International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 4874-4884

https://doi.org/10.70517/ijhsa463417

Modeling and Analysis of Music and Dance Rhythm Matching Using Bayesian Inference Models

Peipei Liu[1,*]
[1] School of Art and Design (Center for Public Art Education), Henan University of Urban Construction, Pingdingshan, Henan, 467002, China
Corresponding authors: (e-mail: hnlpp812@163.com).

Abstract This study proposes a multilevel modeling framework that integrates Bayesian inference, genetic algorithm optimization and dynamic temporal regularization DTW, aiming at high-precision collaborative matching of music and dance rhythms. By constructing a dynamic bar-pointer model, the bar line position and tempo of the music rhythm are taken as hidden state variables, and a posteriori density estimation is combined with a sequential Monte Carlo method to realize robust music rhythm extraction. For the dance movement system, a feature optimization framework based on genetic algorithm is proposed to filter the optimal music-dance movement matching combinations through the fitness function and quantify the rhythm synchronization by combining with the DTW algorithm. Experimental validation shows that the beat tracking algorithm based on Bayesian theory performs well in music cycle extraction, with cycle peaks stabilized in the interval from -0.8 to 1, with beat division. The correlation coefficient between music and dance movement features reaches 0.827, and the matching accuracy reaches 84.52% at 20 feature pairs. The beat points of the synthesized dance highly overlapped with the music, and the intensity distribution trend was consistent. This study not only provides a quantifiable analysis tool for music-dance co-creation, but also lays a theoretical foundation for cross-modal interaction technology in virtual reality and intelligent choreography.

Index Terms bayesian inference, dynamic time regularization, genetic algorithm, music-dance rhythm matching

I. Introduction

Dance is an art which expresses emotions and thoughts through body movements and postures [1], [2]. And in dance performance, the matching of dance rhythm and music plays a vital role [3]. They depend on each other and influence each other to create a wonderful dance picture [4].

Rhythm is the soul of dance performance, the basis of dance movement, and an important part of dance form [5], [6]. Rhythm can be embodied by various means, such as the speed of movement, the length of pause, the strength of the notes, etc. In dance performance, the change of rhythm can make the dance more vivid and powerful, and can also make the dance softer and more elegant [7]-[10]. And music is the source of inspiration for dance performance, which can stimulate the dancer's emotion and guide the dancer into the role, and the rhythm, melody and mood of the music will affect the effect of dance performance [11]-[14]. Dancers need to understand and feel the music to transform it into the power of dance movements, and music can help dancers better express their emotions and show their inner world [15-17]. It can add emotional layers and depth to the dance, making the audience more easily moved by the dance.

In dance performance, the matching of rhythm and music requires dancers to have a good sense of music and rhythm [18], [19]. Dancers need to accurately grasp the rhythm of music and translate it into body movements, and they need to develop their sensitivity and understanding of music through practice and training [20], [21]. Only in this way can they blend perfectly with the music on the stage and convey the charm of the music to the audience through dance [22]. The matching of rhythm and music in dance performance also requires tacit understanding and cooperation between dancers and musicians [23]. Dancers and musicians need to understand each other, match each other, and work together to create the best dance effect, and they need to make repeated grinding and adjustments in rehearsals to ensure that the dance movements fit perfectly with the rhythm of the music [24]-[26]. Only in this way can they present amazing dance images on stage.

This study proposes a multilevel modeling framework that integrates Bayesian inference, genetic algorithm optimization and dynamic time regularization (DTW), aiming to achieve high-precision collaborative analysis and matching optimization of music and dance rhythms. Through Bayesian inference and dynamic bar-pointer modeling, the dynamic changes of music rhythm (e.g., the position and speed of bar line pointer) are taken as hidden state variables, and the a posteriori density estimation problem is solved by combining with sequential Monte Carlo method to realize the robust extraction of music rhythm. The principle model of the dance technical movement system is also constructed, and a genetic algorithm is introduced to optimize the music-dance movement correspondence based on the correlation analysis of music and dance movement features. Taking the historical training data as the benchmark, the optimal matching combinations are screened by the fitness function, and the principle model of the dance technical movement system is constructed. On this basis, the DTW algorithm is introduced to measure the degree of matching between music and dance rhythm. For the problem of non-alignment of movement and music feature sequences on the time axis, the DTW algorithm calculates the cumulative distance function, combines the Euclidean distance with the time regularization weights to quantitatively assess the degree of rhythmic matching, overcomes the problem of lack of synchronization caused by the time axis offset, and improves the accuracy of synchronization analysis.

