석 사 학 위 논 문
Master's Thesis

음악 연주와 사회의 정서적 상황에서 나타나는 공감 비교 연구: 모달리티와 정서가에 따른 공감 정확도 및 심전도 반응을 중심으로

Shared empathic process in music and social contexts: Exploring empathic accuracy and physiological responses across modalities and valence

2024

오 은 지 (吳 垠 知 Oh, Eun Ji)

한 국 과 학 기 술 원

Korea Advanced Institute of Science and Technology

석 사 학 위 논 문

음악 연주와 사회의 정서적 상황에서 나타나는 공감 비교 연구: 모달리티와 정서가에 따른 공감 정확도 및 심전도 반응을 중심으로

2024

오 은 지

한 국 과 학 기 술 원

문화기술대학원

음악 연주와 사회의 정서적 상황에서 나타나는 공감 비교 연구: 모달리티와 정서가에 따른 공감 정확도 및 심전도 반응을 중심으로

오 은 지

위 논문은 한국과학기술원 석사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2023년 12월 12일

심사위원장 이 경 면 (인)

심 사 위 원 이 정 미 (인)

심 사 위 원 박 지 영 (인)

Shared empathic process in music and social contexts: Exploring empathic accuracy and physiological responses across modalities and valence

Eun Ji Oh

Advisor: Kyung Myun Lee

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Master of Science in Culture Technology

Daejeon, Korea
December 12, 2023

Approved by

Kyung Myun Lee
Professor of Graduate School of Culture Technology

The study was conducted in accordance with Code of Research Ethics[1].

[1] Declaration of Ethical Conduct in Research: I, as a graduate student of Korea Advanced Institute of Science and Technology, hereby declare that I have not committed any act that may damage the credibility of my research. This includes, but is not limited to, falsification, thesis written by someone else, distortion of research findings, and plagiarism. I confirm that my thesis contains honest conclusions based on my own careful research under the guidance of my advisor.

MGCT 오은지. 음악 연주와 사회의 정서적 상황에서 나타나는 공감 비교 연구: 모달리티와 정서가에 따른 공감 정확도 및 심전도 반응을 중심으로. 문화기술대학원 . 2024년. 35+iv 쪽. 지도교수: 이경면. (영문 논문)
Eun Ji Oh. Shared empathic process in music and social contexts: Exploring empathic accuracy and physiological responses across modalities and valence. Graduate School of Culture Technology . 2024. 35+iv pages. Advisor: Kyung Myun Lee. (Text in English)

초 록

공감(empathy)은 사회에서 대인간 의사소통을 하고 관계를 맺는 데에 필요한 핵심 요소로, 타인의 감정에 대한 이해를 통해 나타난다. 일상적인 사회적 상황 뿐만 아니라 음악 공연과 같이 사람 간의 상호작용이 일어나는 다양한 맥락에서 공감은 중요한 역할을 한다. 최근 여러 연구들이 음악 공연에서 사람들이 공유하는 정서적 경험을 생체 반응 간 동기화(synchrony) 분석을 통해 확인하였으나, 이때 공감 과정이 어떻게 나타나는지, 그리고 음악에서의 공감 과정이 사회적 상황에서의 공감 과정과 어떠한 연관성이 있는지 직접 확인한 연구는 부족하다. 본 연구에서는 음악 연주와 사회적 상황에서 관찰자 36명이 대상(연주자/화자)에 공감하는 과정을 공감 정확도(empathic accuracy)와 심박수 동기화를 통해 확인하고, 각각의 상황에서 관찰자에게 주어지는 감각 정보와 정서가에 따른 공감 반응을 비교분석하고자 하였다. 다수준 상관관계 분석 결과, 전반적으로 음악에 대한 공감 정확도가 높을수록 사회적 상황에서의 공감 정확도 또한 높은 경향이 나타났으며, 특히 시각 정보만 주어졌을 때 두 상황 간 연관성이 나타났다. 또한 반복측정 분산분석을 통해 사회적 상황에서의 공감 정확도는 음악 상황보다 유의미하게 높으며, 두 상황 모두 청각 정보가 주어진 상황이 시각 정보만 주어진 상황에 비해 높은 공감 정확도를 나타내는 것을 확인하였다. 본 연구는 음악과 사회적 상황 간 공유하고 있을 공감 처리 메커니즘을 확인하였으며, 이를 통해 음악 교육의 효과가 전이되어 사회적 상호작용 능력 향상에 도움을 줄 가능성을 기대해볼 수 있다.

핵 심 낱 말 음악과 감정, 공감, 생체 반응 동기화, 감각 정보, 정서가

Abstract

Empathy is a crucial element in human societies, playing a central role in interpersonal communication and fostering social bonds. Recent studies have explored shared emotional experiences in music performance by analyzing physiological synchrony, but there is a gap in research directly examining the continuous empathic process and its relationship between music and social situations. This study aims to investigate empathic processes in music and social situations by examining empathic accuracy and heart rate synchrony across different modalities (visual-only/audio-only/video-and-audio) and valence (positive/negative). Results from a multilevel correlation analysis indicate a positive association between empathic accuracy for music and social situations, particularly when only visual information is presented. Repeated measures ANOVA also reveals that empathic accuracy in social situations is significantly higher than in music situations, and in both contexts, auditory information is associated with higher empathic accuracy than visual information alone. The study identifies shared empathy processing mechanisms between music and social contexts, suggesting potential transfer effects of music education on improving social interaction abilities.

Keywords Music and emotion, Empathy, Physiological synchronization, Modality, Valence

Contents

List of Tables

List of Figures

Chapter 1. Introduction

Empathy is a crucial component of interpersonal relationships and communication in human society, based on an understanding of the emotions of others. Its importance extends beyond individual one-on-one personal conversations (Blanke et al., 2016) to group communication (Stürmer et al., 2006). In a variety of social settings where interpersonal interactions take place, empathy is processed in real time, and its multifaceted nature results in responses through various channels, including cognitive, emotional, physiological, and neurological.

Recent studies have shown that high empathy enhances synchrony between individuals' behavioral, physiological, and neural responses (Finset and Ørnes, 2017; Dor-Ziderman et al., 2021; Imel et al., 2014; Tabak et al., 2023; Jospe et al., 2020; Zaki et al., 2009). In other words, the more a person interacts with another person, the more similar their response patterns are to a given stimulus. This phenomenon is observed not only in dyadic interactions but also in diverse contexts of social events, such as music concerts, where empathy and social cohesion are fostered through music. Researchers have found synchronized neural activation patterns among audiences experiencing collective pleasure at musical performances, as well as between performers and audience (Ara and Marco-Pallarés, 2020; Hou et al., 2020).

Music is an effective source to trigger strong emotional responses. Shared emotional experiences by engaging in joint musical activities, it is also efficient to enhance empathy and foster prosocial behaviors (Kirschner and Tomasello, 2010; Wu and Lu, 2021). From an evolutionary perspective, these functions of music can be explained by the standpoint that music has come together to promote group cohesion and affiliation in human societies (Cross and Morley, 2009; Savage et al., 2021). In the domain of neuroscience, the concept of mirror neuron systems explains that the neural activity of individuals engaged in music together becomes synchronized through the mimicry of each other's behaviors, leading to an aesthetic experience (Molnar-Szakacs and Overy, 2006).

While there is a possibility for common neural mechanisms between empathy in music and socio-emotional contexts, there is a lack of empirical studies that directly compare the processing of state empathy in both domains. This gap exists primarily because the majority of research on music and empathy has focused on empathy as a trait (Wallmark et al., 2018; Kawakami and Katahira, 2015; Vuoskoski and Eerola, 2012; Stupacher et al., 2022). A study by Wöllner, 2012, which explored the time-series relationship between a performer's expressed emotion and the audience's emotion recognition, is constrained by the limitation that participants were repeatedly shown the same music performance video under multiple conditions.

Most recently, Tabak et al., 2023 conducted a comparison of real-time empathy processing in music and society using the Empathic Accuracy (EA) task paradigm (Zaki et al., 2008; Ickes, 1993). In this task, participants were exposed to both the recordings of piano improvisation and the videos of storytelling, in which the person in the video talks about emotional experiences from their autobiographical memories. They were simultaneously instructed to provide behavioral responses indicating their real-time empathic reactions to the presented targets in sound sources/videos. Overall, the researchers observed a moderate, positive association between empathy in both music performance and dyadic interactions. Notably, there was a robust relationship with positive valence stimuli. This study showed an initial exploration of behavioral evidence for the association of empathy processing in music and society. However, a limitation

of this study was that differences between modalities were unable to be identified, as music was only presented in audio format, and storytelling was only presented as an audiovisual stimulus.

In the present study, our objective is to partially replicate the investigation conducted by Tabak et al., 2023, aiming to explore whether continuous empathic responses to musical performances are linked to empathic processes in dyadic interactions. Furthermore, we seek to identify the variables associated with these connections based on emotional valence (positive, negative) and modality (video-only, audio-only, and video-and-audio). To achieve this, we will employ the EA task paradigm, collecting real-time behavioral and physiological empathy responses through electrocardiogram measurements and analyzing them in a time domain.

The EA task will be divided into two phases: 1) capturing a video of a *target* expressing emotions through music performance or storytelling and collecting their emotional data, and 2) presenting the video to an *observer* who will continuously infer the emotions of the *target* in the video. This will allow for a comparison of the emotional data between the *target* and the *observer*. Additionally, we will explore whether individual variables are correlated with empathy.

The research questions for this study are as follows:

- RQ1. Are individuals who demonstrate high empathy in socio-emotional situations also proficient in understanding emotions in music contexts?
- RQ2. Does the degree of empathy vary based on the emotional valence expressed in emotional situations or the sensory information provided to the observer?
- RQ3. Which modality (visual/auditory) plays a primary role in inferring emotions in natural emotional situations (music/social)?
- RQ4. Are individual factors (e.g., interoception awareness) associated with empathy at the state level?

Chapter 2. Literature Review

2.1 Empathy

Empathy is a multidimensional construct (Decety, Lamm, et al., 2006; De Vignemont and Singer, 2006), commonly discussed in terms of cognitive empathy and emotional empathy. Each type of empathy is known to operate through distinct neural mechanisms (Shamay-Tsoory, 2011). *Cognitive empathy*, often referred to as 'mentalizing' or 'theory of mind,' involves the ability to accurately infer other people's thoughts, feelings, and intentions, understanding them from their perspective within a given context (Ickes, 2009; Shamay-Tsoory, 2011). Empathic Accuracy (EA), a form of cognitive empathy, involves the precise identification of a target's internal states in real-time at the state level, contrasting with trait-level measured by self-reports. EA plays a crucial role in interpersonal communication, stemming from the concept originally grounded in guidance for therapists in clinical settings. It regards the therapist's attitude to establish a healthy trust relationship, or 'rapport,' with clients (Rogers, 1957). In addition, deficits in EA have been identified in clinical populations such as schizophrenia or autism spectrum disorders, providing evidence of a relationship between EA and social interactions (J. Lee et al., 2011; McKenzie et al., 2022; Rum and Perry, 2020).

Experimental tasks designed to measure EA ability typically employ a paradigm in naturalistic settings, initially conceived by Ickes, 1993 and further developed by Zaki et al., 2008. In this paradigm, an *observer* views a video in which a *target* discusses an emotional event they have experienced. The observer is then engaged in a continuous response task, inferring the target's emotions from the video. A behavioral EA score can be calculated by analyzing the association between the target's self-reported emotional time-series data and the observer's inferred time-series data.

