ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on the Mechanism and Path of Modern Technology Assisting Music Aesthetic Experience - Based on the Perspective of Perceptual Style Presupposition Construction

Xiangkui Fu[1,*] *and Haipeng Du*[2]

[1] Jiaying University, Meizhou, 514015, Guangdong, China
[2] Hechuang (Beijing) Technology Co., Ltd., Beijing, 100088, China

SUMMARY: *The continuous development of digital technology, artificial intelligence and multimedia interaction is driving the transformation of music aesthetic experience from single auditory reception to multi-modal collaborative perception. From the perspective of perceptual style presupposition construction, combined with the theory of Gestalt perceptual organization, this paper analyzes the role mechanism of music map, music animation, performance video, dynamic staff and intelligent interactive presentation on aesthetic understanding, auditory organization and experience optimization around the technical links of audio feature extraction, visual coding mapping, dynamic spectrum surface synchronous presentation and interactive feedback regulation. It is verified by experiments with 48 music samples and 96 subjects. The results showed that the comprehensive experience score under the comprehensive preset condition reached 85.8 points, which was 15.5 points higher than that without preset. The comprehensive atlas improved the aesthetic understanding of non-music professionals by 17.6 points. The average comprehensive score of multi-path collaboration was 87.4, which was better than that of each single path. The research provides operable theoretical basis and practical reference for music aesthetic guidance, teaching communication and technology application in the digital age.*

KEYWORDS: *Music aesthetic experience; Perceptual style preset; Visual aid; Modern technology*

1 Introduction

The continued evolution of digital technology, artificial intelligence and multimedia delivery is reshaping the way music is perceived, understood and shared. In the past, music appreciation relied more on hearing itself, and the receiver tended to gradually grasp the melodic line, rhythmic organization, voice part relationship and emotional tension through repeated listening. The intervention of technology has broadened the channels of music aesthetics, but also raised new questions: through what internal mechanism modern technology affects the aesthetic experience of music, and how it should play a positive role in specific applications without making the technology display obscure the music ontology, this issue still needs to be further clarified.

Music aesthetic experience is not the mechanical reception of external sound, but the result of the active organization, association and integration of music elements in the process of perception. Melody direction, rhythm density, strength change, timbre contrast and

*kaer336699@163.com
https://doi.org/10.65102/is2026323 Published: 30 April, 2026

structure level, only when they are endowed with some internal order in the consciousness of the receiver, can they be condensed from scattered sounds into a comprehensible whole state. Especially in the face of works with strong polytonality, complex structure or unfamiliar style, without necessary perceptual guidance, listeners are easy to stay in the level of "hearing the sound" but difficult to form a clear overall impression, which will affect the depth and quality of music experience. It can be seen that one of the key links of music aesthetics lies in whether listeners can form appropriate perceptual expectations and structural prejudgments before or during listening.

Based on this, from the perspective of perceptual style presupposition construction, this paper discusses the function logic and realization path of modern science and technology to help music aesthetic experience, focusing on the analysis of music atlas, music animation, performance video, dynamic staff and other visual auxiliary methods, and combines technical implementation and experimental verification. To investigate the specific effects of different pathways in aesthetic understanding, auditory organization, and experience optimization. It was hoped that this paper could provide an analytical framework for music aesthetic research that took into account both theoretical interpretation and technical practice, and also provide reference ideas for the update of music appreciation, teaching communication and aesthetic guidance in the digital age.

2 Related work

In recent years, there has been an increasing number of studies focusing on the intervention of modern technology in music acceptance and music understanding. Lima and other systems have sorted out the main types, expression logic and application scenarios of music visualization, and pointed out that visualization has expanded from simple sound presentation to multiple levels of analysis, teaching and interactive experience [1]. Georges and Seckin further applied network diagram, multidimensional scaling analysis and support vector machine to music information visualization, indicating that data structuring and visual mapping can strengthen the audience's grasp of music relationship and style differences [2]. In more detailed applications, Itoh et al. proposed an interactive pitch trajectory visualization method for multiple singing versions of the same song, which provides support for comparative listening of complex music information [3]. Park, on the other hand, discussed the auxiliary value of dynamic visualization of strength in music learning from the perspective of practice and teaching scenes, indicating that visual cues can help learners perceive the state of music movement more intuitively [4].

At the same time, artificial intelligence technology is pushing music research from analysis to generation, recommendation and adaptive support. Mycka and Mańdziuk summarized the latest progress of artificial intelligence in the field of music, pointing out that intelligent algorithms have been deeply involved in music creation, recognition, classification and user interaction [5]. Yakura et al. built an automatic background music recommendation system in work scenes, which reflects that the music experience is shifting from static playback to context-oriented personalized matching [6]. Grekow used recurrent neural network and pre-training model to carry out music emotion recognition research, indicating that the computational model has been able to capture music emotion characteristics more stably [7].

At the perceptual and aesthetic level, previous studies have begun to pay attention to the influence of visual, auditory and multi-sensory coordination on music experience. Talamini et al. compared the differences in auditory and visual imagery between music professionals and

non-professionals, and revealed the correlation between music understanding and mental representation ability [8]. Starting from the singing situation, Lange et al. verified the role of multi-sensory integration in music emotion perception [9]. Franěk and Petružalek discussed the audio-visual interaction between music and natural environment, and further explained that the visual field would affect the formation of music experience [10]. In terms of educational application, Liu et al. pointed out through systematic review and social network analysis that music education research supported by mobile technology is continuously growing, and interactivity, portability and media integration have become important trends [11]. Wang's research shows that different gamification material designs can significantly affect music learning effects, indicating that the form of technical media itself has become an important variable in shaping music experience [12].

In general, the existing research has provided rich results for music visualization, intelligent recommendation, emotion recognition and multimodal music education. However, most of the work focuses on technical performance, tool application or learning effectiveness, and there is still a lack of systematic explanation of how technology can improve the music aesthetic experience by constructing listeners 'perceptual expectations and perceptual organization. In particular, there are still few studies that explain the mechanism of modern science and technology from the perspective of "perceptual style presupposition", and incorporate music atlas, music animation, performance video and dynamic staff into a unified analysis framework, which also constitutes the entry point for further discussion in this paper.

3 Music aesthetic experience mechanism and technical path design based on perceptual style presupposition construction

3.1 The concept definition of musical perceptual style presupposition and the generation mechanism of aesthetic experience

Music aesthetic experience requires listeners to actively integrate melody, rhythm, strength, timbre and structural cues in the perceptual process, and then form a perceptual style with a sense of integrity. The so-called musical perceptual style presupposition refers to the initial perceptual framework formed by the listener according to genre information, title hints, past experience and visual auxiliary cues before the listener formally grasp the overall shape of the work. The framework does not determine the entire content of the musical experience, but affects how listeners allocate attention, identify key points, and understand the relationship between the local and the whole. For works with clear structure and familiar style, this presupposition is often done unconsciously. In the face of works with complex polyphonic textures, frequent rhythm changes or unfamiliar styles, presupposition construction will directly affect the quality of subsequent aesthetic experience.

To facilitate the description of the preset formation process, the preset strength can be expressed as follows:

$$P_0=\sum_{i=1}^{n}\omega_i c_i \sum_{i=1}^{n}\omega_i=1 \tag{1}$$

Here, P_0 represents the initial preset strength, c_i represents the effective value of type i cue information, which mainly includes genre cue, title semantics, visual icon, performance action

cue and previous listening experience, ω_i is the weight of each cue.

Music aesthetic experience is a continuous chain composed of music information input-perceptual style presupposition formation -music elements organization and integration -overall style generation -aesthetic experience output. When listeners are exposed to music, they first obtain peripheral information such as title, performance scene, visual atlas and dynamic spectral surface, which will generate some expected contour before formal listening. When the sound really enters the perceptual system, the preset will participate in the recognition of melody direction, rhythm grouping, strength change judgment and structure level induction.

