

Music Perception: Exploring the intrapersonal and cross-cultural judgment on music and its application in audiovisual materials

Xinyi Wang

Received May 15, 2023

Accepted July 26, 2023

Electronic access October 15, 2023

Music perception has been the subject of much systematic investigation. Several studies have documented the use of music in eliciting emotions and generating interpretation and its application in audiovisual materials. In the current review of the literature, two topics will be explored: a) the process of music perception to evoke intrapersonal judgment in audiovisual material, and b) the cross-cultural studies of music perception. This review found evidence that first, the top-down processes of schemas and the bottom-down processes of musical regularities work together to evoke emotional arousal; Second, music may direct specific emotions and alter audiences' attention towards designed motives; Third, the congruency of audio and visual materials may impact audiences' intrapersonal judgment and perception, namely the moral judgment toward protagonists, the expectation of film continuation, and the false memory. Furthermore, the findings indicate that there are significant differences in terms of music judgments among Western and non-Western cultural groups, for instance, Asian, Balinese, and Hindustani cultural groups, even though familiar contexts and cultural universality in psychophysical cues exist.

Keywords Music Perception, Music-induced emotion, Audiovisual, Cross-cultural perception.

Introduction

While music is a part of many people's daily lives, a thorough review of music's impact on our emotion and perception is rarely discussed among the general population. Prior research, including Vitouch's study, revealed that intrapersonal judgment can be affected by the emotion indicated in the music¹. Cheerful music generated more optimistic expectations, whereas, music designed to evoke sadness generated negative expectations. Moreover, cross-cultural studies suggested that the interpretation of music fluctuates between cultures since there are tonal system differences regionally. Balkwill and Thompson's investigation into interpretations of music demonstrated that Western listeners have difficulties identifying peaceful emotions in Hindustani raga melodies due to typical psychophysical components in the music excerpt². However, Kessler et al. found commonalities in cross-cultural music perceptions, implying that similar patterns and tonal cues are apparent across tonal hierarchies³.

Different perspectives and findings in the area of research are presented in this literature review. The following sections will detail how music elicits intrapersonal judgment and interpretation. The paper will explore several studies that examine musical regularities and components as factors to trigger musical expectancy and classify music-induced emotions as grounds for analyzing their correlating interpretation. How the congruency of music and audiovisual material manipulates

audiences' association is demonstrated as well. This guides the application of music perception in audiovisual materials which the present literature lacks. The review encourages future research to probe more possibilities into how music triggers perception and enhances viewing experiences. Furthermore, the Cross-cultural Perspective section investigates if listeners can understand music and perceive emotions in an unfamiliar tonal system and explores the cultural universality and specificity of the psychophysical cues in music. Considering the lack of organized review for cross-cultural studies, it is vital to compile a body of work into one section independently.

Music Perception and Intrapersonal Judgement

Musical components such as tone and harmony work together to make segments of melody. They are like sentences in an article. Some sentences grab readers' attention and make people want to continue reading. Melody does the same. In a piece of music, melody forms structural regularities and generates music expectancy⁴. Music expectancy is a key factor that induces emotions, also known as music-induced emotions. These emotions are different from emotions people experience in their daily life. They are triggered solely by music and can elicit interpretation and intrapersonal judgment. In this section, the process by which music provokes judgments will be outlined, step by step, along with their corresponding empirical studies and theories.

Musical Expectancy

Since the late 19th century, a body of research centered around the role of expectancy. Musical expectancy was investigated from different contexts, including temporal and melodic expectancies⁴, melodic and harmonic expectancies on perception and performance^{5, 6}, and the implication-realization model of musical expectancy⁷. Music expectancy is a process whereby the listeners' expectation of continuation for music is violated, delayed, or confirmed by a specific feature in the music, resulting in music-induced emotions⁸. There are two underlying mechanisms through which music expectancy can be evoked, the long-term process and the short-term process. The long-term process requires prior musical experience, whereas the short-term process only depends on musical regularities.

Narmour helped to develop an explanatory model in 1990, the Implication-Realization model (i.e., IR model). The key finding of the model suggested that the emotional syntax in music is a product of the interruption and suppression of two expectation systems—the top-down system and the bottom-up system⁹. To understand the two systems, two important components should be acknowledged, namely implicative intervals and realized intervals. The Implicative interval is the last melodic interval, whereas the realized interval is the melodic interval between the last tone of the implicative interval and the one that followed.

In the literature, the bottom-up process explained the short-term, automatic, and unconscious generation of expectations. Five underlying principles explained the different distribution of melodic intervals (size and direction of an implicative interval) that apply to points in music when the expectation for continuation is strong¹⁰. Principle 1 was termed *registral direction* which suggested that small intervals imply a continuation of realized interval in the same direction, and large intervals imply a continuation in the opposite direction. Figure 1 has shown that intervals smaller than a perfect fourth were expected to proceed upward if the implicative interval ascends, and intervals larger than a perfect fourth were expected to proceed downward if the implicative interval ascends. New findings that emerged in 2015 discussed registral direction with cognitive energy and emotions. That is, “in any event, the perceptual change from up to down makes for a less tense sense of directional modulation and requires less cognitive energy to process than down to up.”¹¹

Principle 2 of *intervallic* difference assumed that small implicative intervals implied similar-sized realized intervals and large implicative intervals expected smaller realized intervals. Krumhansl¹² defined similar-sized and smaller-sized depending on the registral direction.

For small implicative intervals, if it goes in the same direction, the realized intervals would have the same size plus or minus a minor third; On the other hand, if it goes in the oppo-

site direction, the realized intervals would have the same size plus or minus a major second. For large implicative intervals, with the same direction, smaller meant smaller by more than a minor third; With an opposite direction, smaller meant smaller by more than a major second. Examples are shown in Figure 2.

The melodic structures of the IR model were mainly obtained by the first two principles, and three additional principles were implicit in the model. Registral Return was the third principle. It explained that the interval formed by the first tone of the implicative interval and the second tone of the realized interval is proximate¹². The size of proximity was sorted into two types: Exact registral return and near registral return. Exact registral return was defined to have the same of these tones. Take C4-A3-C4 as an example in Figure 3a, the first and last tones are both C4. Near registral return is the case when these tones are within a major second. C4-G3-B3 would be a good example displayed in Figure 3b as C4 and B3 are within a major second.

Principle 4 was *Proximity*. Different from the proximity size of the registral return, proximity suggested the independence of the size and direction of the implicative interval. It implied that realized intervals were no larger than a perfect fourth. For instance, C4-E4-A4 is weakly proximate and all larger intervals beyond that would be non-proximate. Finally, Principle 5 was *Closure*. It was explained that the closure is strongest when a) the implicative interval is large and the realized interval is smaller and b) the registral direction of implicative and realized intervals are different.

Different from the short-term bottom-up process, the long-term top-down process requires the interpretation of information with evidence of prior knowledge. Narmour explained that the top-down system allowed listeners to “constructively match and compare representative schemata to current input”, and that “schemata range from highly instantiated parametric complexes within a style to extremely generalized structuring of the elementary materials of a style”⁹. Even though the bottom-up process and the top-down process perform through different systems, they do not conflict with each other. The interaction of the two processes results in the phenomenon of suppression, whereby a learned formal schema of the top-down process inhibits the bottom-up implications and the various parameters belonging to it⁹.

