International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 4857-4864

https://doi.org/10.70517/ijhsa464329

Research on Emotional Regulation and Expressiveness Enhancement in Music Conducting Empowered by AI Algorithms

Wenyun Shen[1,*]
[1] Communication University of Zhejiang, Hangzhou, Zhejiang, 310000, China
Corresponding authors: (e-mail: 15558002187@163.com).

Abstract AI technology promotes the intelligent development of music conducting art. This paper focuses on the innovative application of AI technology in music conducting, and proposes dynamic gesture semantic classification technology based on dynamic temporal regularization (DTW), which utilizes the scale invariance of the conductor's gesture trajectory to complete modeling and recognition. Through normalized data processing and trajectory point downsampling method, the spatial deviation is eliminated, and the gesture template matching system is constructed by combining the DTW algorithm. Experiments show that the localization accuracy of the five gesture trajectory points of the proposed model exceeds 0.95, with good localization ability. The average recognition accuracy reaches 99.31%, close to 100.00%, higher than 92.76% of the comparison model. The gesture classification accuracy is 0.902, precision is 0.936, recall is 0.954, and F1 value is 0.967, which is better than the comparison model. The model has the best performance when the number of templates is 7.

Index Terms AI technology, dynamic temporal regularization, semantic classification of gestures, template matching, music conducting gestures

I. Introduction

As an essential part of musical expression and performance, the art of music conducting has a crucial role to play [1]. It serves as a link between the substance of the music and the musicians, and it is also the most essential element other than the composer to convey the musical emotion directly to the audience [2], [3]. The core task of the conductor is to realize the complete artistic presentation of the musical work through coordination and harmony [4]. Conductors not only control the rhythm and tempo of the music through gestures, but also convey their interpretation of the music through complex body language and expressions [5]-[7]. An outstanding conductor not only needs to have exquisite technical skills, but also needs to have a deep understanding of the overall structure of the music as well as a keen insight into each voice part [8]-[10]. This means that the comprehensive quality of music conducting art relates to the level and effect of the musical performance of the performing orchestra and has an important role in music performance [11].

And for a long time, the function of the music conducting art is not clear, and it does not play the role of emotion regulation and performance of music guidance well, which hinders the development of music performing art. At the same time, the progress of technology also provides a new way for the improvement of the art of conducting. With the development of music analysis tools supported by artificial intelligence technology, conductors are able to study the works more deeply and explore different ways of interpretation, which not only improves the technical level of the performance, but also provides the audience with a richer listening experience [12]-[14].

This paper systematically researches the way of integrating the principles of gesture movement design and dynamic time regularization algorithm to improve the artistic expressiveness of music conductor's gestures and to enlighten the emotional presentation of choral performers. Based on the gesture movement design theory of motion trajectory analysis, we analyze the mapping relationship between intensity change and music emotion. Aiming at the goals of artistic design, the dynamic time regularization algorithm is proposed to solve the scale invariance and temporal offset problem of gesture trajectory. Through normalization processing and downsampling of trajectory points, a standardized feature space is constructed, combined with template matching to achieve semantic classification of gestures and precise mapping of gesture semantics. The gesture recognition classification effect of the constructed model is examined through model training and testing experiments.

II. Choral emotion regulation based on dynamic gesture semantics

II. A. Gesture Action Design

II. A. 1) Principles and characteristics of gestural movements

Principles of gesture movement: First, accuracy. From the perspective of the chorus conductor, its main responsibility is to guide all the chorus members, auxiliary personnel to scientifically and reasonably judge the direction of the conductor's gestures, in the full expression of the linear emotions in the musical works, at the same time, the basic knowledge of the conductor more clearly passed to the members, through the accurate expression of the movement and emotion can help members of the overall improvement of the musical expression. Second, standardization. Conductor in chorus rehearsal or performance, need to strictly follow the relevant normative gesture action standard or action illustration for the conductor, to maximize the avoidance of swinging blindness, once the scene command confusion, will affect the normal performance of the entire choir. Third, naturalness. The conductor is like the brain, all the musicians must follow its state playing, to always maintain a good mental outlook state, to maintain a proper body language posture, to members of the natural energy, in order to ease the overcoming of pressure and tension, all the energy to focus on the chorus performance.