To further characterize the level of perceptual organization, the degree of organization of musical information can be written as follows:

$$G=\frac{1}{T}\int_0^T \frac{\rho(t)\cdot\sigma(t)}{1+\mu\kappa(t)}dt \tag{2}$$

Here, G represents the overall organization degree in unit time, ρ(t) represents the listener's perception level of music continuity, σ(t) represents the matching degree between the preset and the actual sound, κ(t) represents the processing load caused by the complexity of the work, μ is the complexity adjustment coefficient. On this basis, the aesthetic experience level can be expressed as follows:

$$E_a=\ln(1+P_0GI)-\delta\varepsilon \tag{3}$$

Here, Ea represents the level of aesthetic experience, I represents the degree of emotional involvement, ε represents the deviation between preset and actual perception, and δ is the deviation suppression coefficient. Equation (3) reveals that there is a coupling relationship among presupposition strength, perceptual organization and emotional involvement. When the three are promoted together, the music experience will be significantly enhanced, but if the preset is inaccurate and the deviation is too large, the overall feeling will be weakened. It can be seen that the reason why modern technology can help music aesthetics is that it can provide listeners with clearer structural clues and more stable expectations before and during formal listening, thus improving the integration efficiency of musical elements.

Thus, musical perceptual style presupposition can be seen as an important mediator connecting external technical support with internal aesthetic generation. At one end, it connects external auxiliary methods such as map, animation, performance video and dynamic staff, and at the other end, it enters the perceptual organization process of the listener, which ultimately affects the integrity, hierarchy and emotional depth of the aesthetic experience. Figure 1 shows the generation mechanism of musical aesthetic experience.

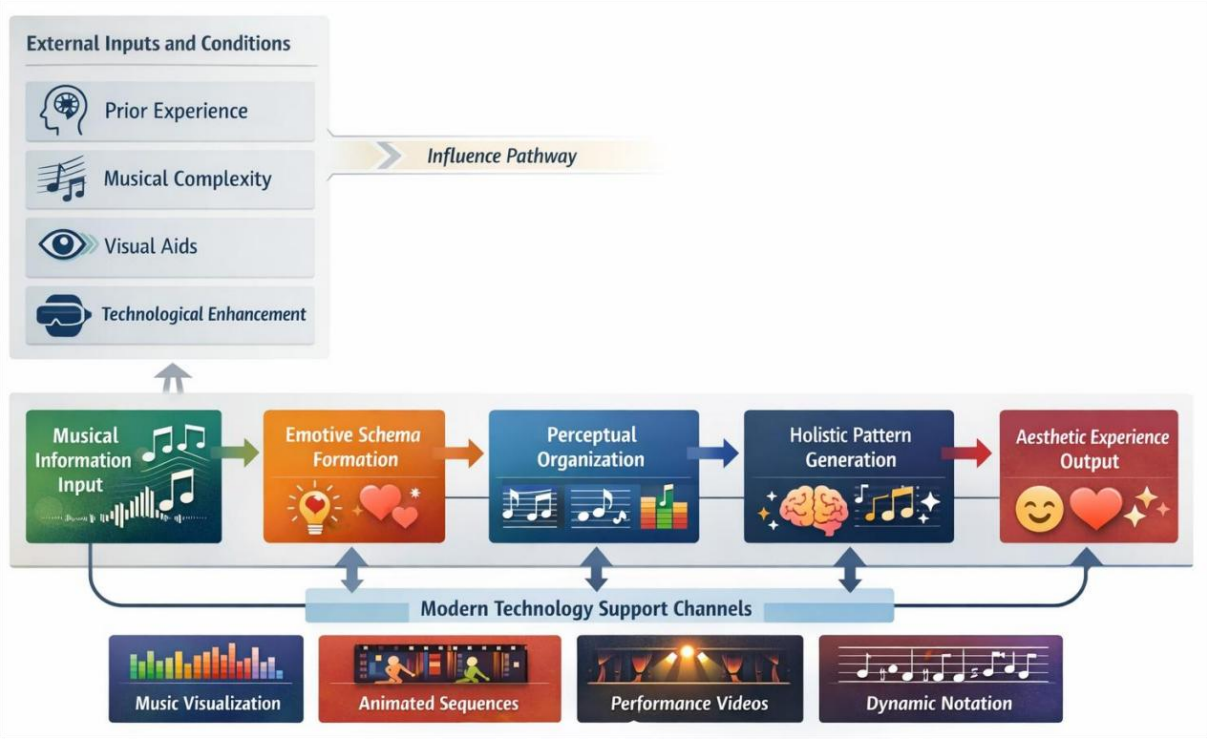

Figure 1: Diagram of musical aesthetic experience generation mechanism

3.2 Analysis of music aesthetic processing mechanism based on Gestalt perceptual organization

According to Gestalt perception theory, when receiving external information, subjects usually do not stop at identifying isolated elements one by one, but spontaneously integrate scattered information into an overall structure with a sense of order according to the principles of proximity, similarity, continuity and closure. This theory is used to explain the aesthetic process of music and has strong adaptability. After music enters the auditory system, information such as pitch, rhythm, strength, timbre, and voice level are not naturally presented in the consciousness in the form of "complete works". Listeners still need to gradually form an overall grasp of phrases, paragraphs, theme relationships and emotional trends through active organization. An analogy between visual perception and auditory perception is shown in Figure 2.

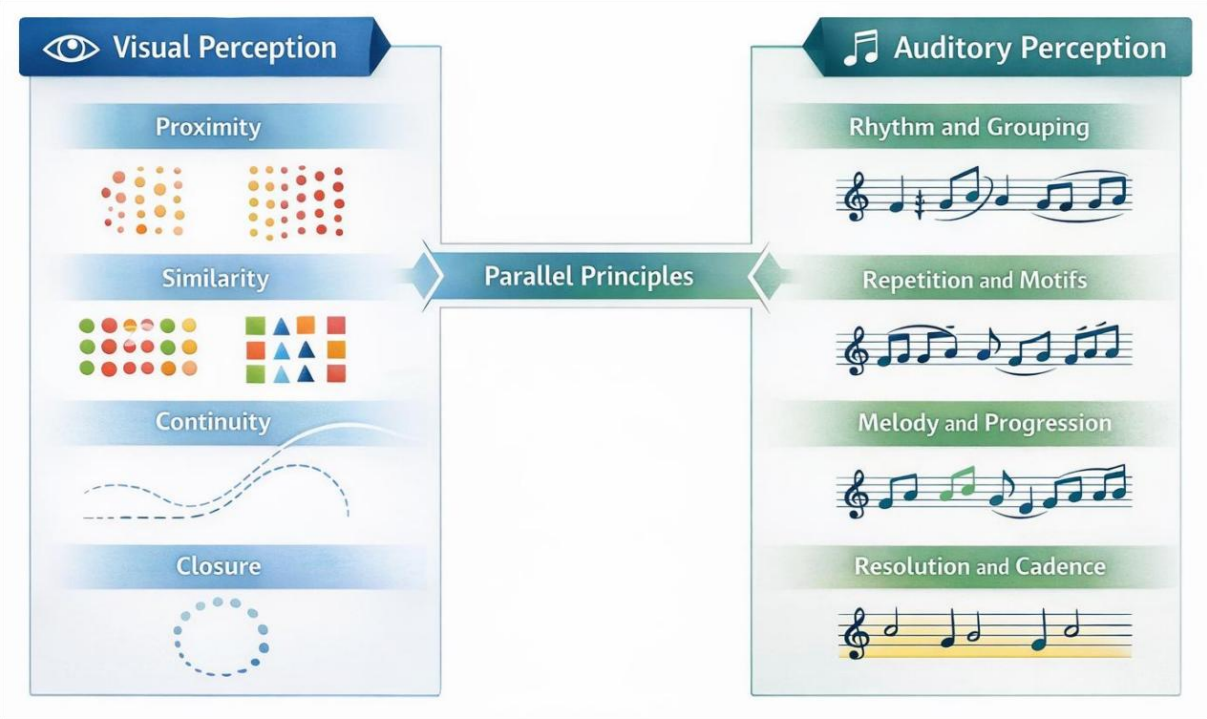

Figure 2: Schematic diagram of the analogy between visual and auditory perception

In visual perception, the reason why a number of points, lines and block surfaces can be recognized as a kind of figure depends on whether the subject captures the internal relationship between each visual unit. In music perception, the reason why a number of instantly occurring sounds can be understood as melody, motivation, texture or structural levels also depends on whether the subject can detect organizational cues between adjacent sounds. Although the two kinds of perception rely on space and time respectively, they have remarkable consistency in processing mode. Proximity aggregation in vision can correspond to rhythm groups and short clauses in music. The morphology in vision is similar, which can correspond to motivation repetition, timbre reproduction and timbre classification in music. The direction of vision is continuous, which can correspond to melody extension, harmony advancement and voice part movement. The tendency to close in vision can correspond to terminations, paragraph regression, and tension release. It is precisely because of this structural connectivity that visual AIDS may assume the function of pre-organization and pre-cue in music appreciation.