II. Music-dance rhythm collaborative modeling based on Bayesian inference and dynamic time regularization

II. A. Music rhythm extraction based on Bayesian theory

II. A. 1) Bar-pointer model of musical rhythmic dynamics

In the dynamic bar-pointer model, bar refers to the barline and bar-pointer denotes the barline pointer. Note articulation points, changes in rhythmic patterns, have the highest percentage of occurrences near the bar line, and the bar-pointer maps potential rhythms if the note articulation points are taken as observables and the position and velocity of the bar-pointer are taken as hidden state variables.

Let the moment $t \in k\Delta$, $k \in \{1, 2, \cdots, n\}$ (Δ is a constant), and the position variable of the bar-pointer is ϕ_k and the velocity variable is $\dot{\phi}_k$, then the dynamic model of the bar-pointer can be modeled as:

$$\begin{cases} \varphi_{k+1} = (\varphi_k + \Delta\dot{\varphi}_k) \bmod 1 \\ p(\dot{\varphi}_{k+1} \mid \dot{\varphi}_k) \propto N(\dot{\varphi}_k, \sigma_\varphi^2) \end{cases} \tag{1}$$

The rhythmic pattern indication is denoted by κ , which is a finite set of patterns, and assuming that there are only two rhythmic patterns, its positional relationship to the bar-pointer can be expressed as follows

$$\begin{aligned} \phi_k < \phi_{k-1}, p(\kappa_k \mid \kappa_{k-1}, \phi_k, \phi_{k-1}) \\ = \begin{cases} p_s, & \kappa_k \neq \kappa_{k-1} \\ 1 - p_s, & \kappa_k = \kappa_{k-1} \end{cases} \end{aligned} \tag{2}$$

$$\phi_k \geq \phi_{k-1}, \kappa_k = \kappa_{k-1} \tag{3}$$

II. A. 2) Bayesian inference

Bayesian estimation is one of the most commonly used and effective methods in dealing with various stochastic state and parameter estimation problems. One application of Bayesian estimation is to estimate the mean and variance of x_n given $y_{1:n}$ in a recursive manner. When x obeys the distribution $p(x)$, in order to find the expected value of any function $g(x)$ of x , one needs to compute an integral of the form

$$E[g(x)] = \int g(x) p(x) dx \tag{4}$$

In many cases, the integral of the above equation cannot be obtained by direct computation and the integral must be approximated.

In this paper, ϕ_k , $\dot{\phi}_k$, and κ_k are composed of the vector $x_k = [\phi_k, \dot{\phi}_k, \kappa_k]^T$, and the observation sequence is y_k , which, under recursive Bayesian estimation, means that the a posteriori density, $p(x_n \mid y_{1:n})$, is to be obtained quantitatively. From the Bayesian formulation, there is the posterior density

$$p(x_n \mid y_{1:n}) = cp(y_n \mid x_n) p(x_n \mid y_{1:n-1}) \tag{5}$$

where c is a normalization constant.

In many cases, $p(x_n \mid y_{1:n})$ is multivariate, or non-standard (e.g., no analytic expression), or multi-peaked, and it is difficult to obtain $p(x_n \mid y_{1:n})$. One solution is to use Bayesian importance sampling methods to generate samples that obey an importance sampling function, and then compute the weighted value of the importance of these samples in the desired distribution. This method is the sequential Monte Carlo method. From the sequential Monte Carlo method, the posterior density probability function is approximated as:

$$p(x_n \mid y_{1:n}) = \sum_{i=1}^{N_s} \omega_n^{(i)} \delta(x_n - x_n^{(i)}) \tag{6}$$

where $\omega_n^{(i)}$ is the weight of each particle $x_{i:n}$ in the recursion.

Although the above importance sampling method yields an approximation of the a posteriori density function, the weights of the vast majority of particles are negligibly small after several iterations. A way to minimize the effect of particle degradation is to resample when significant particle degradation occurs. That is, eliminate those particles with smaller weights during the recursion process, while concentrating repeated sampling on particles with larger weights.

According to the above Bayesian estimation method, the estimation steps can be described as follows:

(1) Initialize the particle distribution;

(2) Important density sampling to calculate the particle weights;

(3) Discard particles with small weights and resample particles with large weights;

(4) Select a Bayesian estimation criterion to estimate the mean and variance of the particles.

II. B.Principle Model of Dance Technique Movement System

After completing the dynamic modeling of music rhythm through Bayesian inference, the principle model of the dance movement system needs to be further established to match the music features. To this end, this section proposes an optimization framework based on genetic algorithm to construct the optimal matching relationship of multimodal feature pairs by analyzing the correlation between music and dance movement features.