The main characteristics of gesture movement: the so-called “gesture movement” is essentially a depiction of the basic trajectory of the graphic line, usually, the graphic pattern consists of three parts, that is, the beat line, the beat point, the line of reflection. In the process of shaping and presenting the image of a musical work, the main source of expression of the choir is based on the points and lines of the conductor's gestures, and through the effective fusion of the two, the melody is constructed into a very rich and colorful music.

In addition, in the design of choral conductor gestures, the difference in the size of the strength of the action affects the final design effect to some extent, and due to the interdependence of the sound and the strength of the two sides of the slightest change may have a serious impact on the design standard of the conductor's action. Therefore, the different strengths and weaknesses of the work, to some extent, make the conductor in the process of action amplitude changes in the size of the final determination of the strength of the force can follow the “from weak to strong” “from urgent to slow” changes in the design, not only to enhance the musical works of the This not only enhances the mobility and diversity of the musical work, but also gives the work more color and emotional changes along with the change of intensity.

II. A. 2) The artistry of gesture movement design

The most essential performance of choral conductor's gesture design originates from “a pair of hands mastering language”, which is able to dig out the deeper meaning of music works. The choral conductor's expression of music mainly relies on the design of the conductor's gestures, and from another point of view, the design of gestures provides a platform for the choral conductor and even all the members to express their own emotions. Therefore, how to use AI technology, the gesture movement for scientific and reasonable design is particularly important, whether it is the overall grasp of the performance process, or the expression of the musical story, the attitude towards art and processing skills will be greatly enhanced, through the gesture movement systematic design and planning, not only to show the connotation and charm of the music works, and to realize the innovation of the music works in terms of emotional expression.

II. B. Dynamic gesture semantic classification technique

II. B. 1) Data processing

The semantics of dynamic gestures is generally not affected by the range of motion, which is due to the differences in the users' gesture habits, arm extension, etc., resulting in different ranges of motion of the gestures, but in real life this does not affect the communication of the semantics, and this characteristic of dynamic gestures is called scale invariance. Taking the gesture of four beats in music conducting as an example, there are obvious scale differences between the trajectory P and the trajectory Q in the direction of x, but both of them express the same semantics. In order to eliminate or mitigate the effects of scale and position, all trajectory points are normalized in the x and y directions, respectively.

In the start and end positions and turn positions, the gesture movement is slower and the trajectory points are relatively dense, and the normalized statistical parameters are more likely to be affected by these positions. In addition, the input data dimension of the DTW-based classification model is not fixed and is positively correlated with the number of trajectory points, and locations with dense trajectory points will be regarded as having higher role weights. In fact, the normalized motion trajectories of the two gestures roughly match each other, and the main difference comes from the location where the trajectory points are densely packed. Therefore, the trajectory points are downsampled according to the Euclidean distance, and only the trajectory points whose interval distance is larger than the threshold are retained, and then the normalization process is done based on the downsampled trajectory points. Figure 1 shows the obtained P, Q trajectories.

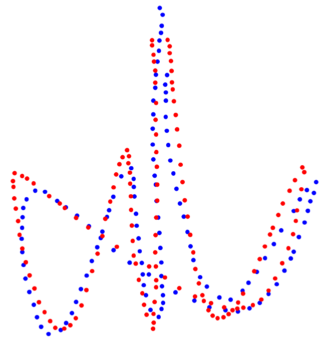

Figure 1: Dynamic gesture trajectories after downsampling and normalization

II. B. 2) Dynamic time regularization

DTW is to find the best alignment strategy between two time series for comparing the similarity of two time series, and is a classical algorithm for template matching. Usually there is a difference in timing length or speed between dynamic gestures and templates, DTW uses the time regularization function to describe the temporal correspondence between dynamic gestures and templates, and the solution objective of the time regularization function is the minimum cumulative distance between the two time series, which is known as the regularized path distance (WPD), and the paths that satisfy the minimum cumulative distance paths are known as the regularized paths (WP).

Figure 2 shows the temporal correspondence obtained by DTW solving for two time sequences A and B. The horizontal axis represents the time, and the vertical axis represents the data features; sequence A is shifted to the left on the horizontal axis as a whole with respect to sequence B, but it has a similar change waveform on the vertical axis, and the temporal correspondence obtained by solving for the regularized path distance through DTW can eliminate the influence of the time sequence shift and focus more on the feature change waveform.