Narmour gave an excellent example to explain this phenomenon (see Figure 4). In measures 5 to 6, listeners would expect A after E-F#-G according to the bottom-up system. However, due to the reverse registral direction from E-F#-G to F# in measures 1 and 2, listeners learned the schematic instantiation and were inclined to mirror it in measure 5. Here the top-up process suppresses the expectancy of the bottom-up process, and expect an F# at the start of measure 6. The composer dramatically put A instead, performing a potent aes-

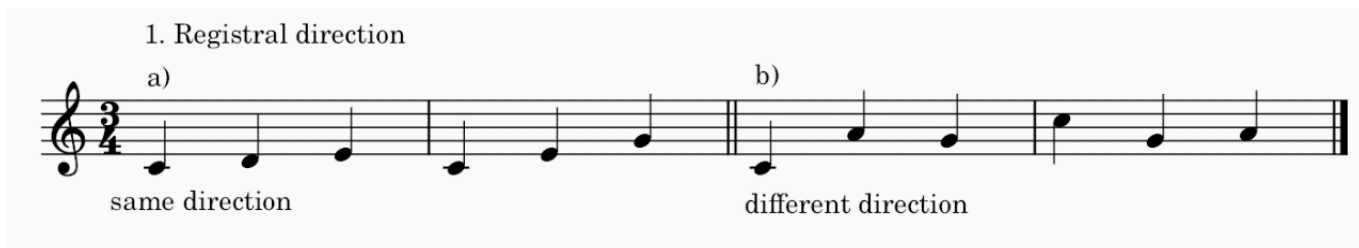


Fig. 1 The First Principle of the I-R model. **a)** The implicative intervals are smaller than a perfect fourth (5 semitones), therefore the realized intervals are expected to continue in the same direction: up-up, down-down, or lateral-lateral. **b)** The implicative intervals are larger than a perfect fourth, so it is expected to go in the opposite direction: down-up, or up-down.



Fig. 2 The Second Principle, the Intervallic difference. **a)** Since they are small implicative intervals and go in the same direction (up-up), their realized intervals would be the same size as the implicative interval plus a minor third (3 semitones) like the first bar or minus a minor third like the second bar. **b)** The small implicative intervals go in the opposite direction as the realized intervals (down-up), therefore the realized intervals should be the same size minus a major second (2 semitones) as the former bar or plus a major second as the later bar. **c)** The implicative interval is large and it goes in the same direction (up-up), so the realized interval has a size smaller than the implicative interval by at least a minor third (3 semitones). **d)** The large implicative with the realized interval going in an opposite direction (down-up) follows an interval smaller sized by at least a major second (2 semitones).

thetic event. This resulted in a “breaking out” of the suppression.

The continuous denial and suppression of implications form interruption and evoke aesthetic surprise. This causes listeners to experience emotional arousal, also known as music-induced emotions. Overall, Narmour successfully demonstrated how music expectancy led to music-induced emotions which set the first step for explaining how music elicits intrapersonal judgment. The next step would be the emotions music evokes.

Music-Induced Emotions

The knowledge of expressiveness in music is not recent. We can date back the exploration of musical meaning to the 1900s, or earlier. Controversy over the problem of distinctiveness in the affective state and interpretation remained a struggle in the area of research, until Kate Hevner’s fundamental attempt in establishing an orderly system that related musical structures and emotions¹³. In her seminal study, she proposed the adjective circle, grouping similar affective states into clusters. Later studies refined and adjusted the clusters, developing a well-rounded classification of music-induced emotions. Lists of terms have been recently exam-

ined by¹⁴, which guided later research into music perception and associations. For the rest of the section, the progression of classification is chronologically outlined to set a foundation in music perception.

Hevner’s adjective circle

In 1936, Hevner summarized a circuit of adjectives compatible with describing various affective states¹³. The eight clusters formed a continuous loop for which every two adjacent clusters shared certain commonalities, and that Cluster 1 and Cluster 8 did not represent extremity. The study allowed 52 participants to rate five different classical music compositions using the adjective circle. The results indicated significant uniformity. Four out of five compositions received a comparable choice of emotional terms, whereby potentially resolved the controversy of distinctiveness. Moreover, the literature offered an important finding: Modality is the most contributing element to its expressiveness, followed by harmony and rhythm of the music; Melody is the least important in expressing affective value. It is indisputable that Hevner proposed a groundwork for music-induced emotional terms. However, several arguable weaknesses need to be noted. Firstly, mod-

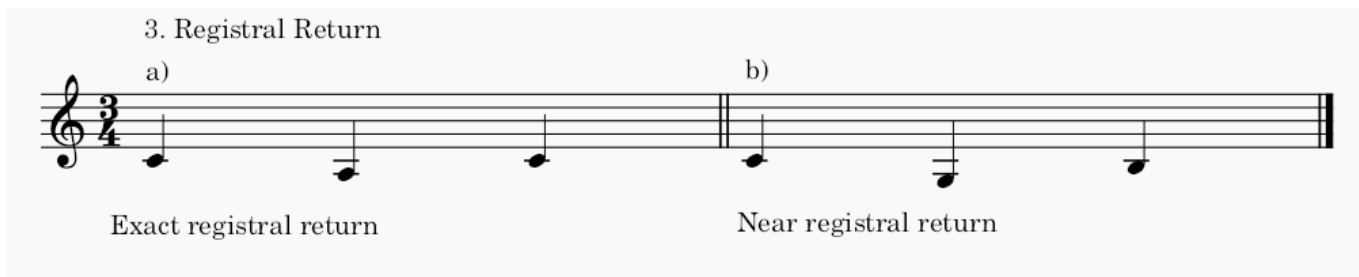


Fig. 3 Registral Return. a) demonstrates exact registral return. b) displays near registral return.

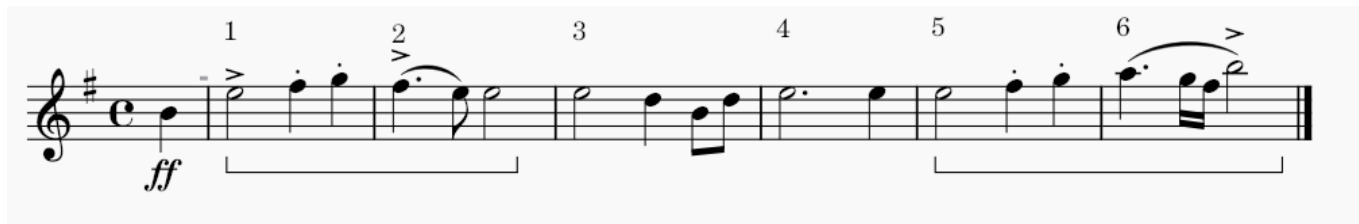


Fig. 4 Dvořák, Symphony No. 9, IV (Allegro con fuoco), mm. 34-39. Adapted from Narmour, E. 1991⁹.

erate emotional states are majorly comprised in the circle, but extreme affective states, for instance, “chaotic” and “tense” are neglected. Secondly, the grouping of terms is too broad. Cluster 4 included both “quiet” and “satisfying”. Quiet is the feeling of calmness and peacefulness, whereas satisfying depicts a more pleasant and delighted feeling. The two terms should not be in the same cluster. Based on the limitations, later research refined the work and provided more detailed categories of words.

Tentative Classification of the Basal Patterns Expressible in Music

Campbell suggested that participants in previous studies used a range of different vocabularies to express similar musical experiences. In his literature Basal Emotional Patterns Expressible in Music¹⁵ in 1942, he modified a set of 12 emotion categories. Categories included Gayety, Joy, Assertion, Sorrow, Yearning, Calm, Tenderness, Rage, Wonder, Solemnity, Cruelty, and Eroticism. A few adjustments to previous work were demonstrated. In Hevner’s adjective clock¹³, “joy” and “gay” were both included in Cluster 6, however, there was a clear difference between the two terms in Campbell’s results. Moreover, Hevner’s Cluster 5, 6, 7, and 8 were regarded to have similar characteristics in experience. Thus, in Campbell’s study¹⁵, these terms were reduced to three groups: “Gayety”, “Joy”, and “Assertion”. The study improved on previous work and provided a more detailed and agreeable classification of emotional patterns, benefiting future works on ex-

pressiveness in music.

