Cognition 239 (2023) 105537

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Aesthetic and physiological effects of naturalistic multimodal music listening

Anna Czepiel[a,b,*], Lauren K. Fink[a,c], Christoph Seibert[d], Mathias Scharinger[e,f], Sonja A. Kotz[b,g]

[a] Department of Music, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[b] Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands
[c] Max Planck-NYU Center for Language, Music, and Emotion, Frankfurt am Main, Germany
[d] Institute for Music Informatics and Musicology, University of Music Karlsruhe, Karlsruhe, Germany
[e] Research Group Phonetics, Department of German Linguistics, University of Marburg, Marburg, Germany
[f] Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[g] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Keywords:
Audiovisual
Physiology
Naturalistic
Neuroaesthetics
Motor mimicry

ABSTRACT

Compared to audio only (AO) conditions, audiovisual (AV) information can enhance the aesthetic experience of a music performance. However, such beneficial multimodal effects have yet to be studied in naturalistic music performance settings. Further, peripheral physiological correlates of aesthetic experiences are not well-understood. Here, participants were invited to a concert hall for piano performances of Bach, Messiaen, and Beethoven, which were presented in two conditions: AV and AO. They rated their aesthetic experience (AE) after each piece (Experiment 1 and 2), while peripheral signals (cardiorespiratory measures, skin conductance, and facial muscle activity) were continuously measured (Experiment 2). Factor scores of AE were significantly higher in the AV condition in both experiments. LF/HF ratio, a heart rhythm that represents activation of the sympathetic nervous system, was higher in the AO condition, suggesting increased arousal, likely caused by less predictable sound onsets in the AO condition. We present partial evidence that breathing was faster and facial muscle activity was higher in the AV condition, suggesting that observing a performer's movements likely enhances motor mimicry in these more voluntary peripheral measures. Further, zygomaticus ('smiling') muscle activity was a significant predictor of AE. Thus, we suggest physiological measures are related to AE, but at different levels: the more involuntary measures (i.e., heart rhythms) may reflect more sensory aspects, while the more voluntary measures (i.e., muscular control of breathing and facial responses) may reflect the liking aspect of an AE. In summary, we replicate and extend previous findings that AV information enhances AE in a naturalistic music performance setting. We further show that a combination of self-report and peripheral measures benefit a meaningful assessment of AE in naturalistic music performance settings.

1. Introduction

There is a clear consensus that listening to music induces aesthetic experiences, with humans augmenting such experiences by optimising the 'where' and 'how' we listen to music, such as in concerts (Sloboda, Lamont, & Greasley, 2012; Sloboda & O'Neill, 2001; Wald-Fuhrmann et al., 2021). Although the aesthetic experience (AE) of music is enhanced in a concert by several aspects (see Wald-Fuhrmann et al., 2021 for an overview), one explored here is visual information. While previous work showed that visual cues enhance self-reported musical evaluation of music performances (e.g., see Platz & Kopiez, 2012 for a meta-analysis), some gaps in the literature remain. Firstly, most studies comparing audiovisual (AV) and audio only (AO) musical performances have been conducted in laboratory settings; to test a more genuine AE, it is imperative to use a more naturalistic situation. Secondly, only two studies so far explored physiological responses between AV and AO musical performances (Chapados & Levitin, 2008; Vuoskoski, Gatti, Spence, & Clarke, 2016), but their findings are contrary to each other. Thus, the current study aimed to specify the link between modality (AO vs. AV), AE, and peripheral physiological responses in a naturalistic music performance setting, i.e., a piano concert.

While the initial study of AE has had a strong philosophical focus, AE

---

* Corresponding author at: Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany.
E-mail address: anna.czepiel@ae.mpg.de (A. Czepiel).

https://doi.org/10.1016/j.cognition.2023.105537
Received 4 July 2022; Received in revised form 31 May 2023; Accepted 24 June 2023