On the other hand, *emotional empathy* is the capacity to experience and understand the emotions of others (Kaplan and Iacoboni, 2006). *Affective sharing* (AS), identified at the state level (Singer and Lamm, 2009; Tabak et al., 2023), involves vicariously experiencing the internal states of others. The mirroring of facial expressions or gestures through the mirror neuron system, particularly involving sensory information, plays a pivotal role in AS. A study by Nummenmaa et al., 2008 measured brain activity during EA and AS and found significantly higher activation in the motor cortices, mirror neuron system, and thalamus during AS compared to EA. Several empirical studies also consistently support the observation of neural or physiological interpersonal synchrony during the experience of AS (Dor-Ziderman et al., 2021; Peng et al., 2021; Golland et al., 2015; Bruder et al., 2012). At the behavior level, AS can be measured by an evaluation of an observer's perception of their own felt emotion(Tabak et al., 2023).

2.2 Emotional Processes in Music

Despite the lack of research on empathic accuracy (EA) and affective sharing (AS) in music, both concepts are likely to play an important role in emotional processing mechanisms in music. When listening to music, the listener's body initially generates sensations in response to the acoustic features of the music. The listener then forms an emotional *perception* of what the music is expressing, representing an *external locus* of emotion. Subsequently, the listener's *felt* emotion, an *internal locus* of emotion

(Schubert, 2013; Gabrielsson, 2001). Aesthetic experiences, including judgments about preferences such as whether music is good or bad, have been proposed as high-level processing occurring after the emotional response (Juslin, 2013).

Recognizing emotions expressed in music is related to empathy, which is the ability to accurately infer and understand emotions. It is important to check empathic responses at this stage since the degree to which a listener empathizes with the musical emotion can influence their subsequent emotional experience during musical activities (Miu and Balteş, 2012). While accurately inferring emotions expressed in music may not be a necessary skill in everyday music listening, situations where the listener directly interacts with the performer (e.g., in a concert hall) may place a premium on the ability to empathize with the emotions conveyed by the performer. This empathetic connection could significantly impact the aesthetic experience, fostering a stronger bond between the performer and the audience and enhancing the sense of immersion in the performance.

Indeed, a study investigating the relationship between EA and AS abilities in both musical and social contexts (Tabak et al., 2023) discovered that both abilities are common in both domains. The researchers conducted on- and offline experiments with a large U.S. sample, revealing robust findings indicating that the more accurate the observers' inferences of the performer's emotions were, and the more consistent the observers' feelings were with the performer's, the higher the degree of empathy for the storyteller.

Moreover, evidence for cardiac synchronization has been observed among listeners who share similar emotional experiences while listening to music, providing support for the notions of social bonding and physiological synchronization facilitated by music (Czepiel et al., 2021; Bernardi et al., 2017). Consequently, we anticipate a heightened level of synchronization between performers and audiences when they interact by sharing the same emotions through affective sharing.

2.3 Emotional Valence and Modality Interactions

Empathy exhibits variability based on the valence (positive/negative) of encountered emotional stimuli. While the brain's mechanisms for processing pleasurable (positive valence) or aversive (negative valence) emotional stimuli share some similarities (Lindquist et al., 2016), there is a suggestion that they may utilize different pathways (Hayes et al., 2014; Nummenmaa and Calvo, 2015). Biased attention towards specific valences throughout developmental trajectories has been consistently observed (Kauschke et al., 2019; Goh et al., 2016; Feyereisen et al., 1986; Nasrallah et al., 2009), and behaviorally measured empathic accuracy has been demonstrated to vary depending on stimulus valence (Zaki et al., 2009; Tabak et al., 2023).

Additionally, the affective processing of stimulus pleasantness may be influenced by the modality presented to the observer. Modality-specific affective processing posits that the processing of certain emotions is more dominant depending on the sensory system involved. The study by Shinkareva et al., 2014 provides evidence for modality-specific processing by comparing modality-specific and modality-general processing hypotheses, utilizing functional magnetic resonance imaging (fMRI) to measure responses to visual and auditory stimuli. Meta-analyses and reviews of various neuroscience studies, such as Satpute et al., 2015 and Miskovic and Anderson, 2018, have further supported the modality-specific hypothesis.

It is well-established that auditory cues play a significant role in accurately decoding others' emotions, particularly when watching storytelling videos (Jospe et al., 2020; Hall and Schmid Mast, 2007; Gesn and Ickes, 1999). Nonetheless, nonverbal information may also influence empathic accuracy to

some extent, as studies have indicated increased accuracy when verbal cues are added to nonverbal cues and integrated (Zaki et al., 2009; Hall and Schmid Mast, 2007). Regarding physiological responses that may reflect the process of affective sharing, a study by Jospe et al., 2020 analyzed differences across three modalities: video-only, audio-only, and video-and-audio. They found that heart rate synchrony between the target and observer was significantly higher when only visual information was presented.

While the effects of modality on the ability to precisely recognize a performer's expressed emotions in music performance have not been extensively explored, facial expressions have been identified as more influential than audio information (Livingstone et al., 2015). Most studies have discussed the association between listeners' emotional experiences and *visual dominance* (E. Coutinho and Scherer, 2017; Vuoskoski et al., 2014; Vines et al., 2011; Vines et al., 2006). In the present study, we anticipate an interaction between modality differences and valence bias processing, hypothesizing the emergence of context-dependent modalities for both music performance and storytelling.

2.4 Interoception and Empathy

Interoception awareness (IA) refers to the ability to sensitively perceive changes in internal bodily states and is linked to self-representation and embodiment through body awareness (Craig, 2010; Herbert and Pollatos, 2012). As empathy involves blurring boundaries between self and others, encompassing the sharing of feelings and understanding another person's emotions from their perspective, it has been suggested that empathy is closely connected to IA, known to modulate the self/other distinction (Tajadura-Jiménez and Tsakiris, 2014). IA is often assessed through tasks where individuals mentally count their heartbeats (Schandry, 1981). In a study by Imafuku et al., 2020, the relationship between heartbeat counting accuracy and the frequency of spontaneous facial mimicry during eye contact was explored. The findings indicated that higher IA was associated with a greater frequency of facial mimicry in social contexts, providing evidence for an association between IA and affective sharing experience based on the mirror neuron system.

However, Baiano et al., 2021, in a review of various studies on IA, argues that IA is associated with perspective-taking ability or cognitive empathy. Additionally, a study involving individuals with autism spectrum disorder suggested the possibility that IA is indirectly related to empathy (Mul et al., 2018). Nevertheless, contradictory results have been reported. For example, a study encompassing various empathy-related tasks, including self-reports of empathy, affective sharing, and perspective-taking (Ainley et al., 2015), did not find any significant association between IA and empathy. Building on previous research, this study aims to reexamine the connection between IA and empathy abilities, employing a distinct experimental paradigm to measure empathy.

Chapter 3. Method

The experiment comprised two phases: 1) a video-recording session to collect data from the *targets*, and 2) a video-watching session to obtain *observers*' empathic accuracy and cardiac responses. Initially, the targets participated in recording either a piano improvised performance or a storytelling session based on their autobiographical memories. They subsequently watched their own videos, assessing their emotional states during video recording while wearing electrocardiography (ECG) electrodes. In the next phase, observers— a distinct group from the targets— continuously rated the inferred emotions of the targets presented in the videos. ECG responses were also recorded during the observers' ratings, adopting procedures similar to those of the targets.

All procedures followed the Empathic Accuracy Task (EAT) paradigms outlined in previous literature (Zaki et al., 2008; Jospe et al., 2020; Tabak et al., 2023). The Korea Advanced Institute of Science and Technology IRB approved the study design.

3.1 Phase One: Targets' Data

3.1.1 Participants

Participants were recruited for two different types of recording, music performance ($n = 8$; female 4; M age = 26.88, $SD = 1.73$) and storytelling ($n = 9$; female 5; M age = 25.56, $SD = 2.79$). Musicians were undergraduate students currently enrolled in or holding a bachelor's degree in Western classical music composition (M years of formal musical training = 14.13, $SD = 5.72$; M years of piano experience = 10.25, $SD = 3.28$). Participants for storytelling were recruited via the university's online bulletin board. As this study was conducted in the Republic of Korea, all participants were recruited exclusively from the Korean population.

The inclusion criteria were as follows: 1) participants aged 20 to 30 with normal vision and hearing (to match the similar age range of the university student participants for phase two), 2) no impairments in daily life due to discomfort in hand movements, 3) no history or current presence of neurological/psychiatric disorders, 4) not currently under medication for neurological/psychiatric conditions, and 5) experience in improvised performances (music condition only). All participants provided informed consent for video recording and the use of their video in subsequent studies.

3.1.2 Video Acquisition

Participants were instructed to prepare semi-improvised piano performances (music) or storytelling of autobiographical memories (social), expressing three basic emotions: joy/happiness, sadness, and anger. These emotions were chosen to replicate the earlier study by Tabak et al., 2023, which investigated empathic accuracy in the interplay between music and social situations. Drawing on findings from prior studies indicating that listeners can discretely perceive these emotions in musical performances (Akkermans et al., 2018; MacGregor et al., 2023), this study recorded videos specifically for joy/happiness, sadness, and anger. These emotions were later separated into positive (joy/happiness) and negative (sadness and anger) valence.

In the musical context, performers were instructed to convey their emotions, incorporating non-verbal expressions. For storytelling, participants were encouraged to freely narrate their personal experiences as if sharing a story with a friend. All storytelling was conducted using Korean honorifics, and participants had access to a summarized script for reference during the recording.

The video recordings featured participants' upper bodies, seated in front of the camera against a black background. Piano performance shots included footage of the keyboard. To minimize the impact of clothing, participants were instructed to wear comfortable white or gray tops. Each video, lasting between 1 to 2 minutes, was recorded in a randomized sequence. A stopwatch on the left monitor facilitated timing, and participants could opt to re-record if desired. All recordings were made using Canon's EOS 5D Mark IV Full Frame DSLR camera with EF 24–105 mm f/4L IS II USM zoom lens set at 50 mm. Piano performances were accompanied by Casio's CDP-120 Portable Digital Piano, utilizing its built-in speakers. The video was acquired with the format of mp4, specification of 1920 x 1080, 25.00 fps, and an audio sampling rate of 48 kHz.

3.1.3 Video Selection

A total of 24 music clips and 36 storytelling clips were produced. For the later experiment with observers, which required repeating each condition twice, it was necessary to select six videos for each pair of situations (music, storytelling) and emotion (joy/happiness, sadness, anger) (see Table 3.1). Five individuals, including researchers and colleagues, participated in the stimulus selection process. The criteria for selection included 1) ensuring that each video visually and auditorily expressed the intended emotion (e.g., if the video aimed to convey joy, it should not appear excessively expressionless or calm) and 2) confirming the absence of any noise that could interfere with watching the video (e.g., nail sounds during piano playing). Additionally, considering the potential impact of gender on empathy, an equal distribution of male and female participants was maintained for each emotion category (e.g., 3 female targets out of 6 music-joy videos). In the end, 18 music performance clips (identity: 3 females) and 18 storytelling clips (identity: 4 females) were selected. Since there were no restrictions on using all three videos of the same individual, the identity of individuals in the videos comprised a total of 14 people.