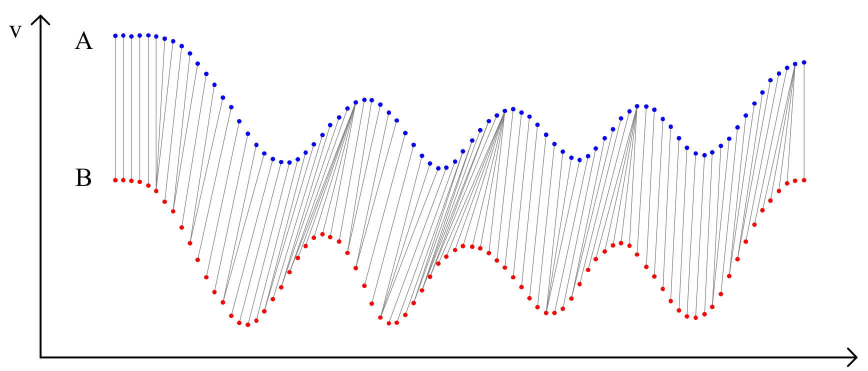

Figure 2: The normalized paths of the two time series

Assuming that the template trajectory satisfies $P=\{p_0,p_1,\ldots,p_{m-1}\}$, and the input trajectory satisfies $Q=\{q_0,q_1,\ldots,q_{n-1}\}$, construct a lattice matrix with $m\times n$, and the element corresponding to the i th row of the matrix and the j th column of the matrix is $D_{i,j}$, which denotes the distance of the regular path obtained by comparing P's first i trajectory points and Q's first j trajectory points to get the regularized path distance, which is calculated as:

$$D_{i,j}=\begin{cases} Dist(p_i,q_j), & i=0, j=0 \\ Dist(p_i,q_j)+D_{i,j-1}, & i=0, 0<j<n \\ Dist(p_i,q_j)+D_{i-1,j}, & j=0, 0<i<m \\ Dist(p_i,q_j)+\min\{D_{i-1,j},D_{i,j-1},D_{i-1,j-1}\}, & 0<i<m, 0<j<n \end{cases} \quad (1)$$

where the distance function $Dist$ uses the Manhattan distance function, calculated as:

$$Dist(x,y)=\sum_{i}^{n}\left|x_i-y_i\right| \quad (2)$$

where x and y are the trajectory points involved in the comparison, and n is the feature dimension of the trajectory points, respectively. The regularized path distance of trajectories P, Q is denoted as:

$$D_{dtw} = D_{m-1,n-1} \tag{3}$$

The dynamic gesture trajectory points are characterized by two-dimensional coordinates, and the DTW algorithm is used to solve for the regularized path distance between the two two two-dimensional dynamic gesture trajectories in Fig. 1, and Fig. 3 shows the obtained regularized paths. The smaller the regularized path distance is, the more similar the two trajectories are. In addition, the distance value is also related to the number of trajectory points involved in the comparison, and the greater the number, the greater the cumulative distance, so the average distance of all input trajectory points is used as a similarity measure, and the average regularized path distance is denoted as:

$$\overline{D}_{dtw} = \frac{D_{dtw}}{n} \tag{4}$$

The classification method based on template matching requires the collection of standard dynamic gesture trajectories as template data, and then the average regularized path distance D_{dtw}^{j} between the input gesture trajectories and each template is calculated sequentially, $j \in \{1,2,\ldots,l\}$, and the closest dynamic gesture category is obtained based on the minimum distance:

$$hand_{cls} = \underset{j \in \{1,2,\ldots,l\}}{\arg\min}\{D_{dtw}^{j}\} \tag{5}$$