In order to characterize the degree of correspondence between visual and auditory cues, the cross-channel tissue mapping value can be defined as follows:

$$K=\sum_{i=1}^{n} \omega_i \cdot r_i \tag{4}$$

where K represents the overall mapping level between visual information and music information, r_i represents the matching degree of type i correspondence, covering dimensions such as direction, density, level, strength and rhythm contour, ω_i represents the weight of each dimension, and satisfies $\sum_{i=1}^{n} \omega_i=1$,when the visual diagram and the internal structure of music show high consistency in multiple dimensions. The listener will build up the overall impression faster, and the uncertainty in subsequent processing will be reduced.

On this basis, the degree of aggregation of perceptual organization can also be expressed

as follows:

$$O=\frac{1}{m}\sum_{j=1}^{m}\phi_j\cdot(a_j+b_j+c_j+d_j) \tag{5}$$

Here, O represents the intensity of perceptual organization, a_j, b_j, c_j, d_j represent the emergence levels of proximity, similarity, continuity and closure in cue group j, respectively, and ϕ_j is the importance coefficient of different cue groups. The clearer the structure of the work is and the more stable the cue is, the higher the degree of aggregation of perceptual organization is, and the easier the listener's grasp of the music style is to become complete.

However, the complexity of a musical piece continuously affects the organizational process. In the face of works with dense rhythm, complex voice parts, tonal drift or rapid material development, listeners need to invest more attention resources to maintain perceptual stability in the dynamic flow. The processing load under this condition can be written as follows:

$$L=\gamma_1 x+\gamma_2 y+\gamma_3 z \tag{6}$$

Here, L represents perceptual processing load, x represents voice part complexity, y represents rhythm change density, and z represents structural turning frequency. This equation reflects that the more complex the musical material is, the more stress is placed on the auditory organization. Without sufficient external cues, listeners are easy to fall into the accumulation of local information, and it is difficult to refine the overall outline, which will limit the coherence and depth of the music experience.

The intervention of modern scientific and technological means can just play a regulatory role in this node. Music graph can transform melody fluctuation, intensity change and time advance into intuitive graphics. Music animation can indicate the tension trend through morphological expansion, direction flow and speed change. The conductor movement, performance posture and facial expression in the performance video can strengthen the musical turning and emotional trend. Dynamic staff can provide listeners with a more explicit structural prediction by virtue of the note position, rhythm arrangement and syntactic boundaries that appear in advance. After the intervention of technology, the perceptual system obtains additional support, and the overall processing efficiency can be further expressed as follows:

$$E=\frac{\alpha K+\beta O+\mu S}{1+L} \tag{7}$$

Here, *E* represents the efficiency of music aesthetic processing, and S represents the degree of stability of technical auxiliary cues. It can be seen that the Gestalt perceptual organization theory provides a solid theoretical fulcrum for explaining the aesthetic processing of music. In the process of music acceptance, listeners always need to use some structural principles to integrate the continuous flow of sound in time into an overall style that can be understood, experienced and remembered. The value of modern technology is that it can transform the original implicit organizational relationship into a visible, sensible and predictable cue system, helping listeners to identify the internal order of music more quickly and complete the transition from local sound to overall perceptual style more steadily. The following sections will further combine music graph, music animation, performance video and dynamic staff to analyze how these technical paths participate in the construction of

aesthetic presupposition in concrete practice.

3.3 Aesthetic presupposition construction method supported by music atlas

The value of music atlas is to transform the melody fluctuation, rhythm density, strength change and structural level, which are mainly dependent on auditory grasp, into visual cues that can be directly viewed. Before the listener formally enters into the deep feeling of the work, the listener first obtains a certain overall impression through the graphic outline, and then brings this impression into the listening process, and the organization efficiency of the musical elements will be improved. For the works with clear structure, the atlas can strengthen the existing perception. For works with dense texture and rapid development, atlas helps to compress the threshold of understanding, so that listeners can capture the trend of the theme, the focus of the paragraph and the change of tension earlier. The construction process of aesthetic presupposition supported by music atlas is shown in Figure 3.

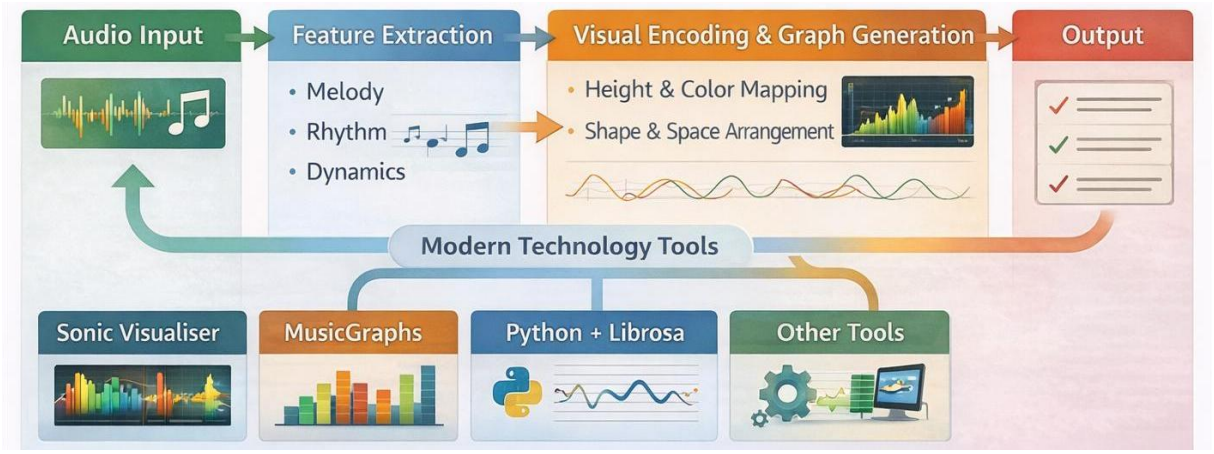

Figure 3: Flow chart of aesthetic presupposition construction supported by music atlas

From the perspective of technical implementation, the generation of music graph first relies on the decomposition and feature extraction of audio information. The system needs to identify pitch contour, rhythmic density, intensity variation, timing distribution, and syntactic boundaries from continuous sound waves, and then map these sound parameters into visual attributes such as position, length, color, thickness, and spatial density. The comprehensive musical characteristics at a certain moment can be expressed as follows.

$$Q(t)=\alpha_1 p(t)+\alpha_2 r(t)+\alpha_3 d(t)+\alpha_4 n(t) \tag{8}$$

Here, Q(t) represents the comprehensive musical representation at time t, p(t) represents the pitch trend, r(t) represents the rhythm density, d(t) represents the strength, and n(t) represents the timbre saliency.

After the feature extraction is completed, the mapping rules from sound to vision need to be established. The ascending melody can correspond to the ascending graphics, the rhythm encryption can correspond to the denser arrangement of points and lines, the strength enhancement can correspond to the deepening of colors or the enlargement of graphics, and the turning point of paragraphs can correspond to the change of visual partitions. If we write the visual encoding as follows:

$$Y(t)=\beta_1 h(t)+\beta_2 c(t)+\beta_3 s(t)+\beta_4 g(t) \tag{9}$$

Here, Y(t) represents the visual expression intensity at time t, h(t) represents the height change of the graph, c(t) represents the color change, s(t) represents the size change of the graph, and g(t) represents the spatial distribution characteristics. Through this process, musical relationships that originally flow through time are compressed into visual structures that can be viewed simultaneously, and listeners are able to form a sense of overall contour more quickly.

After the atlas enters the stage of aesthetic presupposition construction, its function focuses on "advance suggestion". When faced with an atlas, listeners first form preliminary judgments about the movement of the music, such as how the melody will unfold, whether the tempo will tend to be tense, where the strength will change, and where the passage is likely to turn. The preset activation level elicited by the atlas can be expressed as follows:

$$W=\frac{1}{T}\int_0^T \left[\mu_1 u(t)+\mu_2 v(t)+\mu_3 z(t)\right] dt \tag{10}$$

Here, W represents the preset activation level, u(t) represents the melody contour clarity, v(t) represents the rhythmic structure discrimination, and z(t) represents the paragraph boundary saliency. The clearer the map is, the more stable the visual cue is, and the easier it is for the listener to establish a more complete perceptual expectation before listening.