3.1.4 Self-Report Measures

The emotional assessment comprised two parts: an immediate summary rating right after each video recording and a continuous rating after completing all recordings. Participants were instructed to assess their emotions during the video recording on a 9-point scale in both evaluations. All evaluations were carried out using the Psychopy software.

In the initial rating, emotions during filming were evaluated in three dimensions, arousal (1:*calm* - 9:*excited*), valence (1:*negative* - 9:*positive*), and dominance (1:*weak* - 9:*strong*). The second evaluation took place while participants watched the recorded videos. Using keyboard arrow keys, participants responded to the valence scale (1:*very negative* - 9:*very positive*) based on their emotions during recording. At the start, a red circle was positioned above 5 points, and participants moved it left towards 1 point for more negative emotions or right towards 9 points for more positive emotions. Recordings were made at a sampling rate of 2 Hz (two timestamps per second), with each level consisting of 20 notches (e.g., intervals between 1-2 points divided into 20 steps).

Table 3.1: Types and number of video stimuli selected for the observer experiment. For each pair of emotional situations, six videos were selected, consisting of three male and three female targets. Considering the additional *modality* factor to be introduced in Phase Two, 18 videos for both music performance and storytelling were chosen based on the experimental design to present two videos (one male and one female) for each condition. This decision aligns with the balanced Latin square design of this study. Each observer viewed different videos across all conditions, resulting in a total of 36 videos. Since there were 36 participants in Phase Two, each video was presented 12 times per condition across all participants. For instance, the *music-joy* video from *target01* was presented 12 times in the video-only condition, 12 times in the audio-only condition, and 12 times in the audiovisual condition. VO = video-only; AO = audio-only; VA = video-and-audio.

	Emotion	Number of stimuli				Total
		Total	VO	AO	VA	
Music	Joy/Happiness	6 (m3/f3)	2 (m1/f1)	2	2	18 videos (m9/f9)
	Sadness	6	2	2	2	(identity: m4/f3)
	Anger	6	2	2	2	
Social	Joy/Happiness	6	2	2	2	18 videos (m9/f9)
	Sadness	6	2	2	2	(identity: m3/f4)
	Anger	6	2	2	2	

3.1.5 ECG Data Acquisition

Electrocardiographic signals were recorded at a sampling rate of 500 Hz during continuous emotional ratings and before the ratings to acquire baseline ECG data. The physiological responses of the targets were measured using a Compumedics-Neuroscan SynAmps RT system with Curry 8 data acquisition software. The setup included a Cedrus StimTracker connected between Psychopy and Curry 8. The modified limb lead II configuration was used to measure ECG responses, with three disposable electrodes placed at the right ankle (ground), left ankle (+), and right collar bone (-). The impedance of all electrodes was kept below 40 kΩ.

3.1.6 Procedure

Participants were instructed in advance to prepare musical performances or share autobiographical experiences expressing three distinct emotions at least 3 days before their scheduled laboratory visit. Upon arrival, participants completed consent forms, received instructions regarding the experimental procedure, and confirmed their consent for the subsequent utilization of recorded videos.

The target session consisted of three phases (see Figure 3.1 for the overall procedure). Initially, participants had a 15-minute preparation time before video recording, during which musicians practiced the piano, and storytellers organized their thoughts by making notes on paper to recall events. After the preparation, participants confirmed the randomly assigned recording sequence and commenced the filming stage. Following each recording, they participated in a summary rating process, assessing the emotions experienced and the flow states during filming (e.g., sadness recording - rating - joy/happiness recording - rating - anger recording - rating) (note that the flow states score was not reported in this paper). To prevent emotional carryover between recordings, participants were instructed to take a

minimum 5-minute break between sessions to refresh their emotions. Finally, participants, wearing ECG electrodes, underwent a 3-minute baseline measurement. Subsequently, they viewed their three videos in a random order, providing real-time continuous emotion evaluations. Similar to the recording phase, a minimum 30-second break was provided between each video for emotional refreshing. With the exception of the researcher's explanation, all procedures occurred with participants alone in the soundproof booth.

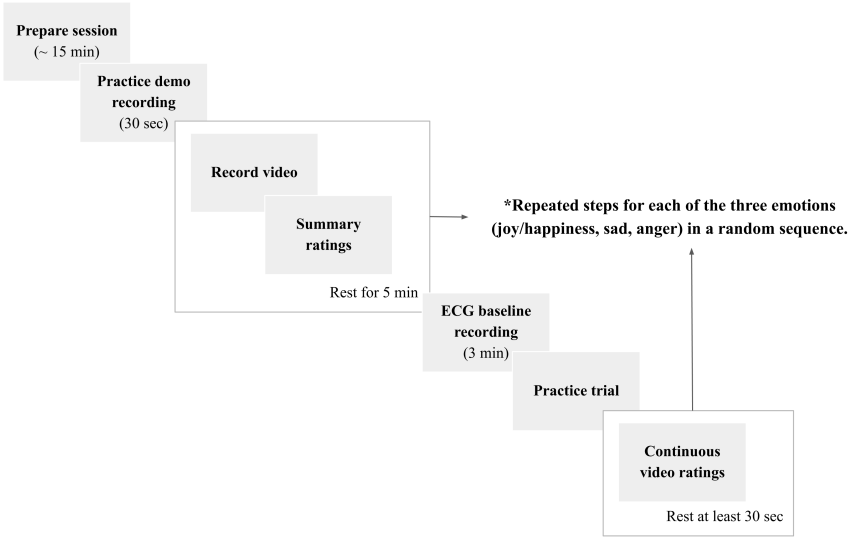

Figure 3.1: The overall procedure of targets' video recording session. Before recording the practice videos for data collection, participants were given 30 seconds to freely perform music or share a story in front of the camera to practice playing or speaking. These practice videos were later used as real-time emotion evaluation practice videos during the evaluation sessions.

3.2 Phase Two: Observers' Data

3.2.1 Participants

Participants were members of the university, recruited via e-mail and online bulletin board. The recruitment was conducted within a separate group from Phase One. To implement a Balanced Latin Square design, we initially recruited 36 participants, aiming for a multiple of the total conditions (3 *emotions* x 3 *modalities*) to be examined in this study. However, due to ECG recording issues during the experiment, one participant's data was excluded, and an additional participant was recruited to maintain the desired balance. In the end, we collected data from a final sample of 36 participants ($n = 36$; female 17; M age = 26.06, $SD = 3.56$). Only native Korean speakers aged 20 or older with normal vision and hearing were recruited. This criterion was essential as participants needed to watch videos presented in Korean across various modalities. Additionally, all participants confirmed the absence of hand movement disabilities, reported no diagnosed neurological or psychiatric conditions, and stated that they were not currently using any medications for psychiatric purposes.

3.2.2 Stimuli

The stimuli for music performance and storytelling consisted of 18 videos each, with six videos per emotion (joy/happiness, sadness, and anger). For all videos used in the experiment, there was no significant difference in the target's physiological responses between positive and negative valence, and behavioral self-report valence scores, assessed immediately after the video recording, were rated significantly higher for positive valence videos than negative valence videos (joy/happiness-sad: $p < .001$; joy/happiness-anger: $p < .001$).

For each emotion, there were two videos (one male and one female) for each of the three modalities: video-only (VO), audio-only (AO), and video-audio (VA). Each video had distinct content, and participants viewed all 36 videos once, with no repetition. This study employed a balanced Latin square design to ensure that each participant watched a unique video in every condition. In total, all 36 videos, including both performances and storytelling, were presented to 36 participants, 12 times in each modality, resulting in 36 presentations (refer to Table 3.1).

Stimuli were presented in an order determined by a balanced Latin square design, and the music and social sessions were conducted on separate days. Both the music and social sessions were counterbalanced across participants. The duration of all videos ranged between one and two minutes. Using Adobe Premiere Pro, they were converted to 30 fps, H.264 codec, and 720 x 480 files. The audio was encoded in AAC format with specifications of 44.1 kHz, 320 kbps, and stereo. Loudness was normalized in accordance with the broadcast standard ITU-R BS.1770-3. In the case of audio-only (AO) videos, they were presented on a consistent black screen with audio alone.

3.2.3 Self-Report Measures

Emotional Assessments

Participants engaged in the Empathic Accuracy Task, a real-time continuous emotion assessment, where they inferred the emotions the target was expressing/experiencing, while they were watching the video (refer to Figure 3.2). The real-time emotional assessment mirrored the target's session in Phase One. Observers were instructed to rate the emotional state of the person in the video on a scale from 1 ('*very negative*') to 9 ('*very positive*') by moving a red dot along a scale with 20 notches between each level. Observers adjusted the dot whenever they perceived a change in the target's emotion. These ratings were recorded at a sampling rate of 2 Hz (timestamps at every 0.5s).

Following video viewing, observers completed questionnaires assessing their own psychological state. The questionnaires included five items evaluating emotions in three dimensions: arousal (1:*calm* - 9:*excited*), valence (1:*negative* - 9:*positive*), and dominance (1:*weak* - 9:*strong*). Additionally, participants rated the degree of flow during video watching ('I was completely immersed in the video'; 1: *not at all* - 9: *very much*) and the level of empathy towards the emotions of the person in the video ('I empathized with the emotions of the person in the video'; 1: *not at all* - 9: *very much*), all on a 9-point scale. This evaluation was conducted to minimize the impact of the observer's emotions on subsequent videos, although it was not reported in this study.

Empathy Abilities

To assess participants' empathy levels, two self-report empathy questionnaires were administered: the Empathy Quotient (EQ; Baron-Cohen and Wheelwright, 2004) and the Interpersonal Reactivity

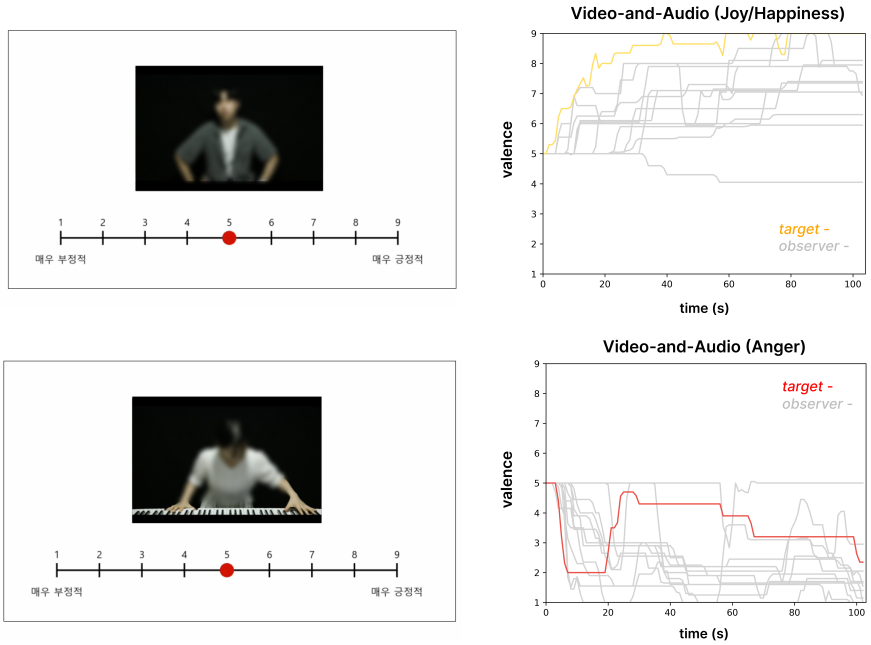

Figure 3.2: The screenshots of the Empathic Accuracy (EA) task and the responses of the target and observers (n = 12): In the *top left*, there is an example of a storytelling video depicting the experience of joy/happiness emotions. The *bottom left* video features the target expressing anger through an improvised piano performance. The *right* graph displays the responses of continuous emotional ratings in VA conditions.

