set to "repeat play" beforehand and the participant selected and played the five musical pieces arranged next to it. The participants wore earphones to listen to the music and could play the musical pieces any number of times and re-evaluate them. During the experiment the participants could take breaks freely. The experiment took about 40 minutes.

Results and Discussion The averaged Q1 scores for each retrieval method are shown in Fig. 4. The vertical axis indicates the average of Q1 scores given by all participants, and the vertical bars indicate standard errors. The horizontal axis represents retrieval methods. The gray rectangles show the averaged evaluation scores for each of the retrieval ranks. Each green rectangle shows the average of all evaluation scores within the retrieval method. We assessed the difference between the average Q1 scores by using ANOVA. There was a significant difference ($F_{(3,236)} = 4.21, p < .05$). We also assessed difference by using Fishers Least Significant Difference (LSD) test and found significant differences ($p < .05$) between ADD (weighted) and the other three methods. Thus the ADD (weighted) method was a retrieval method suitable for searching for musical pieces that dancers can easily dance to.

The retrieval results of ADD (weighted) had the same dance genres as the query more often than other retrieval methods, which increased the evaluation score of ADD (weighted). The dance genres of each retrieval rank for each retrieval method are shown in Tab. 1. Focusing on the top five musical pieces in retrieval results, we find that four out of five musical pieces for ADD (weighted) were Waack, which was as the same dance genre as the query. For the other methods, two out of five, one out of five, and three out of five musical pieces were Waack. The musical pieces that videos in the same dance genre as the query contain got higher evaluation scores.

Next, we focused on the weights. Figure 4 showed that the scores of the weighted methods were higher than those of the unweighted methods and that weighting is effective for retrieving dance music appropriate to dance motion. The calculated weight $W'_\alpha(n)$ is shown in Fig. 5, where the vertical axis indicates the

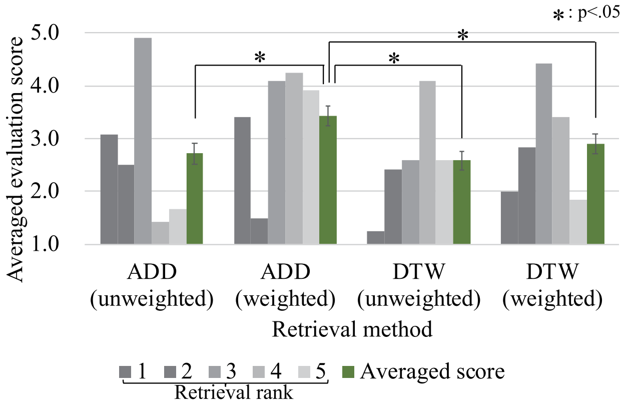

Fig. 4. The averaged Q1 scores for each retrieval method. The ADD (weighted) method was evaluated significantly more highly than the other three methods.

Table 1. Top 5 retrieval results by retrieval methods. P in the table stands for the dance genre Pop, and W stands for the dance genre Waack.

Retrieval method	Dance genre	Retrieval rank				
		1	2	3	4	5
ADD (unweighted)	Waack	W	P	W	P	P
ADD (weighted)		W	P	W	W	W
DTW (unweighted)		P	P	W	P	P
DTW (weighted)		P	W	W	W	P

weight value and the horizontal axis represents the frame numbers in the video used as the query. The movements in the vicinity of frames 240 to 270 with a high weight value were movements like the dancer swinging her arm above her head in long strides. Moreover, the movements in the vicinity of the first 50 frames with a relatively high weight value were movements like the dancer swinging her arm to the left and right in long strokes. Swinging the arm in long strokes, which was Waack's characteristic movement, had been highly weighted. On the other hand, the movements in the vicinity of frames 50 to 75 with a relatively low weight value were movements like simply going backward. Moreover, the movements in the vicinity of frames 200 to 225 with a relatively low weight value were movements like the dancer shaking her waist to the left and right. These movements also occur in various other dance genres. As the above shows, the system successfully weighted the movement peculiar to the dance motion in the query. In contrast, it weighted movement common to other dance genres low.

4.3 Experiment II: Retrieval Performance

Experiment conditions We recruited 12 participants (six male and six female) who were students belonging to a street dance club. All had between 1 and 15 years of dance experience (average = 5.9 years). We compared four dance genres:

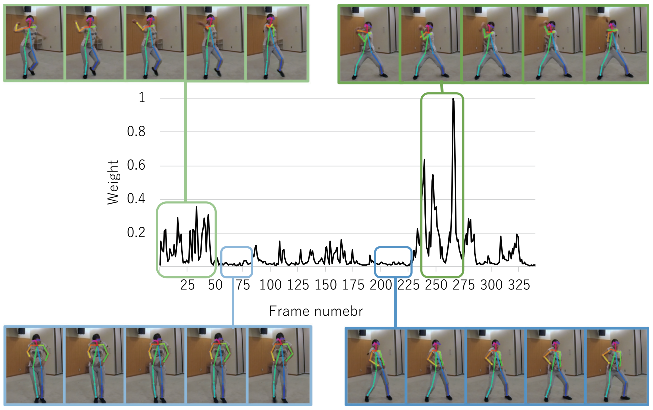

Fig. 5. Waack's characteristic movements were in the vicinity of a relatively high weight value. The movements common to other dance genres were in the vicinity of a relatively low weight value.