Whether music atlas can truly serve aesthetic experience also depends on the degree of consistency between visual expression and real sound. To this end, the matching level between the atlas and the sound can be further written as follows.

$$Z=\frac{\sum_{k=1}^{m} \min(a_k,b_k)}{\sum_{k=1}^{m} \max(a_k,b_k)} \tag{11}$$

Here, Z represents the atlase-sound matching degree, a_k represents the amount of structural information of the KTH visual unit, and b_k represents the amount of structural information of the corresponding sound unit. The closer Z is to 1, the more closely the atlas fits the internal relationship of the music, and the more reliable the presupposition formed by the listener through the atlas.

Therefore, the key to the construction of aesthetic presupposition supported by music atlas is to extract core elements, establish effective mapping, control visual load, and make the movement trend presented by the atlas consistent with the internal organization of music. The listener can quickly acquire the sense of direction, hierarchy and paragraph when entering the work, and the perceptual organization in the subsequent listening is easier to stabilize and expand.

3.4 Aesthetic guidance path supported by music animation and performance video

Music animation and performance video both belong to the path of visual assistance, but the way of entering the aesthetic process of music is not the same. Music animation is more suitable for dealing with abstract structure. It can translate melody ups and down, rhythm advance, strength fluctuation and paragraph turning into tracable motion trajectories, shape changes and color flow, so that listeners can obtain clear dynamic contours before and after listening. The performance video emphasizes more on body movement, expression information and scene atmosphere. The rise and fall of the conductor, the bow movement and key stroke of the player, and the capture of local details by the camera will all have an impact on the audience's emotional entry, attention stay and structure judgment.

From the side of music animation, its core is to transform continuous sound into visual events with a sense of direction and speed. If the rhythm of the animation and the rhythm of the music are consistent with each other, and the upward direction of the melody is consistent with the movement direction of the picture, it is easier for the listener to grasp the development trend of the music. We can write the animation hint strength as follows:

$$A_t=\rho_1 m_t+\rho_2 q_t+\rho_3 l_t+\rho_4 c_t \tag{12}$$

Here, A_t represents the animation prompt intensity at time t, m_t represents the screen motion speed, q_t represents the direction change amplitude, l_t represents the brightness change level, ct represents the color contrast intensity, and ρ_1 to ρ_4 are the weights. The guiding force of animation comes from the joint effect of multiple visual variables. Too strong a single dimension will cause perception deviation, and the overall coordination is more important.

The mechanism of performance video is different from this, and the cues transmitted by performance video are closer to the embodied experience. Viewers can perceive the tension changes of the music from the command gesture, performance posture, breathing state and camera focus. The guided saliency of performance video can be expressed as:

$$V_t=\mu_1 g_t+\mu_2 e_t+\mu_3 k_t+\mu_4 p_t \tag{13}$$

Here,V_t represents the video-guided saliency at time t, g_t represents the amplitude of body movements, e_t represents the intensity of expression changes, k_t represents the effectiveness of lens focus, and p_t represents the correspondence between the performed movements and the musical structure. This path can enhance the sense of presence and help listeners identify paragraph turns and emotional peaks faster.

When two types of paths form synergy, the key variable lies in the time consistency. Animation if too much in advance, easy to disconnect with the actual sound; If the camera switch lag is obvious, the listener's judgment of structural relationship will be weakened. The timing synchronization level can be defined as follows:

$$S=\frac{1}{T}\int_0^T\left(1-\frac{|x_t-y_t|}{\delta}\right)dt \tag{14}$$

Here, S represents the synchronization level between the animation cue and the video cue in the whole time period, x_t represents the key change moment of the animation, y_t represents the key cue moment of the video, and δ represents the allowable deviation threshold. The higher S is, the more consistent the two visual pathways are in their cues to the musical structure, and the more stable the expectations formed by the listener.

On the basis of the synchronization relationship, it is still necessary to determine whether the two types of paths are really complementary. Animation is responsible for refining the abstract outline, and video is responsible for supplementing the action and situation. This degree of complementarity can be expressed as:

$$C=\omega_1 A+\omega_2 V+\omega_3 S+\omega_4 M \tag{15}$$

where C represents the cross-path complementary level, A represents the overall animation cue value, V represents the overall video cue value, S represents the synchronization level, and M represents the common pointing degree of the two to the center of gravity of the music structure.

Collaborative guidance generally does not automatically lead to better aesthetic results,

and cognitive load still needs to be included in the investigation. Too many screen elements, too fast camera changes, and too dense animation movement will squeeze the listener's attention to the music itself. The aggregate load can be written as follows:

$$D=\phi_1 n+\phi_2 r+\phi_3 f+\phi_4 u \tag{16}$$

where D represents the cognitive load level, n represents the number of visual elements, r represents the frequency of shot switching, *f* represents the frequency of animation change, and u represents the amount of prompts to be processed per unit time. When the load exceeds a certain threshold, visual AIDS will transform from support conditions to interference factors, and listeners are prone to aesthetic fatigue and attention distraction.

Based on this, the emotional resonance level can be further written as follows:

$$R=\frac{1}{T}\int_0^T [\eta_1 z_t+\eta_2 h_t+\eta_3 b_t]dt \tag{17}$$

Among them, R represents the emotional resonance intensity caused by multi-path, zt represents the degree of fit between animation emotional color and music emotion, ht represents the infection intensity of performance movement, and bt represents the listener's subjective following level of music emotional changes. If animation and video can jointly promote emotional entry, listeners will be more likely to form sustained emotional engagement beyond structural understanding.

Finally, the effectiveness of aesthetic guidance supported by multi-path collaboration can be expressed as follows:

$$E=\frac{\alpha_1 C+\alpha_2 R+\alpha_3 P}{1+D} \tag{18}$$

Here, E represents the comprehensive efficacy of aesthetic guidance, C represents the level of cross-path complementarity, R represents the intensity of emotional resonance, P represents the stability of the listener's grasp of the music structure, D represents the cognitive load, and α1 to α3 are the weights.

Therefore, the co-design of music animation and performance video should focus on "abstract contour prompt - embodied action reinforcement - timing synchronization control - cognitive load inhibition". For the works with distinct rhythm and clear structure promotion, the participation of the animation path can be appropriately increased to strengthen the overall movement trend. For works with strong emotional tension and rich performance movements, the weight of the performance video can be increased to enhance the sense of presence and emotional appeal.

3.5 Music experience optimization method of dynamic staff and intelligent interactive presentation

The advantage of dynamic staff is that it transforms the original static surface into a visual interface that keeps in line with the progress of the music, so that the listener can get hints of pitch direction, rhythm organization and syntactic boundaries before the sound arrives. In addition, intelligent interactive rendering adds functions such as highlighting, positioning, hierarchical display, local playback and personalized prompts, making the spectrum become an active medium for participating in the organization of music experience. Dynamic staff can play an effective role, one of the key lies in whether the visual cue and the real sound to

maintain a moderate "advance". If the prompt is too early, the listener is easy to produce a sense of waiting; If the prompt is too late, the spectrum will lose its preset guiding significance. The advance display intensity of the spectral surface can be expressed as follows:

$$B=\frac{1}{N}\sum_{i=1}^{N}(t_i^a-t_i^v) \tag{19}$$

Here, B represents the average amount of advance, t_i^a represents the actual sound moment of the ith note, and t_i^v represents the moment when the note is revealed in the interface in advance. When the value is in a reasonable range, it is easier for listeners to establish a stable connection between visual preparation and auditory reception.

The readability of the spectral surface itself also directly affects the experience optimization effect. Disordered note layout, unbalanced scrolling speed, or too wide highlighting range can impair perceived efficiency. The spectral readability can be written as follows:

$$U=\kappa_1 h+\kappa_2 s+\kappa_3 q+\kappa_4 w \tag{20}$$

Here, U represents spectral surface legibility, h represents pitch trajectory clarity, s represents rhythm spacing equalization, q represents syntactic boundary saliency, w represents the control level of the highlighting window, and κ_1 to κ_4 are the weights. The higher the readability, the more stable the dynamic staff can assume the preset prompt function.

The value of intelligent interaction is reflected in the adaptation to the needs of different audiences. Users can actively adjust their own viewing rhythm by pausing, zooming in, looping, switching parts and highlighting. Interactive response efficiency can be expressed as follows:

$$J=\frac{\lambda_1 c+\lambda_2 f+\lambda_3 p}{1+\lambda_4 d} \tag{21}$$

where J represents the interaction efficiency, c represents the operation accuracy, f represents the feedback timeliness, p represents the personalized matching degree, and d represents the interaction delay. The more fluid the interaction, the easier it is for listeners to convert visual cues into stable structural judgments.