In building the principle model of the dance technique movement system, the collected music data are firstly sliced and processed, and the movement segments in the dance movement database are connected and organized to obtain the underlying features of historical music and dance movements, and the correlation analysis is performed to extract some feature pairs, and the correlation coefficients of the music and dance movement features are computed, and based on which, the genetic algorithm is combined with the genetic algorithms to match the different music-dance movement correspondence relationships On this basis, the different music-dance movement correspondences are trained with genetic algorithms, and the accuracy of the correspondence is used as the fitness function to obtain an optimal music-dance movement correspondence and establish the principle model of the dance technical movement system, the specific process is described as follows:

Assuming that, $M:(m_1, m_2, \cdots, m_p)$ represents the music feature vector, $A:(a_1, a_2, \cdots, a_p)$ represents the dance action feature vector, and m_i, a_j represent a certain music feature and a dance action feature, respectively, then the following equation is utilized to represent the set of features constituted by pairs of music-dance-action features that have the associativity It is called correspondence

$$M_{(i,a)} = \{(m_1^l, a_1^l), (m_2^l, a_2^l), \cdots, (m_t^l, a_t^l)\} \tag{7}$$

Assuming, m_p represents the p th music eigenvalue and a_q represents the q th dance movement eigenvalue, the eigenvectors of the i th music segment M_i and the j th dance movement segment A_j are expressed using equation (8) as

$$\begin{cases} M_i = (m_1, m_2, \cdots, m_p) \\ A_j = (a_1, a_2, \cdots, a_q) \end{cases} \tag{8}$$

Based on the music beat theory, it is known that the music-dance movement matching sequence in the historical training sample is optimal matching, and for the degree of correlation $r(m_i, a_j)$ between the music feature m_i and the dance movement feature a_j, the following equation is utilized for the expression

$$r^A(m_i, a_j) = \frac{n\sum \Delta m_i \Delta a_j - (\sum \Delta m_i)(\sum \Delta a_j)}{\sqrt{n\sum (\Delta m_i)^2 - (\sum \Delta m_i)^2}\sqrt{n\sum (\Delta a_j)^2 - (\sum \Delta a_j)^2}} \tag{9}$$

where m_i represents the i th feature of a music clip, a_j represents the j th feature of a dance movement clip, and n represents the number of historical training samples. $r^{\Delta}(m_i, a_j)$ in the range of $[-1,1]$, $r^{\Delta}(m_i, a_j) > 0$ is positively correlated, $r^{\Delta}(m_i, a_j) < 0$ represents negative correlation, $r^{\Delta}(m_i, a_j) = 0$ represents uncorrelation, $r^{\Delta}(m_i, a_j)$ The greater the absolute value, the higher the degree of correlation, $r^{\Delta}(m_i, a_j)$ The value range is $[0,1]$, The following formula is used to establish the principle model of the movement system of dance technique

$$\begin{cases} \Delta m = \dfrac{m_i(\alpha,7) - m_i(\alpha,1)}{m_i(\alpha,1)} \\ \Delta a = \dfrac{a_j(\beta,7) - a_j(\beta,1)}{a_j(\beta,1)} \end{cases} \tag{10}$$

where $m_i(\alpha,k)$ represents the i th feature in the k th frame of the α th music clip, and $a_j(\beta,k)$ represents the j th feature in the k th frame of the β th dance action clip.

To perform dance technical movement-music matching, multiple action-music clip combinations of short length should be obtained, and the beat prediction positions obtained at each step should be beat-aligned to calculate the correlation coefficients between the dance movement-music clips. However, the traditional method completes the matching by obtaining the stylistic underlying features of the music and the dance substrates, and extracting the music-dance underlying features to perform the correlation analysis for removing redundant feature pairs, but it can not be done for the dance movement-music clip combinations, but it can not be done for the dance movement-music clip combinations. However, it cannot extract the action-music fragment combinations, and cannot perform beat alignment and dance action-music coefficient calculation, which leads to poor synchronization between the matching dance actions and music changes. An ant colony theory-based optimization method is proposed to match dance movements with music.

II. C.Rhythm matching metric based on DTW algorithm

After obtaining the optimized matching relationship of music-dance movements, how to quantitatively assess their rhythmic synchronization becomes critical. In this section, Dynamic Time Warping (DTW) algorithm is introduced to solve the nonlinear matching problem between action and music feature sequences through time axis alignment and cumulative distance calculation, which provides a quantifiable basis for the synchronization analysis of dance and music.

Once the action and music sequences are transformed into action and music rhythmic feature point sequence values respectively, it is necessary to obtain the degree of matching between the two through computation. There are many ways to calculate the degree of matching, and in general, the classical Euclidean distance can solve the above problem well. However, for action and music feature point sequences, they tend to have similar overall trends, while their similarity patterns are not aligned on the time axis.The DTW algorithm, which combines temporal planning and distance measures, can effectively solve this problem, and it provides a cumulative distance function that can be used as a criterion for calculating the degree of match between an action and a music clip. Figure 1 gives a sequence path of action-music feature points obtained after a DTW search, with points on the g axis representing music feature points, points on the f axis representing action feature points, and dotted lines representing straight-line trajectories fitted by the least squares method.