Waack, Break, Pop, and Hip-hop. We prepared the video of Waack used in the first experiment. One of the author himself who had 8 years of dance experience was in charge of a Break dancer, and we shot about 13 seconds of that authors Break dance. To prepare other videos, we recruited two more dancers, a Pop dancer and a Hip-hop dancer. The Pop dancer had 3 years of dance experience and we shot about 16 seconds of his Pop dance. The Hip-hop dancer had 15 years of dance experience and we shot about 16 seconds of her Hip-hop dance. Using those videos as queries, we retrieved musical pieces by using the ADD (weighted) method. We denoted the top five music groups in the retrieval results obtained in each of the dance genres as DG-W, DG-B, DG-H, and DG-P. Each music group had five musical pieces.

Procedure At the beginning of the session, participants filled out a pre-study questionnaire on their dance experience. We then gave them a brief explanation of the experiment. Then after the watched one of the randomly selected dance videos, which did not contain music and was used as a query, they were asked to listen to the five musical pieces in each music group and evaluate them on a 5-point Likert scale ranging from 1 for not agree to 5 for totally agree. They were given the music groups DG-W, DG-B, DG-H, and DG-P according to the dance genre of the video. We gave Q1 as the evaluation item. Finally, they were orally interviewed.

We prepared a MacBook Pro (Retina display, 15-inch, Mid 2015) and used QuickTime Player to play the musical pieces and the video. The dance video that was the query was set to "repeat play" beforehand and the participant selected and played the five musical pieces arranged next to it. The participants wore earphones to listen to the music and could play the musical pieces any number of times and re-evaluated them. During the experiment the participants could

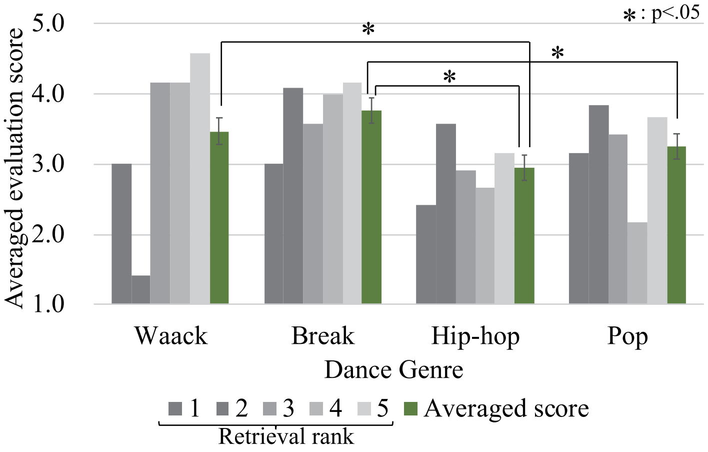

Fig. 6. The averaged Q1 scores for each dance genre.

Table 2. Top 5 retrieval results by dance genre. P in the table stands for the dance genre Pop, W for the dance genre Waack, and B for the dance genre Break.

Retrieval method	Dance genre	Retrieval rank				
		1	2	3	4	5
ADD (weighted)	Waack	W	P	W	W	W
	Break	B	B	B	B	B
	Hip-hop	B	B	B	B	B
	Pop	P	P	P	P	P

take breaks freely. The experiment took about 40 minutes. The participants wore earphones to listen to the music, and they could play the musical pieces any number of times and re-evaluate them. During the experiment the participants could take breaks freely. The experiment took about 40 minutes.

Results and Discussion The averaged Q1 scores for each dance genre are shown in Fig. 6. The vertical axis indicates the average of Q1 scores given by all participants, and the vertical bars indicate standard errors. The horizontal axis represents dance genres. The gray rectangles indicate the averaged evaluation scores for each of the retrieval ranks. Each green rectangle indicates the average of all evaluation scores within the genre. We assessed the difference between the average Q1 scores using ANOVA. There was a significant difference ($F_{(3,236)} = 3.92, p < .05$). We also assessed the difference by using Fishers Least Significant Difference (LSD) test and found significant differences ($p < .05$) between Waack and Hip-hop, Break and Hip-hop, and Break and Pop.