On this basis, the activation degree of dynamic staff to aesthetic presupposition can be further written as follows:

$$E=\frac{1}{T}\int_0^T\left[\mu_1 r(t)+\mu_2 g(t)+\mu_3 m(t)\right]dt \tag{22}$$

Here, E represents the preset activation level, r(t) represents the strength of melodic direction cue, g(t) represents the strength of rhythm organization cue, and m(t) represents the strength of local structure labeling. The higher the value, the easier it is for listeners to form a clear sense of direction and paragraph when entering the work.

In order to measure the optimization effect of intelligent interactive presentation on the overall music experience, the comprehensive effectiveness can be expressed as follows:

$$O=\frac{\alpha_1 B+\alpha_2 U+\alpha_3 J+\alpha_4 E}{1+\alpha_5 L} \tag{23}$$

Here, O denotes the experience optimization effectiveness, L denotes the additional cognitive load, and α_1 to α_5 are coefficients. Furthermore, the stability of the listener's experience during continuous listening can be written as follows:

$$M=\beta_1 O+\beta_2 a+\beta_3 z \tag{24}$$

Here, M represents the stability of musical experience, a represents the sustained attention level, z represents the consistency of structural grasp. If dynamic staff and intelligent interaction can jointly enhance these three dimensions, the sense of hierarchy, coherence and participation of music experience will be significantly enhanced.

The functional configuration of dynamic staff and intelligent interactive rendering is shown in Table 1.

Table 1: Dynamic staff and intelligent interactive rendering function configuration

Module	Core Function	Recommended Parameters	Presentation Mode	Experiential Role
Advance Display Module	Pre-display of upcoming notes	0.8–1.2 s	Rolling display ahead of the current note	Helps form listening expectations before sound onset
Highlight Tracking Module	Localization of the current voice part and main melody	Highlight window of 1–2 measures	Color highlighting + vertical line positioning	Strengthens structural attention
Rhythm Guidance Module	Beat and grouping cues	2–4 beat scrolling unit	Beat markers + rhythmic segmentation	Enhances rhythmic organization
Layered Display Module	Separation of main melody and accompaniment	2–3 display layers	Differentiation by color or transparency	Reduces the burden of complex texture processing
Interactive Control Module	Pause, replay, and loop	Response delay <120 ms	Button control + local dragging	Improves operational fluency
Personalized Adaptation Module	Differentiated settings for different audiences	Font size 110%–160%	Difficulty switching + speed adjustment	Improves adaptability and usability

4 Experimental design and result analysis

4.1 Data source and experimental environment configuration

In order to verify the promotion effect of perceptual style presupposition in music aesthetic experience, this study selected 48 experimental samples from classical music appreciation

materials, covering four types of polytone, singing, dance and lyric works, and controlled the duration of a single piece within 45-90 s, so as to ensure that subjects could complete continuous judgments and subjective ratings within a limited time. The sample audio is normalized by unified loudness, and the video and dynamic atlas material are matched synchronously under the same time axis to avoid the influence of volume differences, screen delays or resolution fluctuations on the experimental results. There were 96 subjects, including 48 music professional group and 48 non-music professional group. The age distribution was concentrated in 19-26 years old, and all of them had normal hearing and basic audio-visual understanding ability. The experiment was completed in a quiet indoor environment, and the background noise was controlled below 35 dB. The playing terminal adopted a 24-inch display and closed monitoring headphones to ensure the stable and consistent visual presentation and sound output. The experimental software integrates four functions: audio playback, dynamic map display, interactive recording and questionnaire scoring, which can synchronously collect the subjects 'stay time, scoring results and operation trajectories under different conditions. To reduce the order effect, the sequence of sample playback was randomized, and short intervals were set between different technical conditions to alleviate perceptual fatigue. The experimental samples and environment configurations are shown in Table 2. On the whole, this study maintained a good balance in sample types, subject composition and equipment conditions, which could provide a relatively stable data basis for subsequent analysis of results.

Table 2: Experimental samples and experimental environment configuration

Item	Configuration	Detailed Description
Number of Music Samples	48 excerpts	Each excerpt lasted 45–90 s and covered four categories: polyphonic, sonata, dance, and lyrical works
Types of Presentation Materials	4 types	Pure audio, music visualization, music animation/performance video, and dynamic staff notation
Number of Participants	96	48 music majors and 48 non-music majors
Age Range	19–26 years	All participants had normal hearing and no obvious audiovisual impairments
Experimental Environment	Quiet laboratory	Background noise was kept below 35 dB with stable lighting conditions
Display Equipment	24-inch monitor	Resolution of 1920×1080 and refresh rate of 60 Hz
Audio Equipment	Closed-back monitoring headphones	Sampling rate of 44.1 kHz with standardized loudness output
Experimental Software	Self-developed interactive platform	Supported playback control, synchronized display, operation recording, and rating collection
Recorded Data	3 categories	Subjective ratings, viewing/listening duration, and interaction behavior trajectories
Presentation Order Control	Randomized	Materials were played in random order with 10 s intervals to reduce fatigue

4.2 Effectiveness analysis of perceptual style presupposition construction

In order to test the promotion effect of perceptual style presupposition on the aesthetic experience of music, this paper compares the performance of subjects under four conditions: no presupposition, basic presupposition, visual presupposition and comprehensive presupposition. Evaluation dimensions included structural understanding, emotional engagement, auditory organization, and composite experience scores. The statistical results of the effectiveness of perceptual style presupposition construction are shown in Table 3. The results show that as the preset information is gradually enriched, the performance of the subjects in all indicators is on the rise. Without preset conditions, listeners rely more on immediate auditory processing, and their comprehension speed is slow and their emotional entry is relatively delayed when facing the segments with complex structures. After adding the basic presupposition, the subjects had a preliminary grasp of the overall outline of the work. Visual presupposition further strengthens the recognition of paragraph boundary, melody direction and rhythm grouping. Under the comprehensive presupposition condition, listeners can form stable expectations faster when entering the work, and the subsequent aesthetic judgment is more coherent.

Table 3: Statistical results of effectiveness of perceptual style presupposition construction

Experimental Condition	Structural Understanding Score	Emotional Engagement Score	Auditory Organization Score	Overall Experience Score
No Presupposition	68.4	71.2	69.1	70.3
Basic Presupposition	74.6	76.8	75.2	75.5
Visual Presupposition	81.3	82.6	80.4	81.5
Integrated Presupposition	86.7	85.9	84.8	85.8

Table 3 shows that the comprehensive experience score under the comprehensive preset condition reaches 85.8 points, which is 15.5 points higher than that without the preset condition. The structural comprehension score increased from 68.4 to 86.7, with an increase of 18.3. The score of auditory organization increased by 15.7 points. This shows that perceptual style presupposition can significantly improve the listener's grasp of the internal relationship of music, and enhance the integrity and coherence of the aesthetic experience.

4.3 Analysis on the improvement effect of music atlas on music aesthetic understanding

Music atlas converts melody ups and down, rhythm density and intensity changes into visible clues, which can provide subjects with a more clear structure outline before the beginning of listening, and can also continue to strengthen their judgment of paragraph hierarchy and movement trend during listening. To investigate this role, this paper compared the no atlas condition, the melody atlas condition, the rhythm atlas condition, and the strength atlas condition with the comprehensive atlas condition, and calculated the score changes in the aesthetic understanding dimension between the music professional group and the non-music professional group, respectively. In the non-atlas condition, subjects mainly relied on immediate auditory processing, and there was a certain lag in structure judgment when facing segments with more layers. After adding the melody map, the melody direction was more intuitive, and subjects were easier to form an overall sense of line. The rhythm map and the strength map continue to supplement the information of time organization and tension change, which further improves the aesthetic understanding. The integrated graph integrates

multi-dimensional cues in the same interface, which significantly enhances the subjects' ability to grasp the internal relationship of music. As shown in Figure 4, the music atlas has a relatively obvious effect on improving the aesthetic understanding of music.