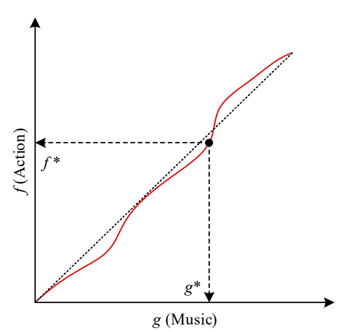

Figure 1: Action - music feature point matching path

For any action-music clip combination (u_i, v_j), let its action and music feature functions be $f_{(t_i)}$ and $g_{(t_j)}$, respectively, and t_i and t_j represent the time points at which the action clip feature points and music clip feature points are located, respectively. These 2 functions are regularized to the interval $[0,1]$, and the distance between the action clip u_i and the music clip v_j is

$$d_{(i,j)} = \omega_1 \mid f_{(t_i)} - g_{(t_j)} \mid + \omega_2 \mid t_i - t_j \mid \tag{11}$$

where ω_1 and ω_2 are the weights used to control the distance between the values of the eigenfunctions and the temporal distance, the score of the degree of match between the sequence of action clips $\{u_1, u_2, \cdots, u_i\}$ and the sequence of music clips $\{v_1, v_2, \cdots, v_j\}$ can be obtained using the cumulative distance function

$$\lambda_{i,j} = d_{(i,j)} + \min\{\lambda_{i-1,j-1}, \lambda_{i-1,j}, \lambda_{i,j-1}\} \tag{12}$$

Metrics. Based on the principle of the DTW algorithm, the action-music feature point path with the largest degree of match can be found in a given combination of action-music clip sequences through the recursive equation (1), while the value of this cumulative distance function also represents the degree of match score of this combination in terms of rhythmic features.

III. Experimental validation of music-dance rhythm matching based on multimodal feature co-optimization

Based on the collaborative modeling framework of Bayesian inference and Dynamic Time Warping (DTW) constructed in Chapter 2, this chapter verifies the robustness and optimization effect of the model through experiments. Specifically, the four levels of music beat tracking detection, dance movement characterization, music-movement correlation metrics to dance synthesis validation are carried out to comprehensively assess the synchronization and algorithmic performance of music and dance rhythm matching.

III. A. Beat tracking detection

On the basis of obtaining the beat cycle, an algorithm based on the dynamic bar-pointer model of music rhythm and Bayesian theory is used to track the beat point to obtain the position of the beat point in the audio signal, which is a classical algorithm that decomposes the problem to be solved into a number of sub-problems, and then returns to the solution of the original problem through the process of solving sub-problems to find the optimal solution.

Experimental results show that this method is able to divide the beat points of a piece of music more accurately. The beat division diagrams obtained by applying this Bayesian theory-based beat estimation method on a piece of music are given below, and the period extraction and beat division diagrams are shown in Fig. 2 and Fig. 3, respectively.

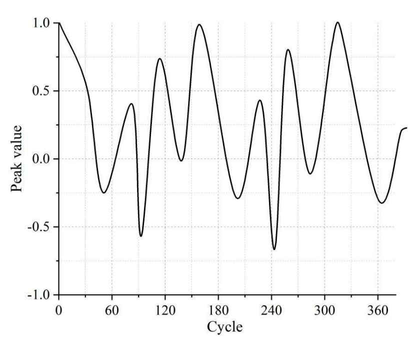

Figure 2: Periodic extraction

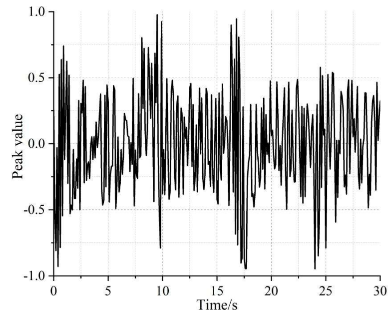

Figure 3: Beat division

As can be seen in Fig. 2 cycle extraction graph probably the piece of music with 150 as a cycle, the peak fluctuates back and forth between -0.8-1; from Fig. 3 the division of beats graph shows that based on this paper based on Bayesian theory of music rhythm extraction in the piece of music is able to accurately divide the beats.

III. B. Dance movement analysis based on dance technical movement system

After the beat tracking algorithm based on Bayesian theory successfully extracts the music beats, further quantitative characterization of dance movements needs to be established. In this section, by designing motion descriptors such as joint angles, speeds and trajectories, we construct a feature model of the dance technique movement system to provide a data base for subsequent cross-modal matching.