Wedin’s Dimension of emotion factors

A few decades later, Wedin attempted to elucidate the dimensional structure of emotional expression using factor analysis¹⁶. In 1969, he finalized the dimensions as Tension-Energy, Gaiety-Gloom, and Solemn Dignity. The Tension-Energy dimension or Dynamics is a bipolar dimension that determines the level of activation and intensity in the music. Its negative pole compiled terms like “pleasing” and “relaxed” while the positive pole included terms like “violent” and “agitated”. The second factor Gaiety-Gloom concerned bipolar emotions like “playful-doleful” and “glad-sad”. Factor III appeared to be unipolar which is resulted by the fact that emotions opposite to “solemnity” and “dignity” were not experienced by listeners in the study. This study differed from previous studies which classify emotional terms in groups and subdivisions. It gave a new insight into classification by incorporating dimensions based on different musical emotion factors. Nevertheless, such classification has one cautious limitation. It fails to distinguish between felt and perceived emotions. Perceived emotion is a sensory or cognitive process that does not reflect what the listener is feeling, since the perception of emotions does not require any emotional involvement^{17, 18}. If studies are not able to differentiate felt emotion from perceived emotion, the emotional terms cannot be regarded as musically induced emotions.

Geneva Emotional Music Scale

More recently, in 2008, the Geneva Emotional Music Scale (i.e., GEMS) succeeded in separating perceived emotions from felt emotions. It also accomplished the objective of collecting the entire spectrum of affective states that can be evoked by different genres of music¹⁴. This systematic investigation comprised four studies. The first two studies employed a 4-point rating scale to identify terms that were most frequently felt or perceived in music, or experienced in extramusical daily life. A 10-min distractor task was adopted to ensure that participants correctly evaluate felt and perceived emotions. 10 emotional factors were derived, and MANOVA and ANOVA factor analyses were conducted for data reduction. Furthermore, what stood out in the result was that reflective emotions (i.e., tender longing, amazement, spirituality, and peacefulness scales) were often experienced by jazz and classical music listeners; Techno and Latin American music listeners more frequently experienced energetic emotions; Pop and rock music listeners got rebellious emotions.

More extensively, Study 3 examined the terms in a larger sample of listeners. One major advantage was that researchers could inspect emotional responses from different age groups and socioeconomic strata. However, an obvious limitation would be the use of focus groups. All studies were conducted in Geneva, Switzerland. This limits the representation of various cultural backgrounds and reduces its universality. Nevertheless, 801 responses were tested through confirmatory factor analytic (CFA) procedures for model construction. As shown in **Figure 5**,

GEMS consisted of 40 enhanced affect terms along with their groupings as first- and second-order factors. The second-order factor of Sublimity consisted of five first-order factors: Wonder, Transcendence, Tenderness, Nostalgia, and Peacefulness which each contained an average of five affective terms. The first-order factors of Power and Joyful Activation were compiled in the second-order factor of Vitality; Tension and Sadness were in the second-order group of Unease.

GEMS provided a well-rounded classification of most music-induced emotions and displayed a clear connection to how the affect states are related to different music genres. It also specified the experience of music from daily life emotions, reporting that people tend to be self-forgetful while listening to music and are usually detached from daily life concerns from the real world. However, no attempt has been made to investigate the cultural difference of affect terms in the classification. This is a common limitation for most studies in the area of research. It limits the standardization and globalization of the scale and lacks the comprehensiveness of the study. Considering cultural influence is typically important be-

cause people from different cultural backgrounds may have different interpretations of the affect terms. Yang and Hu¹⁹ projected a set of mood labels to a 2D space using multidimensional scaling to compare the Chinese dataset and the English dataset. In the study, they pointed out that English native speakers put “campy” and “witty” together with “cheerful” and “amiable”. This is because they thought that the four terms can all be associated with “fun”. On the other hand, Chinese listeners separated the four terms a part because they interpreted the first two terms as neutral and the last two as positive. It is evident that people can perform distinctive judgments based on the same taxonomy even though the semantic meanings of the terms are agreeable. Therefore, future studies of classification must question their applicability to cultural groups.

As music emotions are identified and classified, they are used in various studies to investigate their influence on listeners’ associations with music. Emotions have direct and indirect effects on intrapersonal judgment. Many empirical researches and systematic reviews examined the direct effect in relation to audiovisual materials. Meanwhile, indirect effect involves the Schema Theory and the Congruence-Associationist Model which discussed how music activates schemas and manipulates listeners’ attention to perform new interpretations. In the following sections, how music-induced emotions, directly and indirectly, generate intrapersonal judgment will be elucidated.

Direct Effect of Emotions

After assessing the 24 empirical studies from 1956 to 2018, a systematic review demonstrated that there is a general agreement by which “music can shape the perception of the general mood of a film or evoke a specific mood towards an individual protagonist”²⁰. Chełkowska-Zacharewicz and Paliga²¹ suggested that audiences were able to provoke corresponding associations when emotions were perceived and Vitouch¹ presented that audiences are able to evoke expectations towards the storyline with the help of music and its induced emotions. Thus, the question “How music-induced emotions form associations and intrapersonal judgment in audiovisual materials” may be answered.

Reviewing Vitouch’s study, *When Your Ear Sets the Stage: Musical Context Effects in Film Perception in 2001*¹, the quantitative content analysis examined two versions of the music score: the original score by Miklós Rózsa and a pre-tested fake score of Samuel Barber’s *Adagio for Strings*, op.11. The former was expected to evoke positive expectations and the later was designed to evoke sad emotions. The study hypothesized that the different music excerpts would provoke different plot expectations, most significantly in the

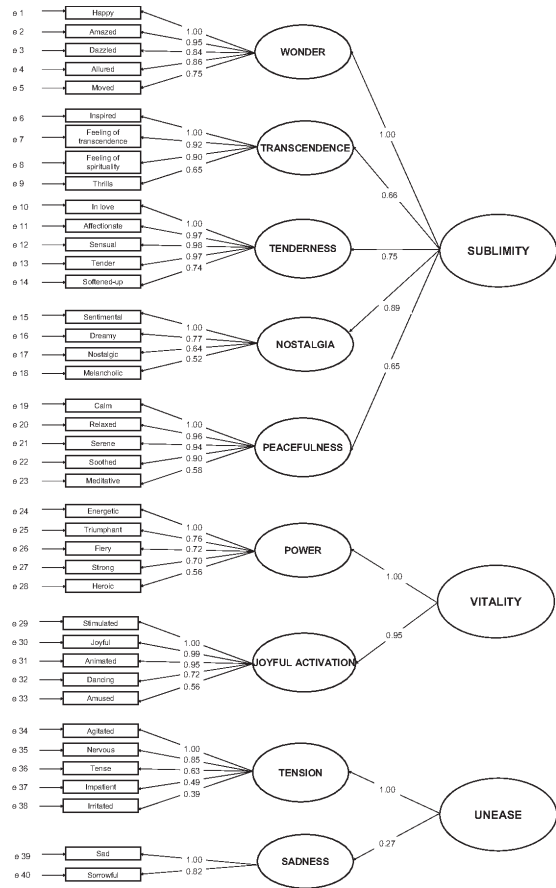


Fig. 5 Taxonomy of musically induced emotions based on Confirmatory factor analysis. Boxes on the left side are the 40 affect terms. 9 factors in the middle are first-order factors. 3 factors on the right are second-order factors. Values on arrows are the standardized parameter estimates. *Reprinted from Zentner, M., Grandjean, D., & Scherer, K. R. 2008.*¹⁴

emotional content of participants' plot continuations. **Table 1** displays the observed frequencies for positive and negative plot continuation against the expected frequencies within 22 indifferent cases. What stands out in this table is that the original (Rózsa) version evoked significantly more positive plot continuations, whereas the Baber version elicited more negative continuations.

Furthermore, the majority of participants mentioned film music as a factor influencing their continuation. This confirmed the effect of music-induced emotions on the viewer's perception of the storyline. However, Vitouch reflected that the music effect was not as strong and clear-cut as expected and that film music was identified more often as a building block than a major determinant¹. Moreover, the study had a small sample size which reduced the significance of the results. Future research can prepare a larger range of samples with more specified music-induced emotions to investigate how expectations vary with their corresponding emotions.