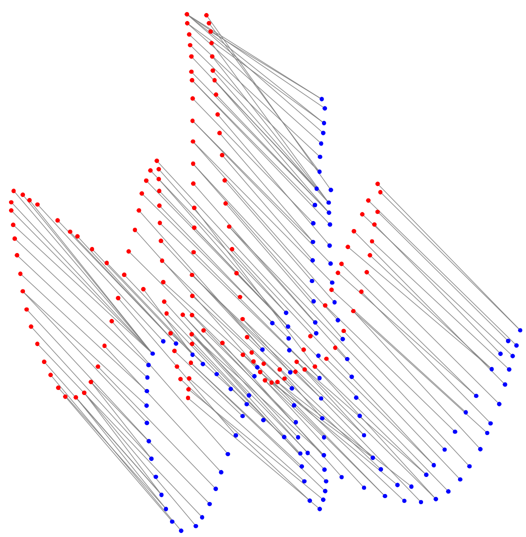

Figure 3: The regular paths of the two dynamic gesture trajectories

II. C. The main means of emotional revelation in choral conducting

II. C. 1) Gestures

Gesture movement in choral conducting is an important communication channel between conductor and chorus members, and hands are also the tools for conductors to express their own feelings, gesture movement has the role of coordinating the members to cooperate, leading the music emotion, and inspiring the members to feel. Different shapes and sizes of gestures complete the chorus beat instructions, dynamic changes, pitch instructions and other command content, excellent gestures can ensure that the music rhythm of the strength of the weak, sharp and slow, pitch rise and fall and other elements of the music to be accurately conveyed. The conductor's right hand grasps the graphic line point of the work, and the right hand movement changes to convey the speed, strength, and beat of the piece; the left hand metaphorically sings the emotion, and depicts the change of the musical color and the fluctuation of the connotation of the emotion of the work. The chorus members can control the progress of music interpretation and the direction of emotion through the observation of whether the conductor's wrist is relaxed, the height and position of the arm, the amplitude of baton swing and the change of the landing point. In choral conducting, the conductor's swinging skills and left and right hand movements will affect the emotional perception of the members, and the gestures that conform to the connotation of the music are the basis for the precise revelation of emotions in choral conducting.

II. C. 2) Facial expressions

Choral conductor in the conductor's facial expression is the first feeling of the members of the music conductor activities, god changes directly on behalf of the conductor in the music of the chorus of the degree of intoxication, able to drive the other members of the singing group immersed in the artistic atmosphere. If the conductor's facial expression is not obvious in the actual conductor, it is not only unfavorable to the revelation of music emotion, but also affects the proximity of the members to the chorus. A rich facial expression and calm and collected conductor's look can not only accurately interpret the artistic situation and thoughts and emotions of the musical work, but also increase the degree of recognition of the members of the group to the music conductor. In order to better convey the emotion in actual performance, the conductor should follow the emotional direction of the work and gesture changes, so that his own expression is rich and varied, such as in the climax of the piece to maintain an uplifting look, and in the progression of the piece to relax the facial muscles appropriately. Among them, the change of mouth and eyes is an important component of facial expression, such as the focus of the eyes to gather and the degree of eye opening and closing reflect the degree of the conductor's intoxication in the music, which will bring emotional rendering to the audience and the chorus members. Generally speaking, the singer squints his eyes to indicate that the singing is about to wind down, and prepares for the weakening or ending of the music in advance, while the staring is the music is about to reach the climax, which requires a strong emotional gestation. Conductors use frequent eye gestures in choral singing to inspire emotion, or use the mouth to emphasize the pronunciation of the key choral nodes, both of which can inspire the members to express their emotions with the correct vocal technique. Conductors should correctly recognize the positive role of their own micro-expressions in stimulating the members to perceive the music emotionally, and drive the members to think emotionally, in order to sublimate the artistic connotation of the music chorus.

II. C. 3) Body Rhythm

Choral performance of the audience on the conductor's facial expression and gesture movement perception is not obvious, but can intuitively feel the conductor's body rhythm, and through the undulation of body rhythm to feel the detailed changes in the music emotion, at the same time, the body rhythm to create a musical atmosphere is also the audience in-depth feeling of the song emotional connotation of the direct way. The body rhythm is different from the dance performance, overly exaggerated swing amplitude is easy to let the audience's appreciation of the center of gravity from the auditory transfer to the visual, can not enhance the emotional enlightenment effect. For this reason, the conductor should pay attention to their own swing amplitude, and at the same time strengthen the thinking of the melody of the music, and be good at adjusting the body posture according to the rhythmic changes of the music, and it is best to do the body with the music to adapt to the natural rhythms of the body movement, the body movement is neither overbearing and can be very good to show the music ups and downs, so that the audience and the group can be able to observe through the conductor's body response to the music can be a clear sense of the emotion, the appropriate body posture and the correct body rhythm can implicitly complete the performance. Appropriate body posture and correct body movement can implicitly complete the emotional enlightenment, so that the members of the group to the music to produce understanding of the conductor to produce emotional resonance.