is currently of great interest in cognitive neuroscience and the neuroscientific subdiscipline of neuroaesthetics. Here, perception, emotion, and appreciation are considered to influence AE (for comprehensive reviews, see Anglada-Tort & Skov, 2022; Brattico & Pearce, 2013; Juslin, 2013; Pelowski, Markey, Lauring, & Leder, 2016; Schindler et al., 2017). Specific to the dynamic nature of music, Brattico, Bogert, and Jacobsen (2013) proposed that the AE of music listening is composed of a chronometry of components: 1) perceptual sensory processes (feature analysis/integration) as well as early emotional reactions (e.g., startle reflex and arousal), 2) cognitive processes (based on long-term knowledge, such as harmonic expectancy), and 3) affective processing (including perceived and felt emotions). A combination of these processes that involve somatomotor processes interacting with the listener themselves (in terms of cultural knowledge, musical expertise, etc.) and external context (e.g., social setting), result in 4) aesthetic responses (emotions, judgements, and liking). Brattico et al. (2013) presented neurophysiological correlates that might reflect these processes. Namely, sensory processes should be reflected in early event-related potentials (ERPs) and in early auditory processing areas (sensory cortices, brainstem). More cognitive ('error' and 'surprise') components should be reflected in the MMN and P300 and non-primary sensory cortices. Finally, (aesthetic) emotion and judgements should be reflected in the late potential component (LPC) and reward and emotion areas in the brain. Research further suggests that (synchronisation of) certain brain oscillations are related to music-evoked pleasure, particularly frontal theta oscillations (Ara & Marco-Pallarés, 2020; Sammler, Grigutsch, Fritz, & Koelsch, 2007; Tervaniemi, Pousi, Seppälä, & Makkonen, 2021), parieto-occipital alpha (Nemati, Akrami, Salehi, Esteky, & Moghimi, 2019), theta (Chabin et al., 2020) and theta phase synchronisation (Ara & Marco-Pallarés, 2020, 2021), as well as the inter-brain synchrony (IBS) of frontal and temporal theta in shared musical pleasure (Chabin et al., 2022).

Although some work has explored music-evoked pleasure with neural responses (electroencephalography, EEG) in the more naturalistic setting of a concert hall (Chabin et al., 2022), measuring brain activity in such settings comes with significant challenges. A more accessible approach, however, has been to measure peripheral physiological responses in naturalistic settings such as theatres (Ardizzi, Calbi, Tavaglione, Umiltà, & Gallese, 2020), concert halls (Egermann, Pearce, Wiggins, & McAdams, 2013), and cathedrals (Bernardi et al., 2017). Peripheral measures include the somatic (voluntary muscle) and autonomic nervous systems (ANS), of which the latter comprises the sympathetic ('fight-or-flight') and parasympathetic ('rest-and-digest') nervous systems (SNS, PNS). In naturalistic settings, previous work revealed (synchronised) physiological arousal responses in audiences occur in relation to surprising, emotional, and structural moments in music such as transitional passages, boundaries, and phrase repetitions (Czepiel et al., 2021; Egermann et al., 2013; Merrill, Czepiel, Fink, Toelle, & Wald-Fuhrmann, 2021). Such peripheral measures are likewise mentioned in the AE chronometry approach (Brattico et al., 2013) as reflecting tension and chill responses (Grewe, Kopiez, & Altenmüüller, 2009; Salimpoor, Benovoy, Longo, Cooperstock, & Zatorre, 2009). However, unlike brain regions (functional magnetic resonance imaging, fMRI) and the latency/polarity of (EEG/ magnetoencephalography, MEG) components, that can be attributed to psychological processes (Kappenman & Luck, 2011), peripheral responses are mainly characterised according to increased/decreased activity, making it more difficult to separate responses relating to distinct sensory, cognitive, and/or aesthetic processes. Thus, rather than taking a superficial understanding that such measures directly index a pleasurable experience, a more thorough biological understanding is required to appropriately interpret the meaning of such measures (see e.g., Fink et al., 2023, for an example in pupillometry).

The current dependent measures of interest, which have also previously been used in research on musical aesthetics (e.g., Grewe et al., 2009; Salimpoor et al., 2009), range from involuntary ANS responses to voluntary motoric control, namely: skin conductance, heart, respiratory, and muscle activity. Skin conductance (SC, also known as electrodermal activity, EDA) measures activation of sweat glands, which are innervated by the SNS only. The heart consists of cardiac muscle (involuntary control), with SNS (via sympathetic nerves) and PNS (vagus) innervations that increase and decrease heart rate (HR), respectively. Typically, HR fluctuates and is measured by different heart rate variability (HRV) measures. These measures can be in the time-domain, for example, the standard deviation between interbeat intervals, or in the frequency-domain, for example, power of certain frequency bands related to SNS and PNS activation. Power at a high frequency (HF, 0.4–0.15 Hz) component is attributed to PNS activity, while power at a low frequency (LF, 0.04–0.15 Hz) component seems to reflect both PNS and SNS influences; thus, the LF/HF ratio is used to represent SNS activity (Malik, 1996; Shaffer & Ginsberg, 2017). Respiratory activity encompasses both involuntary control - where the lungs are innervated by both SNS and PNS, which dilate and constrict the bronchioles, respectively - as well as voluntary control (Purves & Williams, 2001). The somatic (muscle) system consists mainly of skeletal (voluntary) muscle; commonly measured are the facial muscles of zygomaticus major ('smiling') and corrugator supercilii ('frowning'). Although under voluntary control, certain facial muscle responses may be partly unconscious (i.e., occur without attention or conscious awareness, Dimberg, Thunberg, & Elmehed, 2000). Overall, SC, heart, respiration, and facial muscle activity broadly relate to arousal and valence.[1] Higher arousal has been associated with SNS activation, such as increased sweat secretion, increased LF/HF ratio, HR and RR acceleration, and decreased HF power (Di Bernardi Luft & Bhattacharya, 2015; Shaffer & Ginsberg, 2017), while zygomatic and corrugator muscle activity seem to reflect positive and negative valence, respectively (Bradley & Lang, 2000; Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Dimberg et al., 2000; Lang, Greenwald, Bradley, & Hamm, 1993; Larsen, Norris, & Cacioppo, 2003, though see discussion below).