Index (IRI; Davis, 1983). The Korean version of the EQ, developed by J. Kim and Lee, 2010 for reliability and validity, was utilized. The EQ includes three components: cognitive empathy (EQ-CE), emotional reactivity (EQ-ER), and social skills (EQ-SS) (Lawrence et al., 2004). Among the four aspects of the IRI, Empathic Concern (IRI-EC) and Perspective Taking (IRI-PT) were the only components examined in this study. The researcher translated 14 items into Korean.

Musical Experiences

The ability to perceive emotions expressed in music performance videos may be impacted by the level of musical education. To investigate potential connections between musicians' empathy levels and their musical education, we collected additional information on participants' years of formal music education, duration of piano study (year), and the average hours spent practicing their instrument each week.

3.2.4 Interoception Task: Heartbeat Counting

To assess interoceptive ability, participants engaged in a task involving the counting of their own heartbeats. The Heartbeat Counting Task, a commonly employed paradigm for measuring interoceptive accuracy (Dale and Anderson, 1978), was implemented following the procedure introduced by Garfinkel et al., 2015 and adapted by Legrand et al., 2022. Participants were instructed to place their hands on their knees in a comfortable seated position. Upon the initiation of the task, indicated by the 'start' sound and a displayed image on the monitor, participants focused on their bodily sensations, counting their heartbeats without physically checking their pulse. After the image disappeared and the 'stop' sound

was heard, participants entered the counted heartbeats via the keyboard and rated their confidence in the count on a 7-point scale. This process constituted a single trial, and a total of six trials with different time intervals (25, 30, 35, 40, 45, and 50 seconds) were conducted in a randomized order. Following a practice session of 10-second time intervals, the actual trials started, and no feedback on the actual heart rates was provided. All procedures were conducted using the Psychopy program, and participants wore ECG electrodes during the heartbeat counting task to verify the actual heart rates.

The accuracy score for each trial was calculated with the following formula (Legrand et al., 2022; Garfinkel et al., 2015):

$$1 - \frac{|n_{real} - n_{reported}|}{\frac{|n_{real} + n_{reported}|}{2}}$$

The resulting scores for the six trials were averaged to calculate the interoceptive ability index for each participant.

3.2.5 ECG Data Acquisition

The ECG was recorded during both the Heartbeat Task and the real-time emotion evaluation session while participants watched videos. Consistent with the target data in Phase One, three disposable electrodes were attached to the right ankle (ground), left ankle (+), and right collar bone (-) based on the modified limb lead II system. Raw data acquisition was conducted at a sampling rate of 500 Hz using the Neuroscan SynAmps and Curry 8 software. The impedance of all electrodes was maintained below 40 kΩ.

3.2.6 Procedure

The experiment was conducted over two separate days. On day one, participants completed the experiment consent form, musical background questionnaire, IRI, and EQ tests. Following this, they engaged in one of the Empathic Accuracy (EA) tasks, either the music or social session. On the second day, participants underwent the interoception task, followed by the remaining EA task. Each day of the experiment lasted around 1 hour, totaling 2 hours, with an average gap of 3.70 days between the two sessions (ranging from 1 to 12 days). This gap aimed to prevent an excessive cognitive load on participants involved in emotion inference tasks. The order of the two EA tasks (music and social sessions) was counterbalanced.

On the first day, participants completed self-report tests using Google Forms. Afterward, they entered a soundproof booth to attach ECG electrodes. While wearing headphones and maintaining a relaxed state, participants focused on a fixation cross on the monitor for a 3-minute baseline measurement. Following this, participants practiced once to infer emotions from videos and then performed the task for 18 video clips, divided into three blocks of six videos. The practice videos consisted of neutral emotion of 30-second stimuli, designed to avoid evoking emotions (music: a video of playing three different musical scales consecutively; storytelling: a video describing the appearance of one's room). These practice videos featured a fellow researcher not present in the Phase One videos. Before starting the EA task, participants were reminded to assess the target's emotions, not their own. After watching each video clip, they answered five questions about their psychological state and were allowed breaks between videos and blocks.

On the second day, participants were explained the interoception task, wore ECG electrodes, and performed the heartbeat counting task within the soundproof booth. Subsequently, a 3-minute baseline

measurement was conducted with participants via headphones. The following EA task followed the same procedure as the first day. For storytelling videos, participants were encouraged to maintain confidentiality due to the personal nature of the target's stories.

3.2.7 Data Analysis

Behavioral Data

In the analysis, the continuous emotional rating data of the target collected in Phase One and the time-series emotional valence rating data collected during the Empathic Accuracy (EA) task were preprocessed using the same procedure. Firstly, to maintain time points at regular intervals for data with a sampling rate of 2 Hz, interpolation was applied, followed by downsampling to 1 Hz (via Python's *scipy* package). Due to instances of awkwardness at the beginning and end of most target videos, 2 seconds from the start and end of the emotional data were removed for analysis.

To assess how accurately observers inferred the emotions of the target, cross-correlation coefficients (Pearson) between the emotional data of the target and the observer were calculated for each video (RStudio, *tseries* package, 'ccf' function). As the target might react faster to their own emotions while watching pre-recorded videos, cross-correlation was calculated within a lag window of 0 to 10 seconds, obtaining the maximal correlation (Jospe et al., 2020). To perform hypothesis testing on the coefficients' values as observer empathic accuracy scores, all r coefficients were Fisher's r-to-z transformed Pearson coefficients (rZ) for standardization (RStudio, *psych* package, 'fisherz' function). The average value of rZ across two repetitions of each condition for each participant was then used as the dependent variable in the analysis (see Figure 3.3 for examples of the highest and lowest EA score).

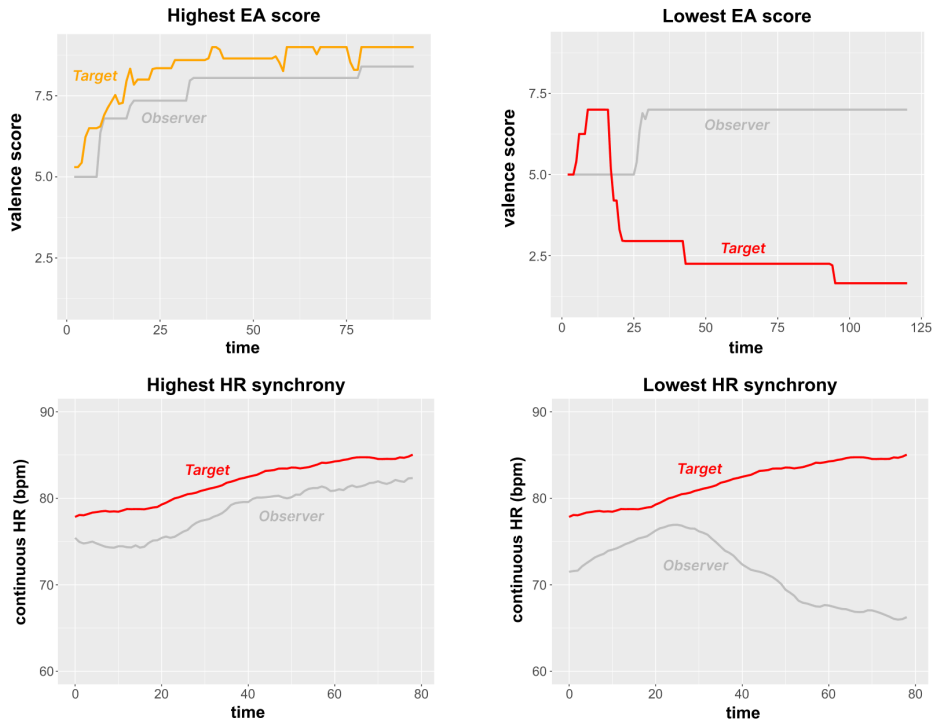

Figure 3.3: Examples of target and observer's ratings. The highest and lowest EA scores (*top*) and the highest and lowest HRS (*bottom*) are drawn in the figure. *Orange* color = joy/happiness stimuli; *Red* color = anger stimuli.

ECG Data

The ECG data underwent preprocessing using the Curry 8 software, including baseline correction and the application of a bandpass filter with a range of 1-50 Hz. In the ECG data, r-peaks were detected by conducting decomposition into 5 levels with the 'sym4' wavelet in MATLAB R2023a, followed by peak detection using the 'findpeaks' function. This process was performed with a 30-second sliding window, and for concurrent analysis with behavioral data, the mean second-by-second heart rate (HR) was computed at a sampling rate of 1 Hz (Jospe et al., 2020).

The analysis process for calculating heartbeat synchrony mirrored the EA score. Firstly, the initial and final 2 seconds of the entire dataset were removed. Subsequently, considering a lag window between 0 and 10 seconds, the maximum cross-correlation coefficient between the target and observer was calculated. For hypothesis testing, Fisher z transform was employed to obtain rZ values, and the average values across repeated conditions were utilized as heartbeat synchrony scores (see Figure 3.3 for examples of the highest and lowest HRS).

Chapter 4. Results

To assess the validity of using the 36 target videos (18 music, 18 storytelling) for the Empathic Accuracy (EA) task, the average target-observer EA score (r coefficient) was examined for each video. For every video, 36 EA scores (12 each for VO, AO, VA) were provided. If the average r coefficient for the 36 scores was less than 0.10, the video was considered to inadequately or ambiguously express the intended emotion and was excluded from the analysis. Consequently, the anger video in the music condition, which not only had the lowest mean in the EA score but also the heart rate synchrony (HRS) mean value (EA score Mean r = 0.06, HRS Mean r = -0.040), was excluded.

The final analysis included 17 music videos (6 joy/happiness, 6 sadness, 5 anger) and 18 storytelling videos (6 joy/happiness, 6 sadness, 6 anger). Repetitive conditions within participants were averaged for statistical analysis, resulting in a consistent dataset of 36 participants, each contributing data for 2 conditions (music, social) x 3 emotions (joy/happiness, sadness, anger) x 3 modalities (VO, AO, VA) with 18 rZ values each. Additionally, as the focus of this study's analysis was on assessing EA and HRS based on emotional valence (not specific emotions), emotions were categorized into positive (joy/happiness) and negative (sadness, anger) (refer to Figure 4.1 for the visualized distribution of all data involved in the analysis). There were no significant correlations between EA and HRS scores (refer to Table in Supplementary material).