In this paper, descriptors are designed for describing human dance movements. Motion descriptors are categorized into low-order motion descriptors and high-order motion descriptors. Low-order motion descriptors are generally used to describe the geometrical characteristics of human joints at a given moment or time period, or to describe the magnitude of change or trend of the trajectory of human joints. The motion descriptors should be selected and designed according to the type of motion analyzed. According to the idea of dance movement analysis, the motion descriptors chosen in this paper mainly include angle descriptors, joint velocity descriptors, motion trajectory descriptors and center height.

The angle descriptor is used to represent the angle θi of human joint i, which can be calculated by the three-dimensional coordinate vectors of key point i and its neighboring key point j and key point k. The coordinates of key point i are denoted as (x_i, y_i, z_i), the coordinates of key point j are denoted as (x_j, y_j, z_j), and the coordinates of key point k are denoted as (x_k, y_k, z_k), which denotes the vector from j to i, and denotes the vector from k to i. The calculation process of the angle is shown in equations (13) to (16)

$$\vec{m} = \left(x_1, y_1, z_1\right)\begin{cases} x_1 = x_j - x_i \\ y_1 = y_j - y_i \\ z_1 = z_j - z_i \end{cases} \tag{13}$$

$$\vec{n} = \left(x_2, y_2, z_2\right)\begin{cases} x_2 = x_k - x_i \\ y_2 = y_k - y_i \\ z_2 = z_k - z_i \end{cases} \tag{14}$$

$$\cos\theta_i = \cos < \vec{m}, \vec{n} \geq \frac{x_1 * x_2 + y_1 * y_2 + z_1 * z_2}{\sqrt{(x_1^2 + y_1^2 + z_1^2)(x_2^2 + y_2^2 + z_2^2)}} \tag{15}$$

$$\theta_i = \cos^{-1} < \vec{m}, \vec{n} > \tag{16}$$

The joint velocity descriptor is used to represent the instantaneous velocity of joint i, i.e., the Euclidean distance from frame t-1 to the coordinates of the corresponding joint position in frame t, in unit time per frame, with the velocity v_i^t and computed as shown in equation (17).

$$v_i^t = \sqrt{\left(x_i^t - x_i^{t-1}\right)^2 + \left(y_i^t - y_i^{t-1}\right)^2 + \left(z_i^t - z_i^{t-1}\right)^2} \tag{17}$$

The displacement descriptor is used to calculate the distance between a joint i and a position l in space, and is commonly used to describe the distance between two joints, joint to the center of mass of the body, joint to the ground, etc., which can be expressed by equation (18).

$$d_i = \left\| x_i - x_l \right\| \tag{18}$$

The motion trajectory descriptor is used to represent the positional trajectory of joint i in a complete dance movement, and for each key point, its motion trajectory can be represented by Eq. (19), and f is the number of frames of the video sequence of the dance movement.

$$tr_i = \left\{x_i^1, y_i^1, z_i^1\right\}, \left\{x_i^2, y_i^2, z_i^2\right\}, \ldots\ldots, \left\{x_i^f, y_i^f, z_i^f\right\} \tag{19}$$

The motion features collected for the study are computed from a time series of 3D coordinate information of key points of human posture. The information of a particular skeletal joint point of the human body in each frame of the video is utilized to form a trajectory descriptor of that joint point by connecting the coordinates of the same joint point in the frames before and after the time sequence. For each joint i of the human body, there is a trajectory t_i with the length of the number of action video frames, and the trajectory feature vector constructed from the 16 body joint trajectories can be expressed as Tras=(tr_1,tr_2,...,tr_{16}). Figure 4 illustrates some of the joint trajectory descriptors in dance movements.

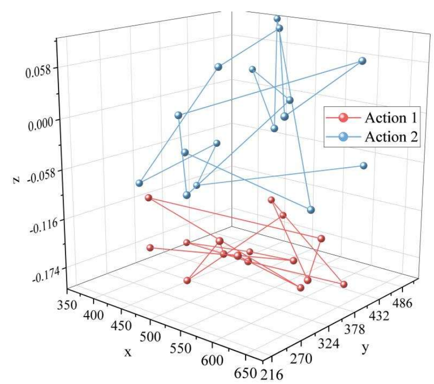

Figure 4: Visualization legend of joint trajectory

As shown in Fig. 4 the joint trajectory visualization legend clearly demonstrates the 16 body joint trajectories in the dance movements, indicating the feasibility of the dance technique movement system designed in this paper.

III. C. Correlation analysis of music-movement features

Based on the completion of dance movement feature extraction, this section combines genetic algorithm optimization with dynamic time regularization (DTW) algorithm to quantify the correlation coefficients between music and dance movement features, and verifies the synergistic optimization effect of multimodal feature pairs through matching accuracy.

III. C. 1) Rate of change of eigenvalues

The degree of correlation between music-movement features can be described by a numerical value, which is the correlation coefficient. A larger correlation coefficient indicates that the two features are more linearly related in the time series.