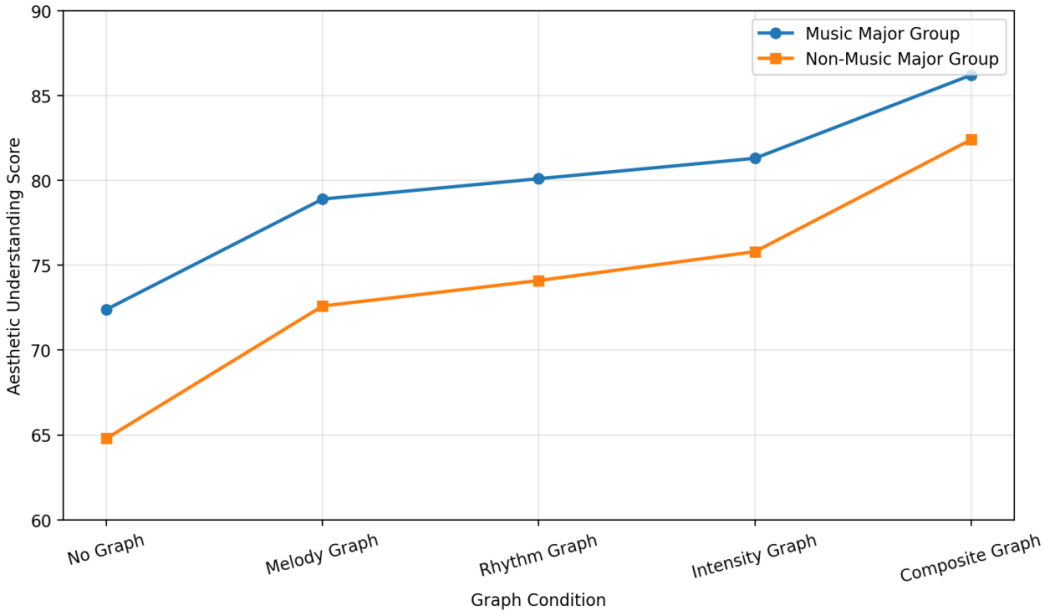

Figure 4: Effect of Music Visualization on Aesthetic Understanding

From the results, the score of aesthetic understanding of the music professional group increased from 72.4 points in the condition of no atlas to 86.2 points in the condition of comprehensive atlas, with an increase of 13.8 points. The non-music professional group increased from 64.8 to 82.4, an increase of 17.6 points. The gap between the two groups in the comprehensive atlas condition was reduced to 3.8 points, indicating that the auxiliary role of music atlas for non-professional audiences was more prominent, and it was also verified that the atlas had a strong positive effect in reducing the threshold of understanding complex works and enhancing structure identification and overall perception.

4.4 Analysis of aesthetic guidance effect of music animation and performance video

Both music animation and performance video can provide intuitive visual clues for listeners, but there are differences in the emphasis of aesthetic guidance between them. Music animation is more likely to highlight melodic movement, rhythm advancement and tension changes, while performance video is more likely to strengthen performance movements, emotional contagion and live atmosphere. In order to compare the two types of paths and their synergies, this paper sets up four conditions: pure audio, audio + music animation, audio + performance video, audio + music animation and performance video collaboration, and makes statistics from three dimensions: attention maintenance, emotional engagement and structural guidance. The aesthetic guidance effect of music animation and performance video is shown in Figure 5.

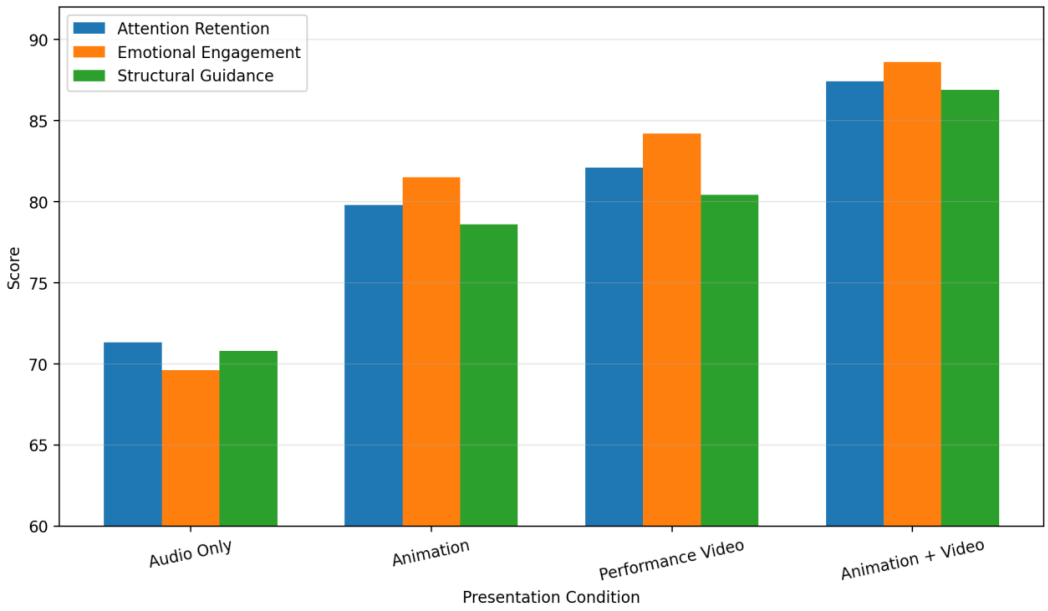

Figure 5: Aesthetic Guidance Effects of Animation and Performance Video

Figure 5 shows that the collaborative condition achieved the highest scores in all three indicators, with attention retention reaching 87.4, emotional engagement reaching 88.6, and structural guidance reaching 86.9, all significantly higher than the audio-only condition. Compared with pure audio, the music animation condition increased 11.9 points in emotional engagement, and the performance video condition increased 9.6 points in structural guidance, indicating that the former was more conducive to activating dynamic feelings, and the latter was more conducive to helping listeners enter the work context and identify key turns.

Further examining the changes in different listening stages, the dynamic advantages of multipath collaboration are more obvious. The variation of the guidance effect of different paths in each listening stage is shown in Figure 6.

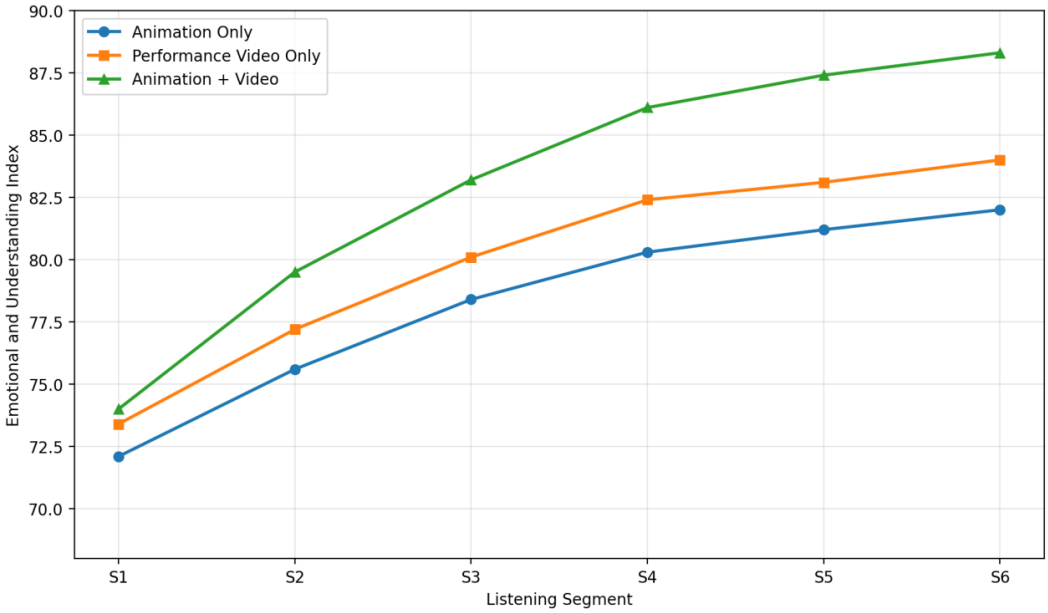

Figure 6: Dynamic Changes in Guidance Effects Across Listening Segments

Figure 6 shows that as the music progresses, the guidance index of the three visual conditions increases. Among them, the collaborative condition always maintains the highest level and reaches 88.3 in the sixth stage, which is 6.3 points higher than that of the animation condition alone and 4.3 points higher than that of the performance video condition alone. It shows that music animation and performance video can enhance the perception of structural cues and the coherence of emotional experience at the same time, so that listeners can maintain a more stable state of aesthetic participation in continuous listening. On the whole, the collaborative path promotes the aesthetic guidance most obviously, which not only improves the efficiency of understanding, but also enhances the depth of experience.