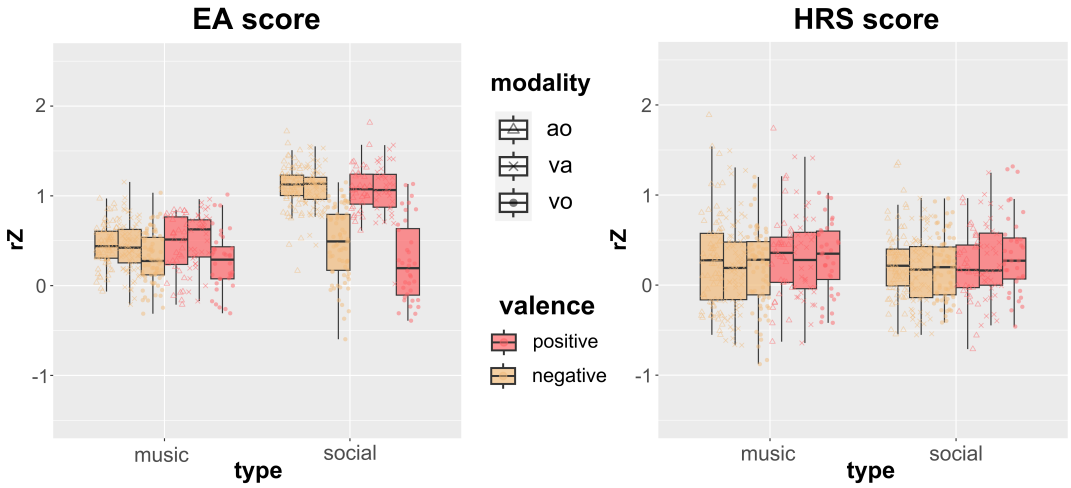

Figure 4.1: Boxplots for the distribution of all data (*left*: Empathic Accuracy score; *right*: heart rate synchrony score). y scale represents the value of rZ. Music (n = 324), Social (n = 324), VO (n = 216), AO (n = 216), VA (n = 216), Positive (n = 216), and Negative (n = 432).

4.1 Demographic Information

No significant age differences were observed between males and females ($t(33.80)$ = -0.29, $p = 0.777$). When comparing Empathy Quotient (EQ) scores, men exhibited significantly higher total EQ, Cognitive Empathy (EQ-CE), and Social Skills (EQ-SS) scores compared to women ($t(33.56) = 3.79$, $p < 0.001$;

$t(33.67) = 4.22$, $p < 0.001$; $t(33.93) = 2.94$, $p = 0.006$). No significant differences were found in EQ Empathic Reactivity (EQ-ER), Interpersonal Reactivity Index (IRI) scores and subscales, Interoceptive Accuracy (IA), years of music training (both formal and piano), and hours of practicing instruments per week. Demographic information for participants is presented in Supplementary material.

4.2 Correlation between Empathy in Music and Social Conditions

To explore the association between empathic accuracy (EA) scores in music and social contexts, multilevel correlation analyses (secondary units: video; primary units: observers)were conducted using the function 'multilevel' in the R package 'correlation' (Makowski et al., 2020). The specific coefficients and p-value for each condition are presented in Table 4.1. Notably, robust positive correlations were identified for both music and storytelling stimuli, with a coefficient of $r = 0.22$ and $p < 0.001$. This pattern persisted even when considering negative emotional videos ($r = 0.25$, $p < 0.001$). Furthermore, a near-significant positive correlation emerged for positive emotional videos ($r = 0.19$, $p = 0.050$). Upon examining different modalities, higher Empathic Accuracy in social situations, specifically when visual information was provided alone, was associated with increased accuracy in music scenarios ($r = 0.21$, $p = 0.029$). This finding underscores a particularly noteworthy connection between empathic accuracy in negative emotional contexts portrayed in visual-only videos across two distinct scenarios ($r = 0.30$, $p = 0.009$). Figure 4.2 illustrates the trends for the correlation of music and social empathic accuracy scores. There were no significant correlations observed in HRS between music and social situations.

4.3 Comparing Empathy in Music and Social Conditions

To examine differences in empathic accuracy (EA) scores (rZ) based on the emotional context and sensory information provided, we conducted a 3-way repeated measures ANOVA with dependent variable as EA score, involving *type* (music, social), *valence* (positive, negative), and *modality* (VO, AO, VA) as factors (see Figure 4.3). The analysis revealed significant 2-way interactions for *type* and *valence* ($F(1, 35) = 4.63$, $p = .038$, $\eta_p^2 = 0.118$), as well as *type* and *modality* after Greenhouse-Geisser correction (GGe $(2, 70) = 1.00$, $p < .001$). Post-hoc analyses with Bonferroni correction for significant interactions revealed higher empathic accuracy in storytelling compared to music for both positive and negative valence stimuli ($p < .001$; $p < .001$). In the post-hoc analysis of the type-modality interaction term, rZ scores for social storytelling were significantly higher than those for music performance in both AO and VA conditions ($p < .001$). However, in the VO condition, the rZ value for social storytelling was significantly lower than in the other AO and VA conditions for both music and storytelling ($p < .001$; $p = .006$). The difference between the VO EA scores for storytelling and music performance was not significant ($p = .168$).

The main effects of *Type* ($F(1, 35) = 284.43$, $p < .001$, $\eta_p^2 = 0.890$) and *Modality* ($F(2, 70) = 119.23$, $p < .001$, $\eta_p^2 = 0.773$) were found to be significant. In the Bonferroni post-hoc analysis, EA was higher in social situations than in musical ones ($p < .001$), and AO and VA conditions were found to be easier for inferring the target's emotions compared to the VO condition ($p < .001$). There was no significant difference in rZ between AO and VA conditions ($p = 1.00$). The 3-way repeated measures ANOVA with HRS rZ as an independent variable did not yield any significant results.

Table 4.1: Results of the multilevel or Pearson correlation between EA scores (rZ) in music and social condition. Multilevel correlations were conducted for data collected from the conditions repeated more than two times. VO = video-only; AO = audio-only; VA = video-and-audio; rep = number of repetitions within a participant; CI = confidence interval. $+p$ value near 0.05, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

			Modality			
			All	VO	AO	VA
		n	324	108	108	108
		rep	9	3	3	3
	All	r	**0.22*** **	**0.21***	-0.11	-0.07
		p value	**< 0.001**	**0.029**	0.253	0.496
		95% CI	[0.11, 0.32]	[0.02, 0.38]	[-0.29, 0.08]	[-0.25, 0.12]
		n	108	36	36	36
		rep	3	1	1	1
Valence	Positive	r	**0.19+**	0.13	-0.23	-0.19
		p value	**0.050**	0.461	0.168	0.278
		95% CI	[0.00, 0.36]	[-0.21, 0.44]	[-0.52, 0.10]	[-0.48, 0.15]
		n	216	72	72	72
		rep	6	3	3	3
	Negative	r	**0.25*** **	**0.30** **	-0.04	0.07
		p value	**< 0.001**	**0.009**	0.770	0.562
		95% CI	[0.12, 0.37]	[0.08, 0.50]	[-0.26, 0.20]	[-0.16, 0.30]

4.4 Modality Dominance

One of the research questions in this study aimed to determine which sensory modality, among visual-only (VO), auditory-only (AO), and visual-auditory (VA), predominantly contributes to empathic accuracy (EA) in naturalistic settings where diverse sensory information is utilized for emotion inference. To address this question, we examined the correlation between unimodal situations showing similar EA responses to multimodal VA situations. The analysis involved selecting models representing VA, VO, and AO responses for each of the 35 videos, with each modality having 12 observers. We computed the maximum cross-correlation coefficients (lag 0-10s) between VA and VO, and VA and AO models for each video, respectively (following the target-observer cross-correlation process), and standardized the coefficients using Fisher's z transform. Models representing each condition were selected by determining the median response of the 12 observers for each timepoint in every video (given the 1 Hz sampling rate). The difference between VA-VO similarity (rZ) and VA-AO similarity (rZ) obtained through this process was analyzed for significance using the Wilcoxon rank-sum test or Student's t-test (based on the results of the Shapiro-Wilk test). Any significant differences would imply that the modality in question is more likely to be used in multimodal situations (VA) compared to the other modality.

The comparison between the two groups revealed that, across all types and valences of stimuli, the auditory-only (AO) condition exhibited more similar responses to visual-auditory (VA) than the visual-only (VO) condition ($W = 968$, $p < 0.001$). This significant difference was observed even when stimuli were categorized into positive and negative valences ($W = 115$, $p = 0.012$; $W = 425$, $p < 0.001$).

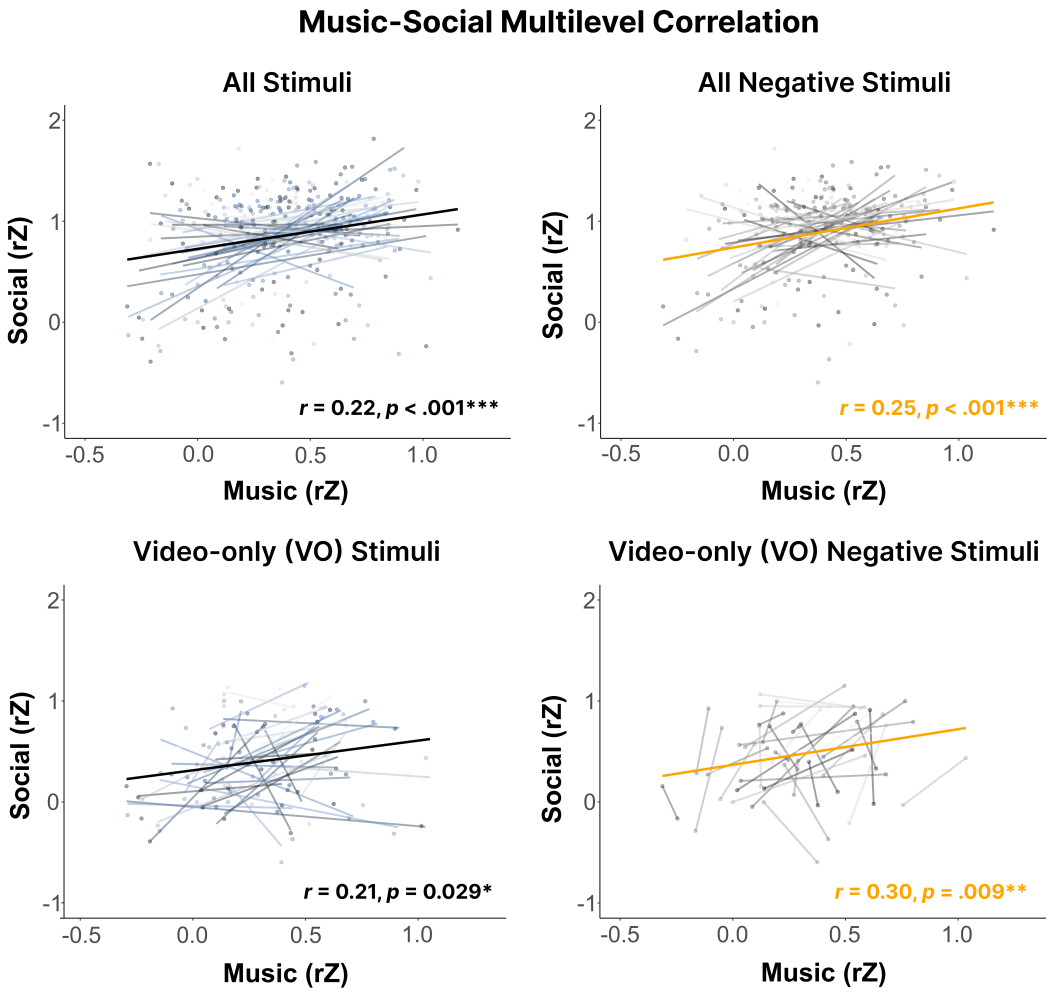

Figure 4.2: The results of multilevel correlation between music and social empathic accuracy scores (*top*: for all stimuli; *bottom*: for video-only stimuli; *left*: for all positive and negative stimuli; *right*: for only negative stimuli).