Around 20 years later, Chełkowska-Zacharewicz and

Paliga²¹ adopted a larger sample (157 participants) to assess the intrapersonal judgment of the seven groups of musical themes. The study aimed to examine the relationship between felt emotions and the correlating associations that emerged from motifs of the movie *The Lord of The Rings*. GEMS scale¹⁴ was used to assess the music-induced emotions across 7 groups of 15 motifs. Results revealed that felt emotions and associations were mostly consistent with the composer Howard Shore's intention for characterizing each motif. The *Hobbits* motif was rated high on Peacefulness, Tenderness, and Joyful activation while associations were mainly about nature and positive emotions; The *Evil* motif had higher ratings on Tension, relating to associations of danger, battle, and negative emotions. One anomaly would be the high reference to associations of power, victory, and majesty even though the ratings of Power were not as significant. This proved the relationship between emotions and interpretations that music-induced emotions can influence audiences' interpretation of audiovisual materials. Furthermore, it gave composers and

filmmakers an approach to shift viewers' interpretation toward what they intended to convey by employing accompanying music.

Notwithstanding, a few limitations remain unresolved in the previous studies. One significant weakness is the choice of audio samples. Herget suggested that the preconditions for the success in which participants can correctly identify the intended meaning in the film music are that the listeners a) can recognize the well-known music or b) is familiar with typical music genres and instrumental stereotypes (and their inherent messages), and c) can interpret the meaning behind by perceiving the music's expressed emotions²². Many studies failed to eliminate the factor of familiarity and the existing musical stereotypes in the tested music samples. Thus, Herget (2020) aimed to address this empirical research gap by examining the effect of conveying meaning for well-known and unknown music²².

To operate different degrees of familiarity, the well-known melodies were chosen from *Titanic* (97% rated familiar) for romantic music and *Mission Impossible* (70% rated familiar) for dramatic music. Unknown melodies were chosen from online music libraries which have similar style (e.g., in terms of instrumentation and tempo) as the selected well-known music without being "sound-alikes". Results have shown in **Figure 6** that as hypothesized, background music with different emotional connotations elicited specific association and guided to different perceptions and interpretations of the film's plot and protagonists. Though, there was no significant effect on the level of familiarity. This means that the perception of music does not require prior experience or schema and is unconsciously and directly generated. Herget recommended future research to examine "whether well-known and unknown music can convey meaning to the same extent in realistic contexts and persuasive communication, such as documentaries and advertising". Since past studies offered advice not to repeatedly use the same well-known background music in film and television practice due to its potential wear-out effects and audience reactance²².

Beyond the direct effect of emotions on interpretation, music can indirectly incent prior experiences in audiences using its emotional characteristic. This helps to elicit interpretations of its corresponding audiovisual materials. DePree²³ summarized that background music and soundtracks usually provide cues for the plot, foreshadow following events, and help audiences elicit interpretations when shots are ambiguous and neutral. One underlying theory to explain this phenomenon would be the Schema Theory.

Indirect Effect of Emotions

The Schema Theory is a branch of cognitive science first proposed by J. Piaget and by F. C. Bartlett. The most im-

portant concept, the schema, is an interpretative framework that organizes memory and experiences into understandable cognitive structures. Bobrow and Norman suggested that memory schemata are constructed using context-dependent descriptions, for which a retrieval mechanism requires two sources of information: the focus schemata (i.e., description) and the context²⁴. Concurrently, a large number of these active schemata interfere asymmetrically in some central resource pools, waiting for processing. Within the hierarchies, low-level computational processing performs the first stage of analysis to interpret and predict available data. Data that low-level processing is incompetent to operate would be shipped to higher processes for interpretation.

With the context of music, the sound of music travels through the sense organs as sensory data to be processed²⁴. Therefore, music, with its characteristic of eliciting emotions, can activate a schema and develop new interpretations based on previous knowledge and experiences. More specifically, the positively or negatively connoted musical schemata from the background music activate positive or negative valence schema.²² Then, they go through low-level computational processing to extrapolate a future scenario in the audiovisual material that is consistent with the implied emotion. However, when the following scene is not as expected due to its incongruity of emotion, a higher-level process is activated. Thus, a new interpretation is generated and it becomes more memorable.

In one of the few empirical studies considering Schema Theory, Boltz (2001) examined the indirect effect of emotions on the interpretation of audiovisual materials⁴. The study targeted how positive or negative music may bias audiences' evaluation of the protagonist's social behaviors. Participants were required to indicate whether they believed the protagonist would hurt the other character and the possible intention and personality of the characters. Results indicated that in all three film clips, participants replied "No" harm and positive intentions in positive music; Negative music received mostly "Yes" harm and negative intentional predictions. Answers varied in the no music condition. This is consistent with Schema Theory which stated that music explains what is going on in the scene by integrating it into an affectively consistent framework. Interpretations are produced, thus affecting the audience's understanding of the film.

A more recent study supported the indirect effect on audiovisual material. Steffens (2018) aimed to investigate the potential influence of music on moral judgment in the context of film reception²⁵. A large sample (252 participants) was tested for their intrapersonal judgment on morality and emotion. As shown in Figure 7, the standardized regression coefficients were significant between the music condition and happiness, as well as the standardized regression coefficients between happiness and perceived rightness of action. Though

Emotional content (EC) of plot continuations, sub-categories combined

		<i>negative</i>	<i>positive or ambivalent</i>	
Róza	Count	8	14	22
	Expected C.	11.0	11.0	(22)
Barber	Count	14	8	22
	Expected C	11.0	11.0	(22)
Total		22 (22)	22 (22)	48

Table 1 Matrix for statistical testing of between-group differences in positive v. negative plot continuations. Reprinted from Vitouch, O. 2001¹

the direct effect of musical condition on the rightness of action in terms of standardized regression coefficients were rather insignificant. The study has supported the third hypothesis for which the induced emotions significantly affect the perceived rightness of action.

In addition to the exploration of judgment to protagonists, Boltz also examined audiences’ selective memory of certain objects in the film with the effect of congruency in audio and visual materials⁴. It appeared that positive objects were better remembered along with positive music and negative items were better remembered along with negative music. Music can guide selective attention toward mood-consistent information and away from other information that is inconsistent with its affective valence⁴, and this is to do with *congruency*. Cohen stated that “Congruence focused on the overall music-visual compatibility on the semantic differential dimension of activity and potency”²⁶. This outlines the basic concept of the Congruence-Associationist Model.

Effect of Congruency

The Congruence-Associationist Model (i.e., CAM)²⁷ is proposed by Marshall and Cohen and discusses how audiovisual materials govern the viewer’s attention and evoke emotions and interpretation at a focal point of *congruency*. In general, the association part is the connection between emotion and music elicited by the audiovisual material, and the congruence part of the CAM model refers to the priority and dominance of attention towards information that performs congruently across two modalities²⁷.

The structure of music and visual material is determined by three dimensions: Evaluative (e.g., good/bad), Potency (e.g.,

weak/strong), and Activity (e.g., active/ passive)²⁸. CAM is processed as follows (**See Figure 8**): First, the music induces emotion based on the Potency and Activity components in the music. Then, the emotion relates itself to the congruent Potency and Activity characteristics of the film. The degree of *congruency* found in the Comparator establishes the Evaluative meaning of audiovisual material. The attention is directed at the Evaluative meaning, also known as the focal point of *congruency*. At last, the attention is associated with a special component in the music to form a new connotation of the film. Consequently, interpretation is altered in the audiovisual material. This model explains how music works with visual content to produce and manipulate interpretation. Apparently, music is a strong influential component in the audiovisual material that determines perception.