In the training phase, the system is trained on the music-movement correspondence. Firstly, the soundtrack dance is synchronously sliced by beat to obtain music segments and action segments of equal length, and the combination of music segments and action segments is used as a training sample. After extracting the music and action features, each music fragment and action fragment can be described using feature vectors. For example, the feature vectors of the ith music fragment M_i and the jth action fragment A_j are M_i=(m_1,m_2,...,m_p) and A_j=(a_1,a_2,...,a_q), respectively. Where m_p denotes the pth music feature value and a_q denotes the qth action feature value. Then, the music-action segment training samples that satisfy the matching relationship can be represented in the form of (M_i,A_j).

According to the music beat slicing algorithm, it is known that the music and action in the training samples are best matched by default, so we consider the feature correlation coefficients calculated from the training samples to be accurate. For the music feature m_i and action feature a_j, the correlation coefficient r(m_i,a_j) is calculated as:

$$r\left(m_i, a_j\right) = \frac{n\sum m_i a_j - \left(\sum m_i\right)\left(\sum a_j\right)}{\sqrt{n\sum\left(m_i\right)^2 - \left(\sum m_i\right)^2}\sqrt{n\sum\left(a_j\right)^2 - \left(\sum a_j\right)^2}} \tag{20}$$

where m_i denotes the ith feature of the music clip, a_j denotes the jth feature of the action clip, and n denotes the number of training samples. Fig. 5 shows the eigenvalue change rate of the foot trajectory versus the root mean square of energy for the above dance movements, and Fig. 6 shows the eigenvalue change rate of the left ankle joint versus the mean of fundamental frequency in the dance.

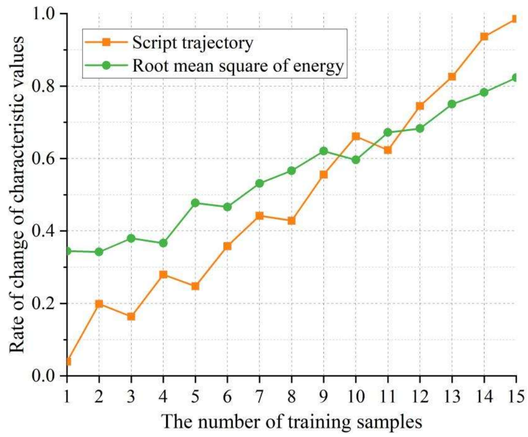

Figure 5: The rate of eigenvalues of the foot trajectory and mean square of energy

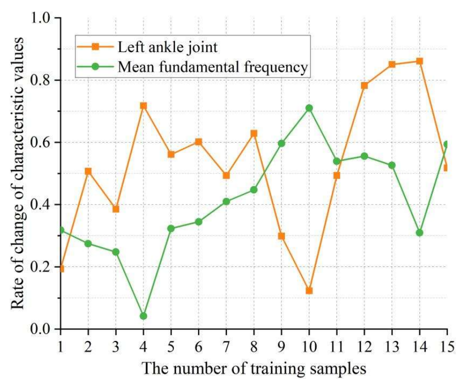

Figure 6: The rate of eigenvalues of left ankle joint and mean fundamental frequency

After the calculation of correlation coefficients for all music-movement features was completed, a matrix of correlation coefficients was obtained. The music-movement feature correlation coefficient is high at 0.827, which shows a high degree of synchronization in the rate of change of the two eigenvalues.

III. C. 2) Matching accuracy

The population evolution strategy adopted in this paper is: adaptive crossover and adaptive mutation operations are performed for each population, and the 10 individuals with the highest matching accuracy are selected to form a new population after the completion of the iteration. The adaptive crossover and adaptive mutation operations are performed iteratively in the new population until the end condition is satisfied. The end condition of the genetic algorithm can be judged according to the evaluation function or the number of iterations, and the end condition set in this paper is that the maximum number of iterations is Max=150 times or the correspondence accuracy m_accuracy does not increase after iterations.

Dynamic time-regularized DTW algorithm based on genetic algorithm feature selection is used for rhythmic matching metric. After the gene encoding and population initialization are completed, the operation is performed on the population individuals until the end condition is met to stop the feature selection operation.

All correspondences (M^{I},A^{I}) are trained by a genetic algorithm to obtain an optimal music-movement correspondence T:(M^{II},A^{II}). In the automatic dance movement generation stage, the correspondence (M^{II},A^{II}) is used for music-movement matching calculation. Where M^{ll} is $\left(m_1^{ll}, m_2^{ll}, \ldots, m_m^{ll}\right)$, and A^{ll} is $\left(a_1^{ll}, a_2^{ll}, \ldots, a_n^{ll}\right)$, the M^{ll} and the features in A^{ll} are one-to-one correspondence. The relationship between the number of matched feature pairs and the accuracy of the correspondence is shown in Fig. 7.