Similarly, when specifically examining social storytelling stimuli, a comparable pattern emerged when comparing the differences in VA-VO rZ and VA-AO rZ. For storytelling videos, the condition where only audio information was provided showed responses similar to audiovisual stimuli ($t = 5.41$, $p < 0.001$). Additionally, when stimuli were distinguished based on emotional valence, the AO condition consistently exhibited significantly higher similarity with VA compared to the VO condition (positive: $t = 3.44$, $p = 0.017$; negative: $t = 4.38$, $p < 0.001$). The statistical analysis results and visualized graphs are presented in Table 4.2 and Figure 4.4.

For HRS, a similar approach was employed to determine representative models for each modality. The differences in correlation coefficients between VA-VO and VA-AO were then analyzed. The results indicated that the synchrony between the observer group's heart rate when only audio was provided in social situations and the heart rate when both audio and visual stimuli were present was significantly higher than the HRS for VA-VO stimuli ($t = 2.30$, $p = 0.028$). As heart rate does not reflect differences based on valence, stimuli were not analyzed based on valence.

Figure 4.3: 3-way repeated ANOVA results: *type* (music performance, storytelling) x *valence* (positive, negative) x *modality* (AO, VA, VO). AO = audio-only; VA = video-and-audio; VO = video-only.

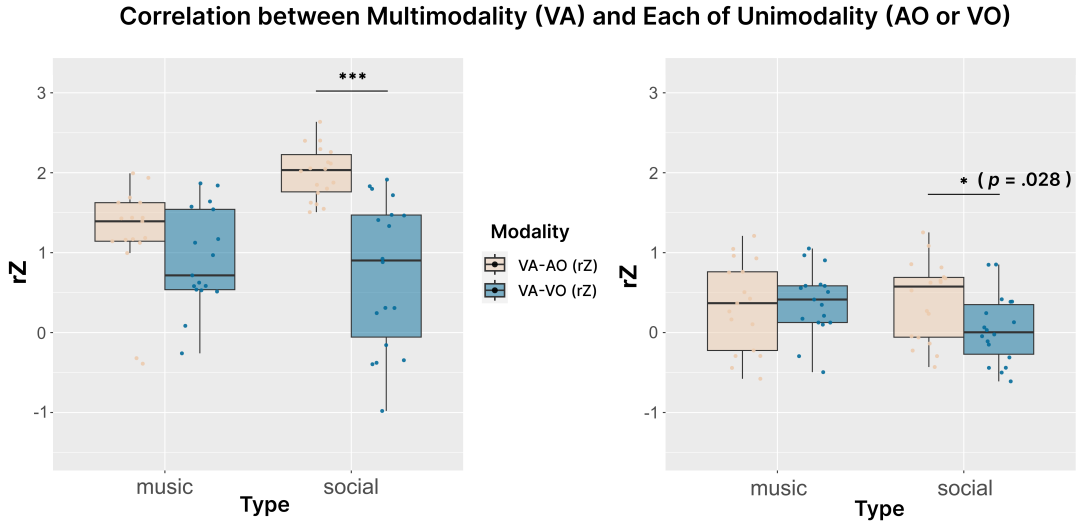

Figure 4.4: Similarity between the responses of multimodality (VA) and each of unimodality (AO or VO). *Left* indicates the boxplot of the correlation coefficients between modalities, resulting from behavioral assessments while *right* boxplot shows heart rate synchrony between modalities. VA = video-and-audio; AO = audio-only; VO = visual-only.

4.5 Individual Variables and Empathy

To examine the association between individual factors (EQ, IRI, IA, musical training) and EA score as well as HRS, Pearson correlation analyses were conducted for each variable with EA and HRS. While no significant correlations were found between all variables and EA scores, HRS showed significant static associations in some conditions. In the case of social storytelling presented as audiovisual stimuli, there was a tendency for higher HRS between target and observer when the observer had higher IRI scores (r

Table 4.2: Results of comparison between coefficients between VA-AO and VA-VO. The coefficients between multimodality, specifically VA, and unimodality, either VO or AO, were computed using the cross-correlation method. These coefficients were then transformed into z values through Fisher's r-to-z transformation. The comparison between the two methods, namely the Wilcoxon rank sum test or Student's t-test, was determined based on the normality test, specifically the Shapiro-Wilk test. EA = Empathic Accuracy; HRS = Heart Rate Synchrony. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

			Wilcoxon rank sum test or Student's t-test			
			All	Music		Social
EA score	All	n	35	17	n	18
		W	**968*** **	190	t	**5.41*** **
		p-value	**< 0.001**	0.122	p-value	**< 0.001**
	Positive	n	12	6	n	6
		W	**115***	25	t	**3.44***
		p-value	**0.012**	0.310	p-value	**0.017**
	Negative	n	23	11	n	12
		W	**425*** **	79	t	**4.38*** **
		p-value	**< 0.001**	0.243	p-value	**< 0.001**
HRS score	All	n	35	17	n	18
		t	1.37	-0.28	t	**2.30***
		p-value	0.175	0.782	p-value	**0.028**

= 0.19, p = .049). Additionally, the Perspective Taking subscale of IRI also showed a near-significant correlation with HRS (r = 0.19, p = .051). There was no significant relation observed between IA and empathy.

Chapter 5. Discussion

In the present study, we aimed to determine how empathy processes in response to emotions expressed in music and social contexts. To achieve this, we employed the established Empathic Accuracy Task paradigm (Zaki et al., 2008) to measure empathy. The real-time behavioral data on empathic accuracy and physiological data associated with affective sharing were collected, comparing the responses based on the type of sensory information and emotional valence provided to the observer.

To determine whether a common empathic mechanism underlies music and social contexts, the multilevel correlations of empathic accuracy were analyzed between each situation. The results revealed an overall trend: individuals proficient in inferring emotions in social contexts tended to excel in inferring emotions in music as well. This outcome aligns with prior research (Tabak et al., 2023), suggesting that empathy in music and social situations may share similar neural processing mechanisms (Wallmark et al., 2018).

In particular, there was a noteworthy association between empathic accuracy for negative emotions in social situations and music performance, despite no significant difference observed between empathic accuracy for positive and negative emotional stimuli. This pattern diverges from the findings of a large-scale study in the United States (Tabak et al., 2023), which indicated stronger music-social empathic accuracy for positive emotions. It's plausible that the association with negative stimuli in this study might be due to Koreans' greater familiarity with the expression of such emotions in music. An analysis of 30 years of Korean popular music data by Jo and Kim, 2023 reported a historical prevalence of music with negative emotions over positive emotions, a trend that has gradually diminished over time. Given the recent shift towards more positive emotional content in Korean music, our study's observation that positive emotions are also likely to be linked to empathic accuracy in both music and society might reflect this evolving cultural context.

Moreover, the age of our participants could have influenced the association between music and social empathy for negative emotions. Existing literature consistently reports that older adults tend to exhibit attentional bias towards processing positive emotions over negative emotions (Fung et al., 2019; Ko et al., 2011). Given that our participants were composed of young adults, with an average age of 26, it's plausible that they were more attuned to the stimuli featuring negative emotions.

It is noteworthy that the capacity to accurately infer and empathize with others' emotions in both musical performances and storytelling videos was more pronounced when exclusively provided with visual information. This implies that individuals adept at utilizing visual cues to interpret emotions are more likely to exhibit higher empathic accuracy for visual representations of emotions in diverse contexts, particularly for negative emotions. This aligns with previous findings highlighting the proficiency of nonverbal information processing in processing general emotional cues (Jacob et al., 2012). This correlation also resonates with the results of Zaki et al., 2008, suggesting that nonverbal cues in negative emotional stimuli are linked to heightened empathic accuracy.

However, in contrast to the study conducted by Jospe et al., 2020, which demonstrated significant physiological synchrony even in the absence of auditory information, our present study did not unveil a disparity in heart rate synchrony between modalities. Furthermore, we did not identify a correlation between empathic accuracy and heart rate synchrony values. Consequently, it is not possible to conclusively assert that individuals who processed visual information of negative emotions also processed the

information with physiological synchrony.

Nevertheless, the results suggest indirect evidence that music and social situations may share certain mechanisms related to visual-emotional inference. In a functional magnetic resonance imaging (fMRI) study, Jacob et al., 2012 discovered that individuals more influenced by nonverbal emotional cues exhibit increased activation in the medial orbitofrontal cortex when processing communicative signals. They argued that those displaying nonverbal dominance are not only sensitive to nonverbal cues but also proficient at processing emotional stimuli in general. In alignment with these earlier findings, it is plausible that individuals in our study who demonstrated greater accuracy in processing emotional information from visual-only stimuli in social situations were also more adept at inferring the emotions of others in a different emotional context, such as music performance. The observation that this ability was particularly pertinent to visual information of negative emotion aligns with the perspective of researchers who interpret the ability to swiftly identify threats in negative emotional stimuli as having evolved due to its evolutionary linkage to survival (Killgore and Yurgelun-Todd, 2004; Morris et al., 1998; Nomura et al., 2004).

The discovery of an association between empathic accuracy for visual information in both music and social-emotional contexts does not necessarily imply higher empathic accuracy solely due to visual information. Upon analyzing the differences among the three variables—the type of emotional context (music, social), valence (positive, negative), and modality (visual-only, audio-only, video-and-audio)—we observed that visual-only (VO) empathic accuracy in both music and social contexts was significantly lower than in all other type-modality contexts. There was no disparity in VO empathic accuracy between music and social contexts, and also no significant difference was observed between the results for audio-only (AO) and video-and-audio (VA) modalities, both of which demonstrated high empathic accuracy.

This supports the argument that auditory information plays a pivotal role in enhancing empathic accuracy in social situations and serves as the primary sense for emotional communication (Jospe et al., 2020; Kraus, 2017; Zaki et al., 2008). On the other hand, considering that the perception of facial expressions in emotion is intricately tied to contextual information (Righart and De Gelder, 2008; Righart and Gelder, 2008; Masuda et al., 2008; T.-H. Lee et al., 2012; H. Kim et al., 2004), we can speculate that the VO condition, involving inferring the emotions of others based solely on visual information without any contextual information about the content, may pose a challenging task, hence the relatively low accuracy.

In alignment with the earlier finding that cognitive empathy is more pronounced in the presence of auditory information, our study revealed that observers' responses in the presence of both audio and visual (VA) were notably similar to their responses in the AO condition. Specifically, when watching storytelling videos, observers' responses in the AO condition bore a closer resemblance to their responses in the VA condition than to their inferences in the VO condition. This trend was also evident when comparing Heart Rate Synchrony (HRS) across modalities for storytelling videos. The heart rate of observers who watched the VA video exhibited significantly higher synchrony with the heart rate of those receiving only auditory information than with those receiving only visual information.

Given that heart rate did not exhibit similarity across modalities for music performance, we approach the results cautiously, interpreting heart rate activation in storytelling as reflective of the individual's experiential narrative flow, rather than a response to the acoustical features of auditory information. This implies that contextual information plays a pivotal role in recognizing and communicating the emotions of others in social situations.