Recent studies have tested CAM and demonstrated its validity. In 2023, Ansani et al. demonstrated that music’s emotional content affects the listener’s mood, directs attention to the point of congruency, and thus affects performance, namely the recognition memory²⁹. The study paired the short visual material On Lockdown³⁰ with three experimental conditions: a happy audio piece (Appalachian Spring - VII: doppio movimento by A. Copland), a suspenseful audio piece (Murder by the Newton Brothers), and a control condition with no music. A vast spectrum of studies has suggested music as an undeniably effective carrier of emotions. Likewise, the study verified their first hypothesis that “music has an impact on the affective state of the recipients coherently with its emotional valence”, conforming to the first step of the CAM. More importantly, Ansani et al. ran the interaction post-hoc analyses to test their third hypothesis for whether music biases remembering coherently with its emotional content. As hypoth-

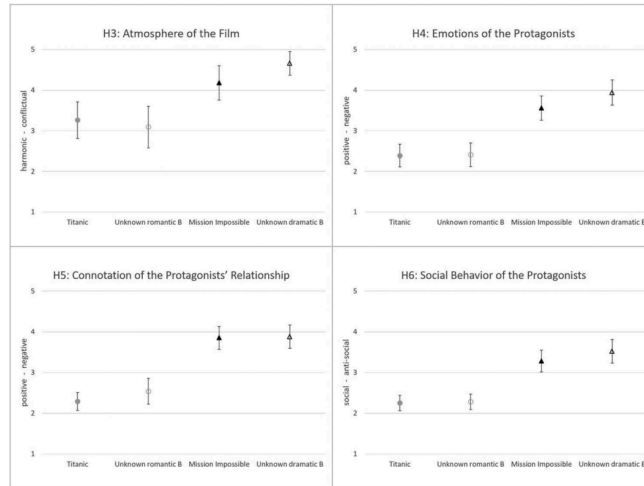


Fig. 6 Arithmetic means on the perception of the film’s atmosphere and the protagonists’ emotions, relationships, and social behavior. Reprinted from Herget, A.-K. 2020.²²

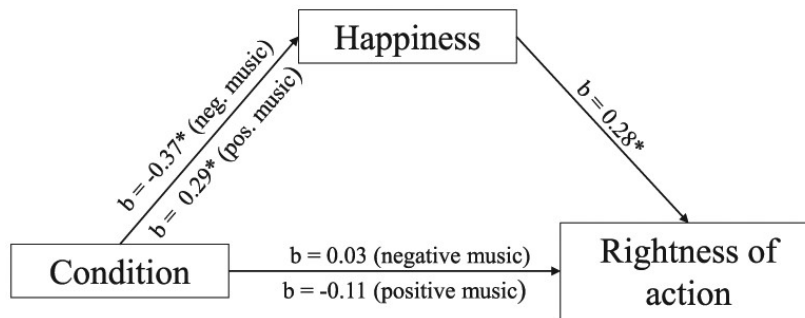


Fig. 7 Indirect effect of music condition on perceived rightness of action, mediated by induced happiness. Reprinted from Steffens, J. 2018.²⁵ Note: *Regression coefficient is significant at the .05 level (two-tailed).

esized, the positive-happy condition established the Evaluative meaning on the point of congruency and elicited a higher amount of positive-valenced false memories than negative-valenced ($p = .026$) and control conditions ($p = .014$). Similarly, the negative-scary condition followed its congruence and elicited a higher amount of negative-valenced false memories than positive ($p < .001$) and control conditions ($p = .031$)²⁹. The positive and negative-valenced false memory recognition scores were not significant in controls, hence proving the effect of congruency, namely the direction of attention in perception (See Figure 9).

In this section, the progress of how music elicits emotions and produces intrapersonal judgment is demonstrated. Starting from the music, the regularities of musical components generate expectancy known as the bottom-up process. By engaging with the top-down process, the suppression and denial generate emotional arousal, thus inducing emotions from music. A large body of literature aimed to identify such emo-

tional arousal and attempted to classify emotions into circuits or spectrums. As typical emotions were recognized, along with the engagement of visual materials, music-induced emotions were found to have direct and indirect effects on listeners’ interpretation and intrapersonal judgment of audiovisual materials. This can be explained by the Schema Theory, the CAM, and many empirical studies. Eventually, audiences may have their moral judgment, prediction, and association for future scenarios impacted by the induction of emotions in music and the congruency with the visual components.

Application in Audiovisual Materials

The above investigation of music, emotion, and intrapersonal judgment suggests several practical applications in audiovisual materials. Firstly, in cinematic art, music can act as an iconic representation, also known as motif³¹. That is, when

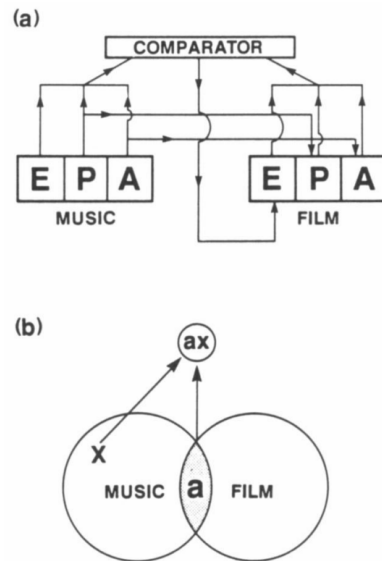


Fig. 8 The Congruence-Associationist Model. Reprinted from Marshall, S. K. and Cohen, A. J. 1988.²⁷ Note: (a) The three dimensions are indicated as Evaluative (E), Potency (P), and Activity (A). (b) The focal point of congruency is indicated as the symbol “a”. The association of music is represented by “x”, and the interpretation is represented by “ax”.

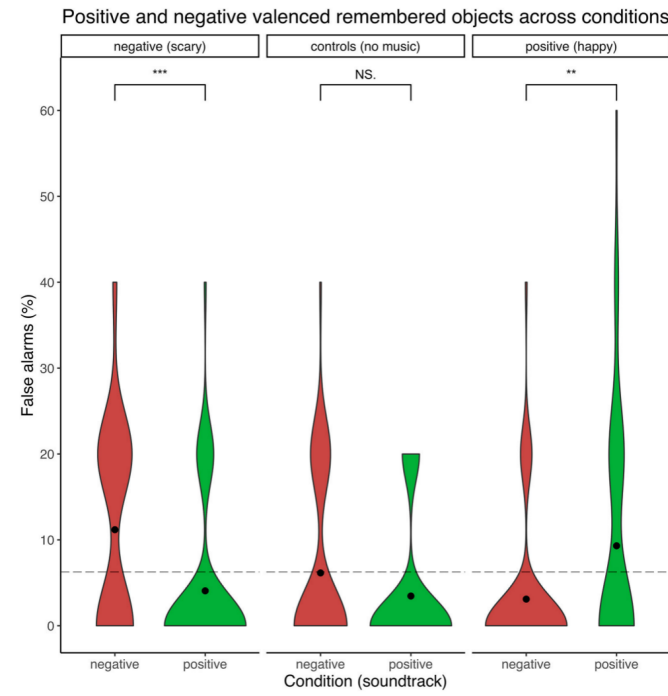
a soundtrack pairs with a specific character or scene, it builds schemata in audiences that every time the music plays, the character or scene appears. This is helpful in movies since it can imply certain characteristics of protagonists using music’s emotional remarks and foreshadowing upcoming scenarios. It is very commonly used in the movie *The Lord of The Rings*, examined in Chelkowska-Zacharewicz and Paliga’s study²¹.

Secondly, since music can evoke intended music-induced emotions congruence to the visual material, a film director can heighten the emotional arousal of a particular scene and its effects on listeners³¹. A typical example would be the Chinese mystery thriller movie *Lost in The Stars*. At nearly the end of the movie when the protagonist He Fei realizes that he has gone into a trap, the dramatic and loud background music arouses audiences’ excitement and stress and amplifies the emotional effect to a peak in the film. This enhances the viewing experience.

Thirdly, music has a persuasive effect on specific attributes of a brand in commercials. Background music that can evoke excitement and a sense of power may influence the emotional process of the consumers to have greater enjoyment and engagement.³² Consumers would be more inclined to have a positive attitude toward the brand or the product, thus increasing their attention and possibly their motive to consume. Apparently, by effectively using the emotional attribute of music and the congruence of audio and visual material, background soundtracks can reinforce the effect of audiovisual medium for various purposes and applications.