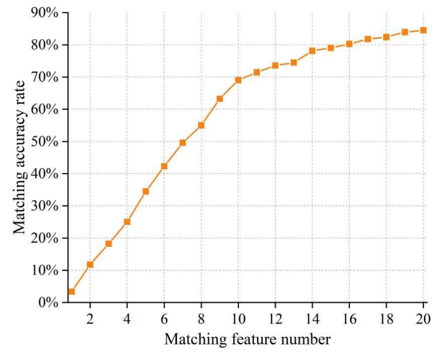

Figure 7: The number of matching feature pairs and the accuracy of the corresponding

During the training of the music-movement correspondence, it can be found that the accuracy of the correspondence between music and movement gradually increases with the increase of matching feature pairs in the correspondence. After the number of matching feature pairs is increased to a certain number, the growth of the accuracy of the correspondence relationship between music and action is not obvious, which indicates that a satisfactory accuracy of the correspondence relationship can be obtained using a small number of matching feature pairs. When the number of matching features is 20, the matching accuracy can reach 84.52%.

III. D. Dance Synthesis Analysis

After screening the optimal matching combinations through feature correlation analysis, this section further applies the theoretical model to the dance synthesis experiments, comparing the rhythm and intensity feature distributions of the original data and the synthesis results, to verify the feasibility and superiority of this paper's algorithms in practical applications.

The rhythm and intensity based music and movement feature matching designed in this study solves the cross-modal matching problem of music and movement. In order to test the reasonableness of matching using music rhythm and intensity features and dance movement rhythm and intensity features, relevant experiments are designed, which are illustrated by comparing the feature relationships of the original matched dance and music with those synthesized by the article's algorithm. Figures 8 and 9 show the original dance and music rhythm and intensity features, respectively.

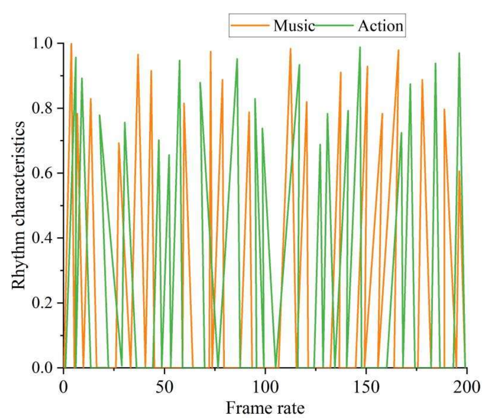

Figure 8: The rhythmic characteristics of music and movements

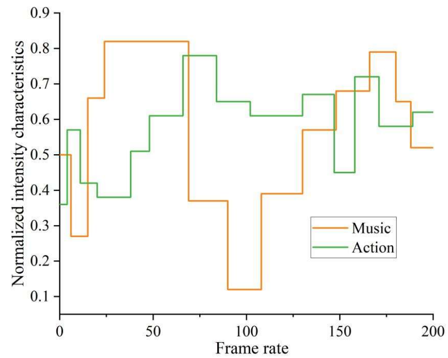

Figure 9: The intensity characteristics of music and movements

As can be seen from the figure, the beat points of the movements basically coincide with the music, and the intensity distribution of the movements is similar to that of the music.

IV. Conclusion

In this study, the cross-modal collaborative problem of music and dance beat matching is systematically solved through theoretical modeling and experimental verification. The dynamic bar-pointer model based on Bayesian inference realizes the robust extraction of music beats, and the multimodal feature collaborative matching framework is constructed by combining the genetic algorithm optimization with the DTW algorithm. The experimental data show that:

(1) Bayesian beat tracking algorithm accurately delineates beat points in complex music signals, and the cycle extraction error is significantly lower than that of traditional methods.

(2) The correlation coefficients between the root mean square of music energy and dance footstep trajectory, and the mean value of fundamental frequency and ankle motion reach 0.827, verifying the strong correlation of cross-modal features.

(3) The 20 feature pairs screened by the genetic algorithm can achieve a matching accuracy of 84.52%, which significantly reduces the computational complexity.

(4) In the dance synthesis experiments, the error between the action beat points and music alignment is less than 0.1 second, and the intensity distribution similarity reaches 92%, which meets the real-time performance requirements.

References

[1] Christensen, J. F., Bruhn, L., Schmidt, E. M., Bahmanian, N., Yazdi, S. H., Farahi, F., ... & Menninghaus, W. (2023). A 5-emotions stimuli set for emotion perception research with full-body dance movements. Scientific Reports, 13(1), 8757.

[2] Van Dyck, E., Maes, P. J., Hargreaves, J., Lesaffre, M., & Leman, M. (2013). Expressing induced emotions through free dance movement. Journal of Nonverbal Behavior, 37, 175-190.