III. Gesture classification test based on dynamic time regularization

III. A. Performance of DTW-based classification models

III. A. 1) Gesture trajectory point localization training

A dataset containing 5000 images of music conductor gesture video frames is selected as the training samples, and 75% of them are used as the training set and the remaining 25% as the test set, and 80 epochs of iterative training are performed to validate the gesture trajectory point localization ability of the DTW-based classification model. Figure 4 shows the confusion matrix of gesture trajectory point localization obtained by the model on the test set. From the confusion matrix, we can see that the five gesture trajectory point localization accuracies of the model reach 0.98, 0.97, 0.97, 0.99, 0.99, all of them are above 0.95 and very close to 1.00, which indicates that the gesture trajectory point localization ability of the DTW-based gesture classification model is good, and it is able to identify the accurate gesture trajectories.

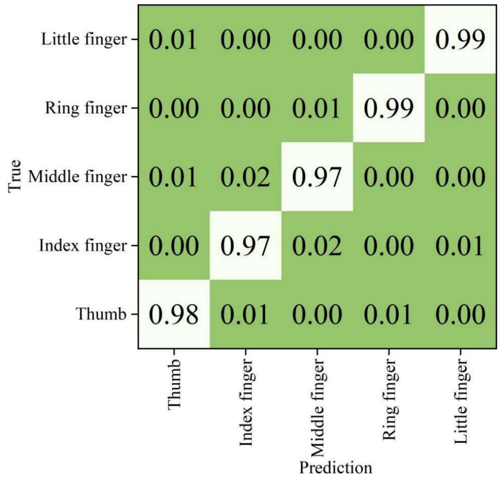

Figure 4: Gesture trajectory point location confusion matrix

III. A. 2) Gesture Recognition Accuracy

The 3750 pictures in the training set are categorized by the video files they belong to, 10 files in total, and the gesture recognition accuracy of the model in each file is counted. The gesture recognition accuracy of the model of this paper in specific pictures is compared with the same type of model CTPGR. Figure 4 shows the accuracy rate of music conductor gesture recognition. The recognition accuracy rate of this paper's model in music conductor gesture recognition in 10 files is more than 95.00%, the highest reaches 100.00%, and the average accuracy rate reaches 99.31%. The highest recognition accuracy of the same type of comparison model is 95.16%, the lowest is only 90.35%, and the average accuracy is 92.76%. From the comparison results of the gesture recognition accuracy of the training set, we can judge that the model of this paper has a higher accuracy of music conductor gesture recognition, which is close to 100.00%.

Table 1: The accuracy rate of gesture recognition for music conductors

File	Prediction	N	S	D	I	Accuracy rate (Ours)(%)	Accuracy rate (CTPGR)(%)
003	370	375	0	3	2	98.67	94.32
006	371	375	0	2	2	98.93	91.51
009	371	375	0	3	1	98.93	93.27
012	375	375	0	0	0	100.00	90.35
015	375	375	0	0	0	100.00	91.22
018	373	375	0	1	1	99.47	95.16
021	375	375	0	0	0	100.00	93.09
024	371	375	0	2	2	98.93	91.26
027	372	375	0	1	2	99.20	92.34
030	373	375	0	1	1	99.47	91.25
Total number	3749	3750	0	13	11	99.31	92.76

III. B. Performance comparison of different models for gesture classification

The experiments in this paper use Accuracy ($\lambda_{Accuracy}$), Precision ($\lambda_{Precision}$), Recall (λ_{Recall}) and F1 value (F1) as the evaluation indexes of the music conductor gesture classification model. The SentiWordNet and TextCNN models are selected as comparison models to compare their performance with this paper's model for the four metrics. Table 2 shows the model classification performance. The music conductor gesture classification accuracy of this paper's model is 0.902, the precision is 0.936, the recall is 0.954, and the F1 value is 0.967, which are all above 0.900. The performance of the 4 indexes of the 2 comparison models did not exceed 0.900, so this paper's model has a better music conductor gesture classification accuracy and can successfully complete the task of categorizing the recognized gestures.

Table 2: Model classification performance

Model	λAccuracy	λPrecision	λRecall	F1
SentiWordNet	0.689	0.724	0.781	0.756
TextCNN	0.756	0.799	0.825	0.810
Ours	0.902	0.936	0.954	0.967

III. C. Effect of different number of templates on experimental results

The number of templates is related to the final effect of matching classification, this section explores the effect of different number of templates on the gesture recognition effect of the model. Table 3 shows the experimental results of gesture recognition classification of this paper's model when the number of templates is 1, 3, 5, 7 and 9. When the number of templates is 7, the precision rate of this paper's model is 0.9877, the recall rate is 0.8798, the accuracy rate is 0.9871, and the F1 value is 0.8865.The values of the four indexes for the other template quantities range from 0.7000 to 0.8000, and all of them are more than 0.8000.Selecting the number of templates to be 7 improves the model's recognition and classification effect of the gestures of music conductor.