A notable observation is that the substantial similarity in responses between the VA and AO condi-

tions was not replicated in the context of music performance. Despite findings indicating lower empathic accuracy in the VO condition compared to the VA and AO conditions, the degree of resemblance between observers' responses to the soundless performance video (VO) and their responses to the video with sound added (VA) did not significantly differ from the similarity observed in VA-AO responses. This implies that vision may be as important as hearing in real-time information processing during the observation of a musical performance. This aligns with longstanding reports in the field of music cognition, underscoring the substantial impact of visual information on the audience's emotional experience during a performance (E. Coutinho and Scherer, 2017; Tsay, 2013; Vuoskoski et al., 2014; Vines et al., 2006).

Empathic accuracy scores were greater for storytelling videos compared to music performance videos. This mirrors the findings of Tabak et al., 2023, which similarly compared music performance to storytelling. In language, words are linked to specific meanings, facilitating a clearer understanding of the speaker's intended message within a given context. In contrast, in music, while each note and chord has a symbol within the musical structure, the interpretation of its emotional meaning is more self-referential (Meyer, 2008; Besson and Schön, 2001). Consequently, accurately inferring the emotion that a performer intends to express becomes more challenging in the context of music.

Furthermore, we observed no correlation between individual traits (including interoception awareness) and cognitive empathy processes, as assessed through real-time behavioral measures. Conversely, we identified static correlations between the Interpersonal Reactivity Index and heart rate synchrony in response to naturalistic social situations (Video-and-Audio, VA). In particular, the Perspective-Taking subscale of the IRI demonstrated a potential connection with physiological responses. The EA task paradigm is crafted to assess the capacity to comprehend the unspoken emotions of others (Ickes, 1993; Ickes, 2001; Ickes, 2016), with perspective-taking representing the ability to immerse oneself in the target's perspective and contemplate how they might have felt. Consequently, higher levels of perspective-taking may be linked to more effective social interaction with the target, as reflected in interpersonal synchronization (J. Coutinho et al., 2021). Given the absence of a correlation between behavioral cognitive empathy states and individual traits, individual traits may be more closely associated with emotional empathy—a bottom-up process manifesting as physiological responses.

Chapter 6. Conclusion

In this study, we investigated the interplay of empathic processing in emotional situations within the domains of music and society. Our exploration aimed to uncover both commonalities and distinctions in empathic processes across diverse contexts. Overall, individuals proficient in emotional reasoning exhibited a heightened ability to discern emotions across different situations. In both musical and social scenarios, auditory cues—particularly those conveying negative emotions—proved instrumental in perceiving and understanding others' emotions. However, when observing a musical performance, visual stimuli also emerged as a substantial factor influencing the audience's emotional experience. The capacity to interpret emotional visual information was identified as a shared ability in both musical and social contexts, with a specific focus on negative emotions.

In addition, in social situations where auditory information was relied upon to grasp others' emotions, individuals with high empathy levels demonstrated increased interpersonal cardiac synchrony during interactions. The significance of this study lies in affirming that common mechanisms for processing empathy, as observed in various social events like music performance and storytelling, may based on visual information. Additionally, the findings suggest that biased attention toward different emotional valences may depend on cultural backgrounds.

Expanding upon the identified modality-specific dominance in the empathic process, there is an opportunity to apply this knowledge for enhancing or modulating individuals' emotional experiences in various emotional situations. For example, within the realm of music performance, where visual information significantly shapes emotional experiences, placing emphasis on the musician's movements during the performance has the capacity to intensify the audience's emotional engagement and immersion. This can be achieved by maximizing the conveyance of the musician's emotions. Especially in the context of music performance platforms utilizing online systems such as augmented reality and virtual reality, incorporating additional visual cues alongside the music has the potential to exert a substantial influence on the audience's emotional experience.

Moreover, we can apply our findings to enhance social competence through music education, capitalizing on the shared empathy mechanisms between music performance and social situations. Engaging in collaborative musical activities provides a potential avenue for the transfer effects of music education, fostering improvements in empathy and interpersonal interaction skills.

This study has several limitations that should be acknowledged. Firstly, the use of piano playing footage in the music videos may have directed observers' attention more towards hand or arm movements, while storytelling videos could have led observers to focus primarily on facial expressions conveying emotions. The mechanisms for recognizing emotions may vary depending on the body part or facial feature conveying emotional information, potentially influencing the results. Future research should consider controlling for the specific body parts involved by comparing performances with vocalists or singers and by directly comparing music and social situations.

Secondly, this study incorporated three types of emotions (joy/happiness, sadness, and anger), and the arousal level of these emotions was not standardized. Unifying the number of emotions based on arousal levels could provide insights into the differences between positive and negative valence.

Thirdly, the study did not account for the potential impact of the observer's personal experiences on empathic accuracy during storytelling. Observers might have displayed greater empathy or biased

responses towards specific values if they perceived a similarity between the autobiographical story and their own experiences. A more nuanced analysis could have been achieved by assessing the level of familiarity with the storytelling content.

Additionally, it is possible that the storytelling videos, designed for individuals unfamiliar with video recording, may have posed challenges for participants in distinguishing between their emotional reactions induced by the story itself and those influenced by the negative emotions introduced during recording. Social situations with lay targets may also not have been as rich in nonverbal emotional expression as music performances with professional musicians. As this difference may have affected empathic responses, future studies should consider the professionalism of the emotional performers. As these differences may have influenced empathy responses, future studies should consider the expertise of the person expressing the emotion.

Furthermore, forthcoming research should take into account the inherent limitations of the Empathic Accuracy Task paradigm. Given that the emotional data gathered from the targets in this study is retrospective and not real-time emotion (i.e., emotions assessed while reviewing their own videos), there is a potential for variations in responses between the authentic emotions experienced by the subjects during the video recording and the self-reported data obtained while viewing the video retrospectively.

Lastly, the study utilized heart rate as an index of physiological response but did not yield significant results. Future investigations should explore and analyze alternative physiological indicators capable of capturing responses in empathic states.

Bibliography

Ainley, V., Maister, L., & Tsakiris, M. (2015). Heartfelt empathy? no association between interoceptive awareness, questionnaire measures of empathy, reading the mind in the eyes task or the director task. *Frontiers in Psychology*, *6*, 554.

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., Busch, V., Lothwesen, K., Elvers, P., Fischinger, T., et al. (2018). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*.

Ara, A., & Marco-Pallarés, J. (2020). Fronto-temporal theta phase-synchronization underlies music-evoked pleasantness. *NeuroImage*, *212*, 116665.

Baiano, C., Job, X., Santangelo, G., Auvray, M., & Kirsch, L. P. (2021). Interactions between interoception and perspective-taking: Current state of research and future directions. *Neuroscience & Biobehavioral Reviews*, *130*, 252–262.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, *34*, 163–175.

Bernardi, N. F., Codrons, E., di Leo, R., Vandoni, M., Cavallaro, F., Vita, G., & Bernardi, L. (2017). Increase in synchronization of autonomic rhythms between individuals when listening to music. *Frontiers in physiology*, *8*, 785.

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, *930*(1), 232–258.

Blanke, E. S., Rauers, A., & Riediger, M. (2016). Does being empathic pay off?—associations between performance-based measures of empathy and social adjustment in younger and older women. *Emotion*, *16*(5), 671.

Bruder, M., Dosmukhambetova, D., Nerb, J., & Manstead, A. S. (2012). Emotional signals in nonverbal interaction: Dyadic facilitation and convergence in expressions, appraisals, and feelings. *Cognition & emotion*, *26*(3), 480–502.

Coutinho, E., & Scherer, K. R. (2017). The effect of context and audio-visual modality on emotions elicited by a musical performance. *Psychology of Music*, *45*(4), 550–569.

Coutinho, J., Pereira, A., Oliveira-Silva, P., Meier, D., Lourenço, V., & Tschacher, W. (2021). When our hearts beat together: Cardiac synchrony as an entry point to understand dyadic co-regulation in couples. *Psychophysiology*, *58*(3), e13739.

Craig, A. (2010). The sentient self. *Brain structure and function*, *214*, 563–577.

Cross, I., & Morley, I. (2009). Music in evolution: Theories, definitions and the nature of the evidence.

Czepiel, A., Fink, L. K., Fink, L. T., Wald-Fuhrmann, M., Tröndle, M., & Merrill, J. (2021). Synchrony in the periphery: Inter-subject correlation of physiological responses during live music concerts. *Scientific reports*, *11*(1), 22457.

Dale, A., & Anderson, D. (1978). Information variables in voluntary control and classical conditioning of heart rate: Field dependence and heart-rate perception. *Perceptual and Motor Skills*, *47*(1), 79–85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in cognitive sciences*, *10*(10), 435–441.

Decety, J., Lamm, C., et al. (2006). Human empathy through the lens of social neuroscience. *The scientific World journal*, *6*, 1146–1163.

Dor-Ziderman, Y., Cohen, D., Levit-Binnun, N., & Golland, Y. (2021). Synchrony with distress in affective empathy and compassion. *Psychophysiology*, *58*(10), e13889.

Feyereisen, P., Malet, C., & Martin, Y. (1986). Is the faster processing of expressions of happiness modality-specific? *Aspects of face processing*, 349–355.

Finset, A., & Ørnes, K. (2017). Empathy in the clinician–patient relationship: The role of reciprocal adjustments and processes of synchrony. *Journal of patient experience*, *4*(2), 64–68.

Fung, H. H., Gong, X., Ngo, N., & Isaacowitz, D. M. (2019). Cultural differences in the age-related positivity effect: Distinguishing between preference and effectiveness. *Emotion*, *19*(8), 1414.

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae scientiae*, *5*(1_suppl), 123–147.

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological psychology*, *104*, 65–74.

Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology*, *77*(4), 746.

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M., & Tan, L.-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in psychology*, *7*, 976.

Golland, Y., Arzouan, Y., & Levit-Binnun, N. (2015). The mere co-presence: Synchronization of autonomic signals and emotional responses across co-present individuals not engaged in direct interaction. *PloS one*, *10*(5), e0125804.

Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, *7*(2), 438.

Hayes, D. J., Duncan, N. W., Xu, J., & Northoff, G. (2014). A comparison of neural responses to appetitive and aversive stimuli in humans and other mammals. *Neuroscience & Biobehavioral Reviews*, *45*, 350–368.

Herbert, B. M., & Pollatos, O. (2012). The body in the mind: On the relationship between interoception and embodiment. *Topics in cognitive science*, *4*(4), 692–704.

Hou, Y., Song, B., Hu, Y., Pan, Y., & Hu, Y. (2020). The averaged inter-brain coherence between the audience and a violinist predicts the popularity of violin performance. *Neuroimage*, *211*, 116655.

Ickes, W. (1993). Empathic accuracy. *Journal of personality*, *61*(4), 587–610.

Ickes, W. (2001). Measuring empathic accuracy. *Interpersonal sensitivity: Theory and measurement*, *1*, 219–241.

Ickes, W. (2009). Empathic accuracy: Its links to clinical, cognitive, developmental, social, and physiological psychology. *The social neuroscience of empathy*, 57–70.

Ickes, W. (2016). Empathic accuracy: Judging thoughts and feelings.

Imafuku, M., Fukushima, H., Nakamura, Y., Myowa, M., & Koike, S. (2020). Interoception is associated with the impact of eye contact on spontaneous facial mimicry. *Scientific Reports*, *10*(1), 19866.