4.5 Analysis on the promotion effect of dynamic staff on aesthetic expectation and auditory organization

By presenting the pitch direction, rhythm grouping and syntactic boundaries in advance, dynamic staff provides support for listeners to establish clearer aesthetic expectations, and also enhances the ability to organize the internal relations of music. In order to compare the effects of different presentation methods, this paper sets up three conditions: static music score, dynamic music score and dynamic music score plus interactive prompt, and makes statistics on the two dimensions of aesthetic expectation and auditory organization. The effect of dynamic staff on the promotion of aesthetic expectations and auditory organization is shown in Figure 7.

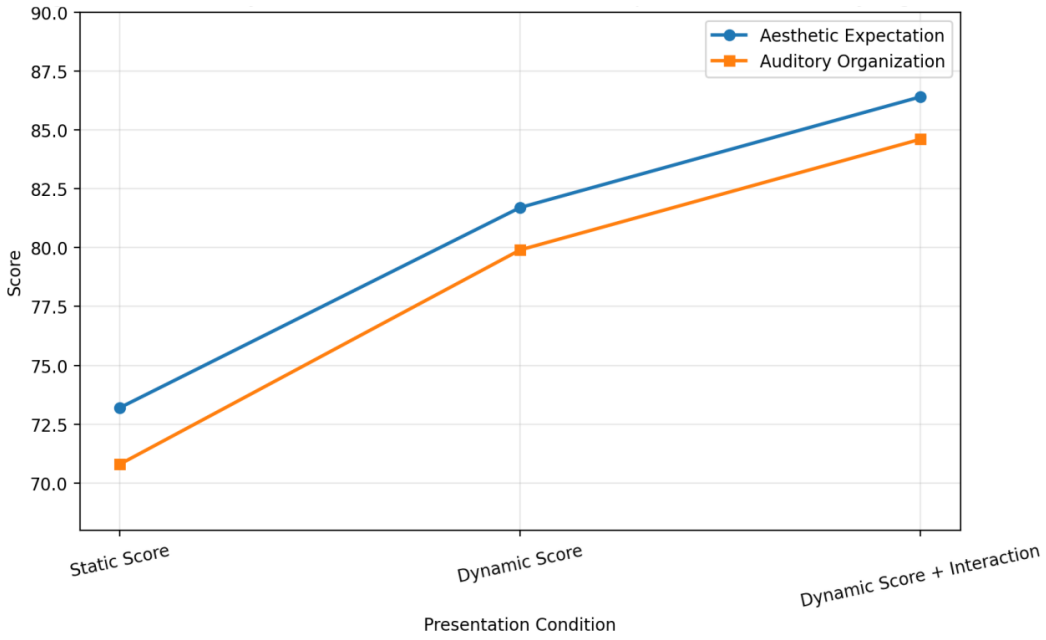

Figure 7: Effects of Dynamic Staff Notation on Aesthetic Expectation and Auditory Organization

Figure 7 shows that both metrics continue to improve with the presentation mode optimization. The score of aesthetic expectation increased from 73.2 in the static music score condition to 81.7 in the dynamic music score condition, and further increased to 86.4 after adding interactive prompts. The score of auditory organization increased from 70.8 to 79.9 and 84.6. The results show that dynamic staff can effectively help listeners to predict the direction of music, and the addition of interactive cues further strengthens their grasp of the

relationship between rhythm hierarchy and structure. On the whole, dynamic staff combined with intelligent interaction has a more obvious supporting effect on music experience.

4.6 Comparative experiments and comprehensive performance analysis of different technical paths

In order to further compare the actual performance of different technology paths in music aesthetic support, this paper incorporates music atlas, music animation, performance video, dynamic staff and collaborative path into a unified analysis framework, and comprehensively evaluates them from five dimensions: structural understanding, emotional engagement, aesthetic expectation, auditory organization and interactive friendliness. The comprehensive performance comparison of different technology paths is shown in Figure 8.

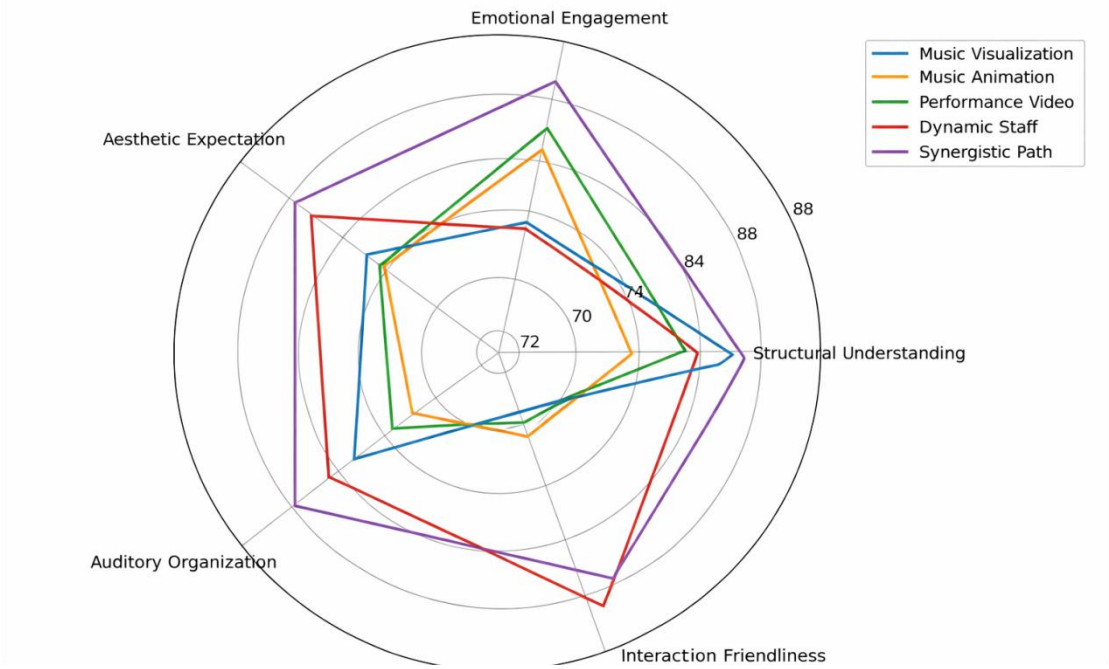

Figure 8: Comprehensive Performance Comparison of Different Technical Paths

Figure 8 shows that the advantage distribution of each technology path has some differences in different dimensions. Music atlas performs well in structure understanding, with a score of 86.2, and has a prominent role in sorting out the internal relationships of complex works. Music animation and performance video have obvious advantages in the dimension of emotional engagement, with scores of 84.7 and 86.1, respectively, which are easier to improve listeners' sense of entry and infection. Dynamic staff has a strong performance in aesthetic expectation and interactive friendliness, in which the aesthetic expectation score is 86.4, and the interactive friendliness score reaches 88.1, indicating that it has high adaptability in advance prompt and interface operation support.

The collaborative path maintains a high level in all five dimensions, and the overall profile is the most balanced. Its structural understanding score was 88.7, emotional engagement score was 89.3, aesthetic expectation score was 87.8, and auditory organization score was 86.9, all of which were in the leading position of each path. Although it is slightly lower than the dynamic staff single path in terms of interaction friendliness, it still reaches 84.5 points, which is at a high level. The results show that the combined use of multi-technology paths can better integrate the structural cue advantages of atlas, the emotional guidance advantages of animation and video, and the expected construction

advantages of dynamic staff, so as to form a more complete aesthetic support effect.

On the whole, a single path often has obvious advantages in a local dimension, but it also has the problem of unbalanced function coverage. The advantages of collaborative path are mainly reflected in the integrity and stability. The average comprehensive score of collaborative path reaches 87.4 points, which is 6.1 points higher than that of music atlas, 8.0 points higher than that of music animation, 6.7 points higher than that of performance video, and 3.4 points higher than that of dynamic staff. It can be shown that in the task of music aesthetic experience optimization, the multi-path collaborative scheme is more conducive to improving the listener's overall grasp ability and continuous experience quality of the work.