In addition, recent research analysed the influence of music on visual scenes through eye tracking and pupillometry. Ansani et al. (2020) found it rather interesting that anxious background music induced more attention to the details in the scene and the control condition (with no background music) was found to stimulate the greatest pupil dilation³³. The study used five metrics for eye tracking: time spent, fixation, revisits, dispersion, and pupillometry. For the metric of fixation, Ansani et al. built two moving areas of interest (mAOIs) to measure the gaze points: the main character’s full body and head, and the almost hidden character, the cameraman. Results have verified their first hypothesis and indicated that the anxious music condition (*The Isle of the Dead* by S. Rachmaninov) showed higher ratings of time spent ($M = 6197, 15\%$) and revisits ($M = 11.07$) on the hidden cameraman than the melancholic music condition (*Like Someone in Love* by B. Evans) (time spent: $M = 4486, 11\%$; revisits: $M = 8.67$)³³. This result draws our attention to the implication in the audiovisual medium that especially for horror, mystery, and thriller movies, anxious soundtracks can be used to trigger audiences’ attention and alertness towards details in the image.

Furthermore, iMotions provided data for pupil dilation by giving the average of both eyes’ pupil sizes in a set time. Examining the two music conditions, the anxious soundtrack (Rachmaninov group) caused greater pupil dilation ($M = 0.27, SD = 0.24$) than the melancholic soundtrack ($M = 0.20, SD = 0.23$)³³. One surprising result could be



found by comparing the percentage change in the conditions in **Figure 10**. The control condition had overall the greatest pupil dilation. Ansani et al. explained the phenomenon that pupil dilation can expose the increasing emotional arousal and cognitive load in participants as the circumstance is dark and silent without any accompaniment of music. Thereupon, a sudden rest in the music or a designed soundtrack blank can be applicable to generate a high state of alert in the audiovisual medium.

The following section outlines the cross-cultural studies conducted relating to music perceptions and induced emotions. Three questions are explored: a) What are the similarities and differences in perceived emotions across cultural groups? b) Can people interpret music from an unfamiliar tonal system? c) What are the cultural universality and specificity of psychophysical cues in music?

Cross-Cultural Perspectives

In recent times, global music has been considered to be a focus of investigation for music perception and cognition. Numerous studies have sought to cross-culturally discuss the similarities and differences of music expectancy and interpretation of music emotions under the prerequisite of musical context. Stevens pointed out that cross-cultural research is necessary

for three main reasons: 1) there is an absence in knowledge of different musical systems and psychological processes, 2) past research was monoculturally based on Western music that they should be evaluated and challenged by a variety of contexts, and 3) prior investigations to the West inhibits the acceptance of theory and empirical findings resulted from diverse cultural settings³⁴.

As mentioned in the previous section, musically induced emotions can be classified into numerous clusters. Related literature and classifications were mostly based on Western music. Even though basic emotional expressions can be universally recognized across cultural groups³⁵, what is different across cultures is the perception and understanding of affective emotional responses to music genres. Hu and Lee³⁶ conducted empirical research comparing the difference in perceived emotions between Americans and Chinese for popular songs across genres. Listeners responded to share more agreements with people from the same cultural background. American listeners agreed more on cheerful and aggressive clusters of emotions, whereas Chinese listeners agreed more on passionate and literate clusters of emotions. Based on the same piece of music, people from different cultural groups may interpret it differently. Likely, Western people are more sensitive to the intended emotions in Western songs, and Chinese people understand Chinese songs better. Therefore, the problem

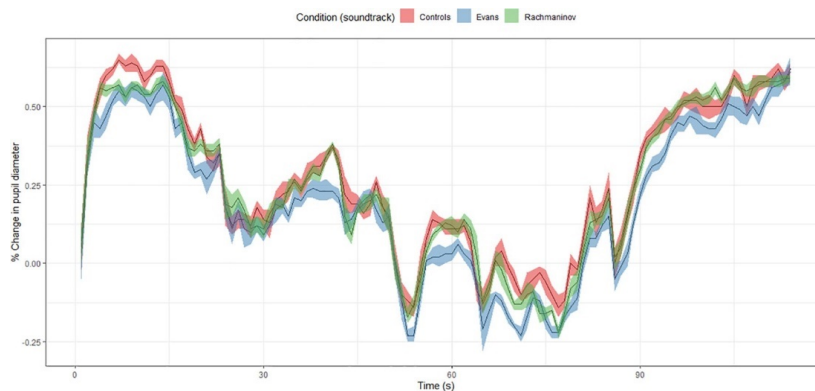


Fig. 10 Percentage change in pupil diameter as a function of time (s) and condition. Reprinted from Ansani et al. 2020.³³

with solely using Western music samples is obvious. There is a clear bias toward the American cultural group as all 30 test songs are Western songs. Possibly, participants in the Chinese cultural group are not able to understand the songs, thus giving false perceptions of music emotions.

Emotional Judgement Across Cultural Groups

In two-year's time, Lee and Hu³⁷ refined their methodology and attempted to investigate the difference in music perceptions between three cultural backgrounds: American, Korean, and Chinese. Korean and Chinese cultural groups are chosen for their unique cultural distinctions and connections. Even though they are both Asians, Koreans are heavily influenced by Western cultures due to America's involvement in World War II. Koreans seem to be an intermediate linkage between Western and non-Western cultures.

The study aimed to test perceived music mood from three perspectives: mood judgment distributions and agreement levels, music (stimuli) characteristics, and listeners' (subjects) characteristics. Participants were given 30 music clips consisting of instrumental, pop, and rock songs with lyrics and asked to select corresponding perceived emotions from five mood clusters used in MIREX (i.e., Music Information Retrieval Evaluation exchange). It was shown that American listeners were more familiar with the songs and Korean listeners indicated twice as familiar as Chinese listeners. The difference in levels of familiarity provided a variable that Americans are more likely to have prior knowledge for some of the music clips and thus perceive emotions based on existing knowledge. Moreover, it was proven by both studies that lyrics may affect the perception of music emotions as American listeners reach a much higher agreement ratio on English vocal songs^{36,37}. Korean and Chinese listeners had approximately the same agreement ratio on vocal music as the lyrics are not in their native language. With these characteristics,

similarities and differences in the choice of emotions were presented. It was shown that Americans and Koreans had a close number of selections for Cluster 2 (e.g., cheerful, fun, sweet, rollicking) than Chinese participants; Koreans and Chinese had a higher rate of selection for Cluster 3 (e.g., literate, poignant, wistful, bittersweet) than Americans; Koreans had much fewer selections on Cluster 5 (e.g., aggressive, fiery, tense/anxious, volatile) even though Americans and Chinese had similar ratings. Regarding genres of music, it was indicated while American listeners had the most agreement ratio on Pop songs, Koreans agreed most on the Easy-listening genre and Chinese on Other genres. Overall, cross-cultural agreements were lower than intra-cultural agreements across all genres of music. Within the three-way comparison, Chinese and Korean listeners had a rather similar perception of musical emotions and American and Chinese listeners displayed the greatest difference among the three cultural groups. The literature³⁷ concluded that it is very necessary to compare multiple cultural groups instead of a two-way comparison between Western and non-Western groups and that the understanding of lyrics and culturally characterized music can affect judgment and perceptions.

Psychophysical Cues Across Cultural Groups

Kessler et al.³ compared the music perception of both Western (i.e., diatonic scale) and Balinese (i.e., pelog and slendro scales) melodies between Western and Balinese listeners. While half of the Balinese listeners had some exposure to Western diatonic music, the other half had no exposure to Western cultural or diatonic scales. The study was based on the probe-tone method³⁸ which listeners had to rate on a scale from 1 to 7 for how well they think the probe tone fits in the preceding musical context. According to past experiences, listeners would give high ratings to probe tones in their familiar scale and lower ratings to unfamiliar non-scale tones. Results

indicated that both cultural groups rated fewer variations in familiar contexts and more within-group variation in unfamiliar contexts. Western listeners had the least success in abstracting tonal hierarchy from the Balinese *slendro* contexts, suggesting that cultural factors have a considerable influence on people's perception and interpretation of music. However, a tonal hierarchy of tonal functions, scale membership, pitch height, and frequency of tones were found to have the same patterns for both cultural groups, indicating that despite cultural learning, human cognitive universality is present.