The Q1 evaluation scores and dance genres of each retrieval rank for each retrieval method are shown in Tab. 2. Using the Hip-hop video as the query resulted in comparatively bad performance: the system returned musical pieces in the Break genre. There are two reasons for this. One is that Hip-hop is more finely divided into many dance genres. The style of the Hip-hop in our input video was Middle Hip-hop, but the Hip-hop dance styles on the database were Style Hip-hop, Girls Hip-hop, Jazz Hip-hop and so on; dance styles slightly different from that of the query. Therefore it was hard for the system to extract the musical pieces of Hip-hop. The other reason is that Middle Hip-hop has some movements similar to those of Break. One characteristic of Break is dancing while keeping your hands on the floor. Before holding their hands on the floor, however, Break dancers dance with movement similar to Middle Hip-hop dance. Therefore the system selected the musical pieces of Break. On the other hand, the system could extract motion similar to the query, which prevented the evaluation scores from markedly decreasing. For these two reasons, the system extracted the musical pieces of Break which were not the same dance genre of the query. In the future, we divide the dance genres into more finely and add various dance genres into the database, which will improve the evaluation score.

Next, focusing on Pop, one sees that the evaluation score of Pop was worse than that of Break and tended to be worse than that of Waack. To find the reason for this, we interviewed the participants who gave a low score to Pop. They said they did so be cause the dance motion used as the query included a “vibration” technique in which the dancers have their bodies move with a rapid trembling motion and a “wave” technique in which the dancers have their bodies move like a wave. Those movements match specific sounds, and the participants judged that musical pieces not including those sounds were not appropriate to those movements. We can solve this problem by using interactive retrieval methods that let dancers adjust parameters according to their purposes. For example, if dancers want to search for the specific musical pieces used in videos containing movements that closely resemble movements like “waves”, the system allows the dancers to search for a narrow range of musical pieces by adjusting parameters to match movements that have a high similarity. Additionally, if users want to search for musical pieces for practicing dance or for dancing in a place like a dance club where many dancers gather, the system allows the users to search a variety of musical pieces in a large range by adjusting parameters to match the movements that have a lower similarity. Users changing the parameters according to the situations could realize a dance music retrieval more suitable for the purpose.

5 Conclusion

We proposed Query-by-Dancing, which is a dance music retrieval system that enables a user to retrieve a musical piece by dancing. We confirmed that the system’s retrieval method is appropriate for dance music, the system can find a musical piece that dancers can dance to easily, and music better for the dancer can be obtained by weighting the importance of dance motions when retrieving videos. Moreover, we conducted the comparative experiments on four dance genres and confirmed that the system gained an average of three points or more evaluation scored for three dance genres, Waack, Pop, Break, and then our proposed method could adapt to different dance genres. We confirmed that that genre is divided more finely even in the same dance genre as the query. It is necessary to add a wide range of dance genres to the database of dance videos. In the future, we plan to propose a more interactive system that allows the users to adjust the parameters of retrieval accuracy.

References

1. Cao, Z., Simon, T., Wei, S., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: the 2017 IEEE conference on Computer Vision and Pattern Recognition (2017)
2. Chen, J., Chen, A.: Query by rhythm: an approach for song retrieval in music databases. In: Proceedings of the 8th International Workshop on Research Issues in Data Engineering: Continuous-Media Databases and Applications. pp. 139–146 (1998)
3. Ghias, A., Logan, J., Chamberlin, D., Smith, B.C.: Query by humming - musical information retrieval in an audio database. In: Proceedings of ACM Multimedia95. pp. 231–236 (1995)
4. Hassan-Montero, Y., Herrero-Solana, V.: Improving tag-clouds as visual information retrieval interfaces. In: International conference on multidisciplinary information sciences and technologies. pp. 25–28 (2006)
5. Instagram: Instagram, `https://www.instagram.com/` (accessed July 30, 2018)
6. Jang, J., Lee, H.R., Yeh, C.H.: Query by tapping: A new paradigm for content-based music retrieval from acoustic input. In: Pacific-Rim Conference on Multimedia. pp. 590–597 (2001)
7. Logan, B., Salomon, A.: A music similarity function based on signal analysis. In: the 2001 IEEE International Conference on Multimedia and Expo. pp. 22–25 (2001)
8. Ltd., S.E.: Shazam, `https://www.shazam.com/` (accessed July 30, 2018)
9. Maezawa, A., Goto, M., Okuno, H.G.: Query-by-conducting: An interface to retrieve classical-music interpretations by real-time tempo input. In: the 11th International Society of Music Information Retrieval. pp. 477–482 (2010)
10. Salvador, S., Chan, P.: Toward accurate dynamic time warping in linear time and space. Journal of Intelligent Data Analysis **11**(5), 561–580 (2007)
11. Smiraglia, R.P.: Musical works as information retrieval entities: Epistemological perspectives. In: the 2nd International Society of Music Information Retrieval. pp. 85–91 (2001)
12. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.: Semantic annotation and retrieval of music and sound effects. IEEE Transactions on Audio, Speech, and Language Processing **16**(2), 467–476 (2008)
13. YouTube: Youtube, `https://www.youtube.com/` (accessed July 30, 2018)