5 Discussion

According to the experimental results, the promotion effect of modern technology on music aesthetic experience has clear hierarchical differences. The performance of music atlas in structure understanding indicates that when melody lines, rhythm groups and intensity changes are transformed into clear visual contours, it is easier for listeners to establish a grasp of the overall trend of the work, especially when entering more complex music materials, this kind of hint can effectively reduce the pressure of understanding at the initial stage. For non-music professional audiences, this advantage is more obvious, indicating that visual processing has a strong realistic value in reducing the threshold of appreciation.

The experimental results of music animation and performance video suggest that the improvement of aesthetic experience not only depends on the clear presentation of structural information, but also is closely related to emotional contagion and physical suggestion. The animation path enhances the sense of movement and tension of the music, and the performance video enhances the audience's emotional entry through the command gestures, performance movements and scene atmosphere. After the synergy of the two, attention maintenance, emotional engagement and structural guidance reached a higher level, indicating that there was a strong complementary relationship between multimodal cues. In this condition, the listener obtains not only the feeling of "understanding" the music, but also the experience of "entering" the music.

The dynamic staff results further illustrate that advance prompts play an important role in music aesthetics. After note direction, beat advance, and syntactic boundaries are presented in a dynamic manner, listeners are able to form aesthetic expectations earlier, and subsequent auditory organization is more stable. After the addition of interactive functions, the adaptability of this path is further improved, especially suitable for scenes such as teaching guide, online appreciation and hierarchical communication. However, it should also be noted that more technical intervention is not always better. Too much visual information, too fast interface changes, and too high frequency of prompts may squeeze the feeling space of hearing itself, making the audience pay too much attention to the screen rather than the internal relationship of the music.

Therefore, the key of modern technology to help music aesthetic is not to constantly increase external stimuli, but to establish a proper cue mechanism around the music itself. Technology should serve the formation of perceptual style presupposition, and help listeners grasp the structure, enter the emotion, and maintain a coherent experience. Future research can further refine the adaptation rules under different work types, different audience backgrounds and different media conditions, so as to make the design of technology path more targeted and operable.

6 Conclusion

This paper focuses on the mechanism and path of modern technology to help music aesthetic experience, and proposes that perceptual style presupposition is an important intermediary connecting external technical support and internal aesthetic generation. Music atlas has outstanding advantages in structure understanding, music animation and performance video are more helpful to enhance emotional engagement and attention maintenance, dynamic staff and intelligent interaction have better performance in aesthetic expectation construction and auditory organization stability. The experimental results show that the score of structural understanding is improved to 86.7 under the comprehensive precondition, and the collaborative path keeps ahead in the four indicators of structural understanding, emotional engagement, aesthetic expectation and auditory organization. It can be seen that the collaborative design of multi-technology paths around music ontology is helpful to improve the audience's overall grasp of the work, continuous experience quality and aesthetic acceptance depth.

Funding

This work was supported by the 2025 Guangdong Provincial Education Science Planning Project (Higher Education Specialization), titled "Research on Intervention Measures for Constructing 'Presupposition' in the Music Reception Domain" (Project No.: 2025GXJK0144).

References

[1] Lima H B, dos Santos C G R, Meiguins B S. A Survey of Music Visualization Techniques[J]. ACM Computing Surveys, 2021, 54(7): 1-29. DOI:10.1145/3461835.

[2] Georges P, Seckin A. Music information visualization and classical composers discovery: an application of network graphs, multidimensional scaling, and support vector machines[J]. Scientometrics, 2022, 127(5): 2277-2311. DOI:10.1007/s11192-022-04331-8.

[3] Itoh T, Nakano T, Fukayama S, Hamasaki M, Goto M. SingDistVis: interactive Overview+Detail visualization for F0 trajectories of numerous singers singing the same song[J]. Multimedia Tools and Applications, 2025, 84: 1057-1077. DOI:10.1007/s11042-024-18932-3.

[4] Park E J. Music dynamics visualization for music practice and education[J]. Multimedia Tools and Applications, 2025, 84: 36145-36161. DOI:10.1007/s11042-025-20637-0.

[5] Mycka J, Mańdziuk J. Artificial intelligence in music: recent trends and challenges[J]. Neural Computing and Applications, 2025, 37: 801-839. DOI:10.1007/s00521-024-10555-x.

[6] Yakura H, Nakano T, Goto M. An automated system recommending background music to listen to while working[J]. User Modeling and User-Adapted Interaction, 2022, 32: 355-388. DOI:10.1007/s11257-022-09325-y.

[7] Grekow J. Music emotion recognition using recurrent neural networks and pretrained models[J]. Journal of Intelligent Information Systems, 2021, 57: 531-546. DOI:10.1007/s10844-021-00658-5.

[8] Talamini F, Vigl J, Doerr E, Grassi M, Carretti B. Auditory and visual mental imagery in musicians and non-musicians[J]. Musicae Scientiae, 2023, 27(2): 428-441. DOI:10.1177/10298649211062724.

[9] Lange E B, Fünderich J, Grimm H. Multisensory integration of musical emotion perception in singing[J]. Psychological Research, 2022, 86: 2099-2114. DOI:10.1007/s00426-021-01637-9.

[10] Franěk M, Petružálek J. Audio-Visual Interactions Between Music and the Natural Environment: Self-Reported Assessments and Measures of Facial Expressions[J]. Music & Science, 2024, 7: 1-18. DOI:10.1177/20592043241291757.

[11] Liu C, Hwang G J, Tu Y F, Yin Y, Wang Y. Research advancement and foci of mobile technology-supported music education: a systematic review and social network analysis on 2008-2019 academic publications[J]. Interactive Learning Environments, 2023, 31(7): 4535-4554. DOI:10.1080/10494820.2021.1974890.

[12] Wang Y H. Exploring the effects of using various designs of game-based materials on music learning[J]. Interactive Learning Environments, 2023, 31(5): 2650-2664. DOI:10.1080/10494820.2021.1894182.

[13] Dasovich-Wilson J N, Thompson M, Saarikallio S. Exploring Music Video Experiences and Their Influence on Music Perception[J]. Music & Science, 2022, 5: 1-18. DOI:10.1177/20592043221117651.

[14] Dasovich-Wilson J N, Thompson M, Saarikallio S. The characteristics of music video experiences and their relationship to future listening outcomes[J]. Psychology of Music, 2025, 53(1): 36-54. DOI:10.1177/03057356231220943.

[15] Damjanovic L, Kawalec A. The role of music-induced emotions on recognition memory of filmed events[J]. Psychology of Music, 2022, 50(4): 1136-1151. DOI:10.1177/03057356211033344.

[16] Braun Janzen T, de Oliveira B, Ventorim Ferreira G, et al. The effect of background music on the aesthetic experience of a visual artwork in a naturalistic environment[J]. Psychology of Music, 2023, 51(1): 16-32. DOI:10.1177/03057356221079866.

[17] Hong Y J, Choi A, Lee C E, et al. Concurrent musical pitch height biases judgment of visual brightness[J]. Psychology of Music, 2025, 53(3): 492-502. DOI:10.1177/03057356231216950.

[18] Kuch M, Wöllner C. Effects of mobile music listening: Patterns of changes in focus of attention, environmental perception, and self-experience[J]. Musicae Scientiae, 2024, 28(4): 740-757. DOI:10.1177/10298649241249427.

[19] O'Donohue M, Lacherez P, Yamamoto N. Audiovisual spatial ventriloquism is reduced in musicians[J]. Hearing Research, 2023, 440: 1-10. DOI:10.1016/j.heares.2023.

108918.

[20] Wang L, Tang X, Wang A, et al. Musical training reduces the Colavita visual effect[J]. Psychology of Music, 2023, 51(2): 592-607. DOI:10.1177/03057356221108763.

[21] de Bruin L R, Merrick B. Innovation, inclusion and engagement in a university music course[J]. Technology, Pedagogy and Education, 2024, 33(3): 347-362. DOI:10.1080/1475939X.2024.2324327.

[22] Black P, Barton G. Digitally mediated collaboration and participation: composing 10,427 miles and 11 hours apart[J]. British Journal of Music Education, 2025, 42(1): 124-135. DOI:10.1017/S0265051724000226.

[23] Hoad C, Johnson H, Rogerson-Berry M, et al. Gender representation in undergraduate music technology education: case studies from Aotearoa/New Zealand[J]. British Journal of Music Education, 2025, 42(2): 245-261. DOI:10.1017/S0265051724000330.