A similar study examined Western listeners of Hindustani music. Balkwill and Thompson attempted to answer two research questions: a) Can people identify the intended emotion in music from an unfamiliar tonal system? b) If they can, is their sensitivity to intended emotions associated with perceived changes in psychophysical dimensions of music?² A selection of 12 raga performances each evoking a target emotion (e.g., joy, sadness, anger, and peace) was presented to a group of listeners raised in Western culture. Listeners were asked to rate on a scale from 1 to 9 how strongly they thought the emotion was conveyed, as well as the psychophysical components (i.e., tempo, rhythmic complexity, melodic complexity, and pitch range) in the music excerpt. Results showed that Western listeners were able to identify joyful and sad emotions in the ragas even though they were not familiar with Hindustani music. This has supported their hypothesis that "listeners enculturated to one tonal system can accurately perceive the intended emotion in music from an unfamiliar tonal system." Furthermore, by comparing to expert listeners who are deeply familiar with the Hindustani culture, results found that naïve listeners demonstrated a high level of agreement for psychophysical cultural cue ratings with expert listeners. Overall, the study revealed that there are universal psychophysical and emotional cues that all listeners can identify no matter what cultural group they belong to.

Timbre, being one of the psychophysical cues that greatly impacts the affect perception, Wang et al. (2021) examined the relationship between timbral properties and perceived affect ratings among Western and Chinese listeners through a linear partial least squares regression (i.e., PLSR)³⁹. Listeners were asked to perceive the three-dimensional affect that was expressed in music, namely the valence, the tension arousal, and the energy arousal. At the same time, MIRTtoolbox and Timbre Toolbox were used to analyze the acoustical features of timbre in the stimuli. Results have shown that there was no significant difference in valence and energy arousal between the perception of Western and Chinese listeners. For both groups, greater valence and higher energy arousal were associated with more spectral variation, more impulsive-type note envelopes (e.g., staccato, pizzicato) with sharp attack, and more dynamic range. However, the acoustic features that

are perceived to affect tension arousal were different between the two cultural groups. Looking at Figure 11, it is apparent that Western participants regarded more spectrum energy variation over time, wider spectral distribution, and noisier sounds with sharp decay as music stimuli with higher tension arousal. Whereas, Chinese participants believed that positive tension arousal is associated with more vibrato sounds with different note durations, greater temporal energy, wider spectrum distribution, and more high-frequency energy.

In the same year, Wang, Wei, and Yang (2021) investigated the cultural universality and specificity of the psychophysical elements overall⁴⁰. The PLSR analysis first examined the entire cross-cultural dataset, and then tested the four datasets independently, namely Chinese solo and ensemble, and Western solo and ensemble. Results indicated that tempo, pulse clarity, articulation, and dynamics greatly influenced emotional perception for all data sets, therefore suggesting their cultural universality. It was agreed across cultural groups that fast tempo, definite rhythm, and performance with staccato conveyed pleasure, meanwhile, greater dynamics range increased people's tension. However, elements with cultural specificity suggested that for Chinese classical ensemble music, register features including higher pitch range significantly impacted the emotional dimensions. A higher pitch range conveyed happiness and activity, while greater pitch variety with the engagement of multiple instruments made Chinese ensemble music more aesthetic. Rhythm complexity was another classical music element. For Chinese solo music, spaciousness negatively correlated with the three dimensions of emotional perception. This means that the smaller the performance area, the greater the emotional effect the music has. Wang, Wei, and Yang reasoned that Chinese musical instruments imitate the human voice, thus smaller space reveals more timbre characteristics and adds more flavor to the overall performance.

For Western music, few musical elements showed impacts on valence, tension arousal and energy arousal. For valence, timbre and rhythm including spaciousness, thickness, brightness, and rhythm complexity had great effects. For tension arousal, the main contributors were loudness range, pitch range, and pleasantness. In addition, both Western classical solo music and classical ensemble music indicated dynamics as the key factors for the perception of energy arousal. This study put a step forward for the exploration of universality and specificity in psychophysical cues. However, the datasets are limited to only classical music across cultural groups. In future work, the annotation of datasets can be expanded to more varied genres of music, for instance, pop and jazz music. Moreover, music from more cultural groups can be engaged, for example, African folk songs, Indian instrumental music, Balinese music, etc. Here, I recommend the UNESCO

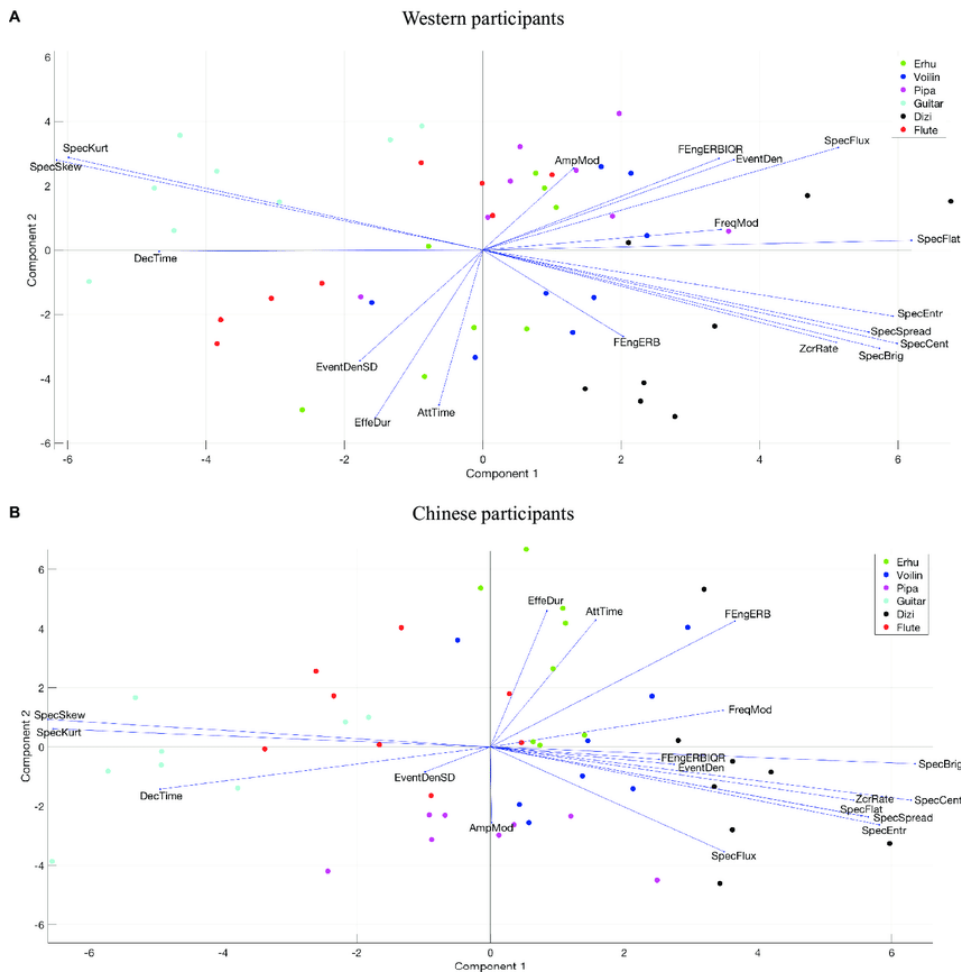


Fig. 11 Loadings and scores across two PCs of PLSR for tension arousal A) Western participants. B) Chinese participants. Reprinted from Wang et al. 2021.³⁹

Collection of Traditional Music⁴¹. It comprises 127 albums of music from around the world, representing more than 70 cultural groups.