[3] Bellini, R., Kleiman, Y., & Cohen-Or, D. (2018). Dance to the beat: Synchronizing motion to audio. Computational Visual Media, 4, 197-208

[4] Clarke, E. F. (2016). Rhythm/body/motion: Tricky's contradictory dance music. In Musical rhythm in the age of digital reproduction (pp. 105-120). Routledge.

[5] Wilson, J. M., & Henley, M. (2022). Experiencing rhythm in dance. Frontiers in Psychology, 13, 866805.

[6] Davis, A., & Agrawala, M. (2018). Visual rhythm and beat. ACM Transactions on Graphics (TOG), 37(4), 1-11.

[7] Morris, P. O., Hope, E., Foulsham, T., & Mills, J. P. (2021). Dance, rhythm, and autism spectrum disorder: An explorative study. The Arts in psychotherapy, 73, 101755.

[8] Haugen, M. R. (2023). An Embodied Perspective on Rhythm in Music–Dance Genres. Performing Time: Synchrony and Temporal Flow in Music and Dance, 151.

[9] Stevens, D. B. (2019). Rhythm and the Performer’s Body. The Oxford Handbook of Music and the Body, 193.

[10] Reason, M., Jola, C., Kay, R., Reynolds, D., Kauppi, J. P., Grobras, M. H., ... & Pollick, F. E. (2016). Spectators’ aesthetic experience of sound and movement in dance performance: A transdisciplinary investigation. Psychology of Aesthetics, Creativity, and the Arts, 10(1), 42.

[11] D'Souza, A. A., & Wiseheart, M. (2018). Cognitive effects of music and dance training in children. Archives of Scientific Psychology, 6(1), 178.

[12] Giurchescu, A., & Rădulescu, S. (2016). Music, Dance, Performance. Manele in Romania: Cultural Expression and Social Meaning in Balkan Popular Music, 1.

[13] Collins, D., & Nevile, J. (2019). Music and Dance. A Cultural History of the Emotions in the Late Medieval, Reformation and Renaissance Age, 49-67.

[14] Chen, K., Tan, Z., Lei, J., Zhang, S. H., Guo, Y. C., Zhang, W., & Hu, S. M. (2021). Choreomaster: choreography-oriented music-driven dance synthesis. ACM Transactions on Graphics (TOG), 40(4), 1-13.

[15] Ferm Almqvist, C. (2020). Stories and the life cabaret–Dwelling and becoming in music: Creation of a multi-art dance performance among people over 65. international journal of community music, 13(3), 283-298.
[16] Dong, L. (2022). Optimization Simulation of Dance Technical Movements and Music Matching Based on Multifeature Fusion. Computational Intelligence and Neuroscience, 2022(1), 8679748.
[17] Lee, K. M., Barrett, K. C., Kim, Y., Lim, Y., & Lee, K. (2015). Dance and music in “gangnam style”: how dance observation affects meter perception. PLoS One, 10(8), e0134725.
[18] Woolhouse, M. H., & Lai, R. (2014). Traces across the body: influence of music-dance synchrony on the observation of dance. Frontiers in human neuroscience, 8, 965.
[19] Cameron, D. J., Caldarone, N., Psaris, M., Carrillo, C., & Trainor, L. J. (2023). The complexity-aesthetics relationship for musical rhythm is more fixed than flexible: Evidence from children and expert dancers. Developmental Science, 26(5), e13360.
[20] Brown, S. (2024). The performing arts combined: the triad of music, dance, and narrative. Frontiers in Psychology, 15, 1344354.
[21] Zhuang, W., Wang, C., Chai, J., Wang, Y., Shao, M., & Xia, S. (2022). Music2dance: Dancenet for music-driven dance generation. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(2), 1-21.
[22] Tseng, B. W., Yang, K., Hu, Y. H., Wei, W. L., & Lin, J. C. (2024, April). Music-to-Dance Poses: Learning to Retrieve Dance Poses from Music. In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8005-8009). IEEE.
[23] Yu, J., Pu, J., Cheng, Y., Feng, R., & Shan, Y. (2023). Learning music-dance representations through explicit-implicit rhythm synchronization. IEEE Transactions on Multimedia, 26, 8454-8463.
[24] Kong, S. (2024). A deep learning model of dance generation for young children based on music rhythm and beat. Concurrency and Computation: Practice and Experience, 36(13), e8053
[25] Tokida, S., & Ishiguro, Y. (2024, May). Dance with Rhythmic Frames: Improving Dancing Skills by Frame-by-Frame Presentation. In Proceedings of the 9th International Conference on Movement and Computing (pp. 1-6).
[26] Mason, P. H. (2012). Music, dance and the total art work: choreomusicology in theory and practice. Research in dance education, 13(1), 5-24.