Although broadly reflecting arousal and valence, peripheral measures have been related to sensory, cognitive, and aesthetic experiences with regard to acoustic/musical stimuli in separate studies. Increased SC and HR patterns have been related to early sensory reactions to an acoustic signal - referred to as an orienting response/startle reflex (Barry, 1975; Barry & Sokolov, 1993; Graham & Clifton, 1966; Roy, Mailhot, Gosselin, Paquette, & Peretz, 2009). Physiological changes occur in response to cognitive music processes such as recognising unexpected harmonic chords (Koelsch, Kilches, Steinbeis, & Schelinski, 2008; Steinbeis, Koelsch, & Sloboda, 2006) and deviant stimuli (in an MMN-like paradigm, Chuen, Sears, & McAdams, 2016; though see Lyytinen, Blomberg, & Näätänen, 1992), which might be enhanced by attention (Frith & Allen, 1983). In more naturalistic music listening, many studies showed that arousing music (faster tempi and unpredictable harmony) increase SC, HR, and RR (Bernardi, Porta, & Sleight, 2006; Coutinho & Cangelosi, 2011; Czepiel et al., 2021; Dillman Carpentier & Potter, 2007; Egermann et al., 2013; Egermann, Fernando, Chuen, & McAdams, 2015; Khalfa, Isabelle, Jean-Pierre, & Manon, 2002; Krumhansl, 1997), though we note this result is not consistent across studies, for reviews see (Bartlett, 1996; Hodges, 2009; Koelsch & Jäncke, 2015). In terms of valence, researchers have shown that zygomaticus activity increases during happy music (Lundqvist, Carlsson, Hilmersson, & Juslin, 2008). However, other work showed it can

---

[1] The two main dimensions of emotion, according to the dimensional model of emotion (Russell, 1980). These terms reflect bipolar continuums: arousal ranging from calm to excitement, while valence varies from negative to positive emotional experience. Such peripheral responses have also been attributed to the discrete (basic) emotion theory, where SNS activation relates to happiness/fear, while PNS activation relates to calmness/sadness). For a more thorough discussion on emotion models, see for example (Barrett & Russell, 2015; Hamann, 2012).

increase during unpleasant (dissonant) music (Dellacherie, Roy, Hugueville, Peretz, & Samson, 2011; Merrill et al., 2021). This conflict suggests that perhaps the activation of the smiling muscle is not just related to valence (see also Wingenbach, Brosnan, Pfaltz, Peyk, & Ashwin, 2020). Peripheral responses have likewise been related to aesthetic experience of music, or least music-evoked "chills" (frissons), which increases SC, HR, RR and EMG (Blood & Zatorre, 2001; Craig, 2005; Grewe et al., 2009; Salimpoor et al., 2009). Hence, evidence suggests that peripheral measures can reflect (a mixture of) the sensory, cognitive and/or preference parts of the AE, rather than being a direct index of AE. Therefore, it is of importance to collect self-report measures to further interpret the peripheral responses to AV and AO comparisons.