Overall, after reviewing the empirical research, a few conclusions can be made: a) the perception of music emotion can vary across cultural groups due to the country’s history and social norms, whilst the lyrics can impact listeners’ understanding of the songs; b) similar patterns can be found in tonal systems across culture, therefore people from different cultures can correctly perceive the intended emotions in the music; c) some psychophysical elements are characterized to have cultural universality including tempo and articulation, and the others are characterized to have cultural specificity, thus cultural groups can have different perceptions for some psychophysical cues.

Conclusion

This literature review has discussed the music perception and intrapersonal judgment of music-induced emotions in audiovisual material, as well as the emotions and psychophysical cues in cross-cultural studies. Major findings reveal that a) familiarity with the music context has no direct effect on the perception of music; b) In audiovisual materials, music can indirectly affect the perceived rightness of action along with the music-induced emotions; c) According to the CAM, music-induced emotion can direct the attention at the point of congruency and elicit interpretation or false memory accordingly; d) tempo, pulse clarity, articulation, and dynamics indicate features of cultural universality, therefore there is a universal agreement on the effect of perception across cultural groups. These findings have significant implications in the audiovisual medium because a) operating suppression and denial of the two pro-

cesses in music allows filmmakers to design emotional effects and trigger intended emotions in the audience; b) arranging soundtrack congruence or incongruence to the visual material can manipulate audiences' judgment and attention toward specific detail or object in the scene; c) applying cultural specificity in the background music can convey cultural identity and indicate settings in audiovisual materials.

Limitations

Notwithstanding the agreeable results found in numerous works of literature, there are reasonable improvements necessary for future research. Firstly, studies that concerned the relationship between music-induced emotions and interpretations often categorize interpretation as positive or negative. There is a lack of detailed explanations for the interpretations. Future research can improve the methodology by specifying the contents of understanding elicited by different emotions. Secondly, most studies were achieved through self-reports done by the participants. This would limit the research because untrained participants might give false responses. A valid improvement would be to apply different methodologies, such as EMG measures, pupillometry, and fMRI scans, to observe physiological unconscious responses. By comparing the self-reports to the physiological responses, the accuracy and reliability of the results can be confirmed. Thirdly, as mentioned in the Cross-Cultural Perspectives section, more cultural music and groups should be engaged in the research. Most studies targeted Asian and Western groups as typical cultural representatives. However, this grouping of cultures is too broad. Future research should consider exploring differences in folk cultures all around the world. This would give a more in-depth investigation of cross-cultural studies.

Future Research

This literature review proposes a few areas for future research: Firstly, future studies should examine the principles of the bottom-up processes in cultural music and test if Narmour's Implication Realization model applies to all music across cultures; Secondly, cross-cultural studies can conduct a systematic review for the cultural universality and specificity of psychophysical cues across 70 cultures in the UNESCO Collection of Traditional Music; Thirdly, future studies can assess how intended emotions are expressed differently in western and cultural folk music.

Acknowledgment

I would like to thank my mentor, Erick Aguinaldo, for guiding me through the preparation process and paper writing, provid-

ing resources, and supporting me during stressful times. His tutoring helped me in comprehending parts of literature I don't understand, clarify my ideas, and structure my paper. I learned a lot from him in my academic writing and critical thinking skills.

References

- 1 O. Vitouch, *Psychology of Music*, 2001, **29**, 70–83.
- 2 L.-L. Balkwill and W. F. Thompson, *Music Perception: An Interdisciplinary Journal*, 1999, **17**, 43–64.
- 3 E. J. Kessler, C. Hansen and R. N. Shepard, *Music Perception*, 1984, **2**, 131–165.
- 4 M. G. Boltz, *Music Perception*, 2001, **18**, 427–454.
- 5 M. A. Schmuckler, *Music Perception: An Interdisciplinary Journal*, 1989, **7**, 109–149.
- 6 M. A. Schmuckler, *Psychomusicology: A Journal of Research in Music Cognition*, 1990, **9**, 122–147.
- 7 E. Narmour, *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*, The University of Chicago Press, 1990.
- 8 C. L. Krumhansl and K. R. Agres, *Behavioral and Brain Sciences*, 2008, **31**, year.
- 9 E. Narmour, *Music Perception*, 1991, **9**, 1–26.
- 10 C. L. Krumhansl, J. Louhivuori, P. Toiviainen, T. Järvinen and T. Eerola, *Music Perception: An Interdisciplinary Journal*, 1999, **17**, 151–195.
- 11 E. Narmour, *Music Perception: An Interdisciplinary Journal*, 2015, **33**, 32–69.
- 12 C. L. Krumhansl, *Music Theory Spectrum*, 1995, **17**, 53–80.
- 13 K. Hevner, *The American Journal of Psychology*, 1936, **48**, 246.
- 14 M. Zentner, D. Grandjean and K. R. Scherer, *Emotion*, 2008, **8**, 494–521.
- 15 I. G. Campbell, *The American Journal of Psychology*, 1942, **55**, 1.
- 16 L. Wedin, *Swedish Journal of Musicology*, 1969, **51**, 119–140.
- 17 A. Gabrielsson, *Musicae Scientiae*, 2002, **5**, 123–147.
- 18 R. Harré, *Emotion and the Arts*, Oxford University Press, 1997, p. 110–118.
- 19 Y.-H. Yang and X. Hu, Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012, 2012, pp. 19–24.
- 20 A.-K. Herget, *Psychology of Music*, 2019, **49**, year.
- 21 M. Chełkowska-Zacharewicz and M. Paliga, *Roczniki Psychologiczne*, 2019, **22**, 151–175.
- 22 A.-K. Herget, *Media Psychology*, 2020, 1–28.
- 23 K. M. DePree, 2017.
- 24 D. G. Bobrow and D. A. Norman, *Representation and Understanding*, 1975, p. 131–149.
- 25 J. Steffens, *Psychology of Music*, 2018, **48**, year.
- 26 A. Cohen, *The Psychology of Music in Multimedia*, 2013, p. 17–47.
- 27 S. K. Marshall and A. J. Cohen, *Music Perception: An Interdisciplinary Journal*, 1988, **6**, 95–112.
- 28 C. E. Osgood, G. J. Suci and P. H. Tannenbaum, *The Measurement of Meaning*, University Of Illinois Press, 1957.
- 29 A. Ansani, M. Marini, I. Poggi and L. Mallia, *Journal of Cognitive Psychology*, 2023, **35**, 59–75.
- 30 C. D. Macdiarmid, *On Lockdown [Animated Short Movie]*, 2020, <https://vimeo.com/435128203>.
- 31 M. G. Boltz, B. Ebendorf and B. Field, *Music Perception*, 2009, **27**, 43–59.
- 32 U. Cuesta, L. Martínez-Martínez and J. I. Niño, *European Journal of Social Science Education and Research*, 2018, **5**, 84.

-
- 33 A. Ansani, M. Marini, F. D'Errico and I. Poggi, *Frontiers in Psychology*, 2020, **11**, year.
 - 34 C. J. Stevens, *Topics in Cognitive Science*, 2012, **4**, 653–667.
 - 35 T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A. D. Friederici and S. Koelsch, *Current Biology*, 2009, **19**, 573–576.
 - 36 X. Hu and J.-H. Lee, International Society for Music Information Retrieval Conference, 2012, p. 535–540.
 - 37 J.-H. Lee and X. Hu, IConference 2014 Proceedings, 2014.
 - 38 C. L. Krumhansl and R. N. Shepard, *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 579–594.
 - 39 X. Wang, Y. Wei, L. Heng and S. McAdams, *Frontiers in Psychology*, 2021, **12**, year.
 - 40 X. Wang, Y. Wei and D. Yang, *Cognitive Computation and Systems*, 2022, **4**, 116–129.
 - 41 UNESCO - UNESCO Collection of Traditional Music of the World, *UNESCO Collection of Traditional Music of the World*, <https://ich.unesco.org/en/collection-00002>, Accessed on [Insert Date].