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., & Atkins, D. C. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *Journal of counseling psychology*, *61*(1), 146.

Jacob, H., Kreifelts, B., Brück, C., Erb, M., Hösl, F., & Wildgruber, D. (2012). Cerebral integration of verbal and nonverbal emotional cues: Impact of individual nonverbal dominance. *NeuroImage*, *61*(3), 738–747.

Jo, W., & Kim, M. J. (2023). Tracking emotions from song lyrics: Analyzing 30 years of k-pop hits. *Emotion*, *23*(6), 1658.

Jospe, K., Genzer, S., klein Selle, N., Ong, D., Zaki, J., & Perry, A. (2020). The contribution of linguistic and visual cues to physiological synchrony and empathic accuracy. *Cortex*, *132*, 296–308.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of life reviews*, *10*(3), 235–266.

Kaplan, J. T., & Iacoboni, M. (2006). Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy. *Social neuroscience*, *1*(3-4), 175–183.

Kauschke, C., Bahn, D., Vesker, M., & Schwarzer, G. (2019). The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias? *Frontiers in psychology*, *10*, 1654.

Kawakami, A., & Katahira, K. (2015). Influence of trait empathy on the emotion evoked by sad music and on the preference for it. *Frontiers in psychology*, *6*, 1541.

Killgore, W. D., & Yurgelun-Todd, D. A. (2004). Activation of the amygdala and anterior cingulate during nonconscious processing of sad versus happy faces. *Neuroimage*, *21*(4), 1215–1223.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of cognitive neuroscience*, *16*(10), 1730–1745.

Kim, J., & Lee, S. J. (2010). Reliability and validity of the korean version of the empathy quotient scale. *Psychiatry investigation*, *7*(1), 24.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and human behavior*, *31*(5), 354–364.

Ko, S.-G., Lee, T.-H., Yoon, H.-Y., Kwon, J.-H., & Mather, M. (2011). How does context affect assessments of facial emotion? the role of culture and age. *Psychology and aging*, *26*(1), 48.

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist*, *72*(7), 644.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the empathy quotient. *Psychological medicine*, *34*(5), 911–920.

Lee, J., Zaki, J., Harvey, P.-O., Ochsner, K., & Green, M. F. (2011). Schizophrenia patients are impaired in empathic accuracy. *Psychological Medicine*, *41*(11), 2297–2304. https://doi.org/10.1017/S0033291711000614

Lee, T.-H., Choi, J.-S., & Cho, Y. S. (2012). Context modulation of facial emotion perception differed by individual difference. *PLOS one*, *7*(3), e32987.

Legrand, N., Nikolova, N., Correa, C., Brændholt, M., Stuckert, A., Kildahl, N., Vejlø, M., Fardo, F., & Allen, M. (2022). The heart rate discrimination task: A psychophysical method to estimate the accuracy and precision of interoceptive beliefs. *Biological Psychology*, *168*, 108239.

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The brain basis of positive and negative affect: Evidence from a meta-analysis of the human neuroimaging literature. *Cerebral cortex*, *26*(5), 1910–1922.

Livingstone, S. R., Thompson, W. F., Wanderley, M. M., & Palmer, C. (2015). Common cues to emotion in the dynamic facial expressions of speech and song. *Quarterly Journal of Experimental Psychology*, *68*(5), 952–970.

MacGregor, C., Ruth, N., & Müllensiefen, D. (2023). Development and validation of the first adaptive test of emotion perception in music. *Cognition and Emotion*, *37*(2), 284–302.

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020). Methods and algorithms for correlation analysis in r. *Journal of Open Source Software*, *5*(51), 2306.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of personality and social psychology*, *94*(3), 365.

McKenzie, K., Russell, A., Golm, D., & Fairchild, G. (2022). Empathic accuracy and cognitive and affective empathy in young adults with and without autism spectrum disorder. *Journal of autism and developmental disorders*, *52*(5), 2004–2018.

Meyer, L. B. (2008). *Emotion and meaning in music*. University of chicago Press.

Miskovic, V., & Anderson, A. (2018). Modality general and modality specific coding of hedonic valence. *Current Opinion in Behavioral Sciences*, *19*, 91–97.

Miu, A. C., & Balteş, F. R. (2012). Empathy manipulation impacts music-induced emotions: A psychophysiological study on opera. *PloS one*, *7*(1), e30618.

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: From motion to'e'motion. *Social cognitive and affective neuroscience*, *1*(3), 235–241.

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*(6684), 467–470.

Mul, C.-l., Stagg, S. D., Herbelin, B., & Aspell, J. E. (2018). The feeling of me feeling for you: Interoception, alexithymia and empathy in autism. *Journal of Autism and Developmental Disorders*, *48*, 2953–2967.

Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: Enhanced sensitivity to negative word valence. *Emotion*, *9*(5), 609.

Nomura, M., Ohira, H., Haneda, K., Iidaka, T., Sadato, N., Okada, T., & Yonekura, Y. (2004). Functional association of the amygdala and ventral prefrontal cortex during cognitive evaluation of facial expressions primed by masked angry faces: An event-related fmri study. *Neuroimage*, *21*(1), 352–363.

Nummenmaa, L., & Calvo, M. G. (2015). Dissociation between recognition and detection advantage for facial expressions: A meta-analysis. *Emotion*, *15*(2), 243.

Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? an fmri study on neural systems for affective and cognitive empathy. *Neuroimage*, *43*(3), 571–580.

Peng, W., Lou, W., Huang, X., Ye, Q., Tong, R. K.-Y., & Cui, F. (2021). Suffer together, bond together: Brain-to-brain synchronization and mutual affective empathy when sharing painful experiences. *Neuroimage*, *238*, 118249.

Righart, R., & De Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: An erp study. *Social cognitive and affective neuroscience*, *3*(3), 270–278.

Righart, R., & Gelder, B. d. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 264–272.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of consulting psychology*, *21*(2), 95.

Rum, Y., & Perry, A. (2020). Empathic accuracy in clinical populations. *Frontiers in Psychiatry*, *11*, 457.

Satpute, A. B., Kang, J., Bickart, K. C., Yardley, H., Wager, T. D., & Barrett, L. F. (2015). Involvement of sensory regions in affective experience: A meta-analysis. *Frontiers in psychology*, *6*, 1860.

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59. https://doi.org/10.1017/S0140525X20000333

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology*, *18*(4), 483–488.

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in psychology*, *4*, 837.

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, *17*(1), 18–24.

Shinkareva, S. V., Wang, J., Kim, J., Facciani, M. J., Baucom, L. B., & Wedell, D. H. (2014). Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. *Human brain mapping*, *35*(7), 3558–3568.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96.

Stupacher, J., Mikkelsen, J., & Vuust, P. (2022). Higher empathy is associated with stronger social bonding when moving together with music. *Psychology of music*, *50*(5), 1511–1526.

Stürmer, S., Snyder, M., Kropp, A., & Siem, B. (2006). Empathy-motivated helping: The moderating role of group membership. *Personality and Social Psychology Bulletin*, *32*(7), 943–956.

Tabak, B. A., Wallmark, Z., Nghiem, L. H., Alvi, T., Sunahara, C. S., Lee, J., & Cao, J. (2023). Initial evidence for a relation between behaviorally assessed empathic accuracy and affect sharing for people and music. *Emotion*, *23*(2), 437.

Tajadura-Jiménez, A., & Tsakiris, M. (2014). Balancing the "inner" and the "outer" self: Interoceptive sensitivity modulates self–other boundaries. *Journal of Experimental Psychology: General*, *143*(2), 736.

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences*, *110*(36), 14580–14585.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*(2), 157–170.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80–113.

Vuoskoski, J. K., & Eerola, T. (2012). Empathy contributes to the intensity of music-induced emotions. *Proceedings of the 12th international conference on music perception and cognition (ICMPC)*, 1112–1113.

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics*, *76*, 591–604.

Wallmark, Z., Deblieck, C., & Iacoboni, M. (2018). Neurophysiological effects of trait empathy in music listening. *Frontiers in behavioral neuroscience*, 66.

Wöllner, C. (2012). Is empathy related to the perception of emotional expression in music? a multimodal time-series analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(3), 214.

Wu, X., & Lu, X. (2021). Musical training in the development of empathy and prosocial behaviors. *Frontiers in Psychology*, *12*, 661769.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological science*, *19*(4), 399–404.

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, *9*(4), 478.

Supplementary Material

Table 1: Results of correlation between observers' empathic accuracy (rZ) and heart rate synchrony (rZ).

Type	Valence		Modality			
			All	Video-only	Audio-only	Video-and-Audio
All	All	r	0.01	0.08	-0.07	0.03
		p	0.885	0.268	0.290	0.720
	Positive	r	-0.05	0.07	-0.18	0.03
		p	0.444	0.553	0.138	0.832
	Negative	r	0.04	0.12	-0.02	0.02
		p	0.268	0.153	0.823	0.824
Music	All	r	0.05	-0.01	0.07	0.07
		p	0.369	0.945	0.482	0.465
	Positive	r	-0.05	-0.17	0.03	-0.05
		p	0.596	0.320	0.865	0.778
	Negative	r	0.11	0.09	0.10	0.11
		p	0.121	0.464	0.415	0.347
Social	All	r	0.02	0.14	-0.15	0.11
		p	0.793	0.137	0.111	0.272
	Positive	r	-0.02	0.24	-0.30	0.23
		p	0.839	0.168	0.080	0.186
	Negative	r	0.05	0.17	-0.09	0.03
		p	0.443	0.152	0.461	0.799

Table 2: Demographic information for participants. The table includes mean and standard deviation values. The results of the Student's t-test between males and females showed that the EQ total, EQ-CE, and EQ-SS scores of males were significantly higher than those of females. EQ = Empathy Quotient; CE = Cognitive Empathy; ER = Emotional Reactivity; SS = Social Skills; IRI = Interpersonal Reactivity Index; PT = Perspective Taking; EC = Empathic Concern; IA = Interoceptive Accuracy. $^{**}p < .01$, $^{***}p < .001$.

		Mean (*SD*)		
		All (n = 36)	Male (n = 19)	Female (n = 17)
Age		26.06 (3.56)	25.89 (3.91)	26.24 (3.23)
EQ				
	Total* **	**40.64 (12.09)**	**46.79 (10.28)**	**33.76 (10.28)**
	CE* **	**11.53 (4.75)**	**14.11 (4.29)**	**8.65 (3.46)**
	ER	10.39 (4.53)	11.68 (4.27)	8.94 (4.48)
	SS **	**6.25 (2.30)**	**7.21 (2.15)**	**5.18 (2.01)**
IRI				
	Total	51.86 (8.20)	53.21 (8.78)	50.35 (7.46)
	PT	26.61 (4.42)	26.89 (4.65)	26.29 (4.27)
	EC	25.25 (4.63)	26.32 (4.84)	24.06 (4.21)
IA				
	Total	0.64 (0.18)	0.63 (0.20)	0.64 (0.17)
Music				
	Formal education (years)	6.26 (4.70)	5.29 (4.28)	7.34 (5.05)
	Piano education (years)	3.72 (4.13)	3 (3.14)	4.53 (4.99)
	Instruments practice (hr/week)	0.44 (1.18)	0.47 (1.31)	0.41 (1.06)