Table 3: Influence of the number of templates on the experimental results

Number of templates	Precision	Recall	Accuracy	F1
1	0.7177	0.7107	0.7708	0.7717
3	0.7806	0.7717	0.7761	0.7761
5	0.7886	0.7767	0.7801	0.7801
7	0.9877	0.8798	0.9871	0.8865
9	0.7581	0.7567	0.7507	0.7577

IV. Conclusion

In this paper, we propose an AI music conductor assistance method based on DTW dynamic gesture classification model and verify its practical performance advantages. In 80 iterations, the model's five gesture trajectory point localization accuracies were 0.98, 0.97, 0.97, 0.99, and 0.99, respectively. The accuracy of music conductor gesture recognition for all 10 video files is higher than 95.00%. The values of the four indicators of classification performance were 0.902, 0.936, 0.954, and 0.967, which were higher than 0.900. When the number of templates is set to 7, the four indexes reach 0.9877, 0.8798, 0.9871, 0.8865, which are all more than 0.8000, and the classification ability of music conductor gesture recognition is better, which can mobilize the performer's emotional expression. The breadth of applicability of the classification model can be improved in the future by combining the cross-modal emotion computing framework.

References

[1] Carnicer, J. G., Garrido, D. C., & Requena, S. O. (2015). Music and Leadership: the Role of the Conductor. International Journal of Music and Performing Arts, 3(1), 84-88.

[2] Kumar, A. B., & Morrison, S. J. (2016). The conductor as visual guide: Gesture and perception of musical content. Frontiers in psychology, 7, 1049.

[3] Morrison, S. J., & Demorest, S. M. (2018). Once from the top: Reframing the role of the conductor in ensemble teaching. Vocal, instrumental and ensemble learning and teaching: An Oxford handbook of music education, 283-300.

[4] Regier, B. J., Scherer, A. D., & Silvey, B. A. (2020). An examination of undergraduate conductor practice behavior. Journal of Music Teacher Education, 29(3), 37-49.

[5] Silvey, B. A., & Fisher, R. A. (2015). Effects of conducting plane on band and choral musicians' perceptions of conductor and ensemble expressivity. Journal of Research in Music Education, 63(3), 369-383.

[6] Claraco, A. V. (2015). The process of nonverbal communication between choir and conductor. Revista Electrónica de LEEME, 36, 49-70.

[7] Silvey, B. A., & Koerner, B. D. (2016). Effects of conductor expressivity on secondary school band members' performance and attitudes toward conducting. Journal of Research in Music Education, 64(1), 29-44.

[8] Rucsanda, M. D. (2020). The role of the Conductor of children choir in the improvement of Interpersonal Communication through Music. Bulletin of the Transilvania University of Braşov, Series VIII: Performing Arts, 13(2-Suppl), 257-266.

[9] Meals, C. D. (2020). The question of lag: An exploration of the relationship between conductor gesture and sonic response in instrumental ensembles. Frontiers in Psychology, 11, 573030.

[10] Nápoles, J., & Silvey, B. A. (2017). Effects of conductor baton use on band and choral musicians' perceptions of conductor expressivity and clarity. Journal of Research in Music Education, 64(4), 474-486.

[11] Montemayor, M., & Silvey, B. A. (2019). Conductor expressivity affects evaluation of rehearsal instruction. Journal of Research in Music Education, 67(2), 133-152.

[12] Karipidou, K., Ahnlund, J., Friberg, A., Alexanderson, S., & Kjellström, H. (2017, May). Computer analysis of sentiment interpretation in musical conducting. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 400-405). IEEE.

[13] Globerson, E., Flash, T., & Eitan, Z. (2021). Space, time and expression in orchestral conducting. Space-Time Geometries for Motion and Perception in the Brain and the Arts, 199-212.
[14] Li, B., Liu, X., Dinesh, K., Duan, Z., & Sharma, G. (2018). Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. IEEE Transactions on Multimedia, 21(2), 522-535.