In terms of modality effects on self-reports, audio information seems to be consistently influenced by performer movement. In one percussion study, pairing visual gestures that created long notes to acoustic sounds of short notes resulted in short sounds being perceived as longer sounding notes (Schutz & Lipscomb, 2007); an effect later shown to be consistent in percussive (but not sustained) sounds when the sound appears after a gesture (Schutz & Kubovy, 2009). In piano performances, one acoustic performance was paired with four videos: one as the original performance and three pianist 'doubles'. Ninety-two out of ninety-three participants perceived differences between the performances, although the sound remained identical (Behne & Wöllner, 2011). With regard to more aesthetic influences, several studies that compared uni- and bimodal versions of music performances showed visual cues enhance a listener's perception of performance quality (Waddell & Williamon, 2017), musical expertise (Griffiths & Reay, 2018; Tsay, 2013), musical expression (Broughton & Stevens, 2009; Davidson, 1993; Lange, Fünderich, & Grimm, 2022; Luck, Toiviainen, & Thompson, 2010; Morrison & Selvey, 2014; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011; Vuoskoski, Thompson, Clarke, & Spence, 2014), perception of emotional intention (Dahl & Friberg, 2007; Vines, Krumhansl, Wanderley, & Levitin, 2006), and felt emotion (Van Zijl & Luck, 2013). As AE is related to the appreciation of performance expressiveness, quality, and emotion (Brattico & Pearce, 2013; Juslin, 2013), this research, as well as a meta-analysis (Platz & Kopiez, 2012), showed that AE increases with additional visual cues. One neuroaesthetic theory that could further explain this enhanced AE postulates that visual information may increase embodied simulation, which subsequently increases AE (Freedberg & Gallese, 2007; Gallese & Freedberg, 2007). Support for this idea comes from studies showing higher activation in the action observation network when viewing movements that are rated as aesthetically pleasing (Cross, 2011).

However, this enhanced AE effect has been mostly assessed in laboratory settings. Recent studies are increasingly exploring such experiences in live concerts (Chabin et al., 2022; Coutinho & Scherer, 2017; Czepiel et al., 2021; Scherer, Trznadel, Fantini, & Coutinho, 2019; Swarbrick et al., 2019; Tervaniemi et al., 2021), where participants report experiencing stronger emotions (Gabrielsson & Wik, 2003; Lamont, 2011); however, Belfi, Samson, Crane, and Schmidt (2021) found that felt pleasure did not differ between live and an audiovisual recording of the same performance. Focusing more specifically on the role of modality, to date only a few studies compare responses to AV vs. AO conditions in naturalistic settings. Compared to eyes-closed conditions, eyes-open conditions increased movement energy and interpersonal coordination, suggesting that visual information may enhance the social aspect of live pop/soul music (Dotov & Trainor, 2021). Coutinho and Scherer (2017) compared emotional responses in a live AV performance to recorded AV, AO, and VO performances of Schubert Lieder, where the live AV condition had significantly higher wonder and significantly lower boredom ratings. Although these two studies highlight the difference between genres and the affordances that visual information can give (focus on seeing other audience members/musicians in popular/classical music, respectively), they essentially show that additional information enhances the (social/emotional) experience. We stress that it is not trivial to replicate findings from the lab to a more naturalistic setting, since, for example, well documented effects of familiarity and body movement on music appreciation found from lab studies were not replicated in a field study (Anglada-Tort, Thueringer, & Omigie, 2019). It is also worth extending Coutinho and Scherer (2017), since they focus on the more emotional part of AE, and only collected data from an AV modality in a naturalistic setting (other modalities were tested in a lab-like setting). The current study thus compares modalities in one naturalist setting to examine more specifically the judgement and preference components of AE.

Two previous studies have compared peripheral physiological responses as a function of modality during music performances and serve as the starting point for the current work. Chapados and Levitin (2008) found that self-reported tension as well as SC were both highest in AV conditions. However, Vuoskoski et al. (2016) found that, although self-reported intensity, high energy arousal, and tension were highest in AV conditions, SC was actually highest in AO conditions. While the discrepancy between these two studies could relate to the different styles and instruments used (which offer different expressive affordances), Vuoskoski et al. (2016) argued that SC might be higher during AO performances due to musical expectancy (Huron, 2006; Juslin & Västfjäll, 2008). More specifically, as visual information increases listeners' ability to predict upcoming musical events, AV stimuli are less surprising. Indeed, this idea is supported by speech studies focusing on the N100, an EEG event-related potential component that reflects early sensory processing, where a larger N100 amplitude can indicate a response to a less predictable stimulus. The N100 component is enhanced in AO (compared to AV) conditions in speech (Klucharev, Möttönen, & Sams, 2003; van Wassenhove, Grant, & Poeppel, 2005), emotional expression (Jessen & Kotz, 2011), as well as non-speech events such as clapping (Stekelenburg & Vroomen, 2007). These findings corroborate the idea that the lack of visual information makes sound onsets less predictable.

Together, this evidence suggests that peripheral responses might be 1) higher in AO conditions if they reflect sensory processing, or 2) higher in AV conditions if they reflect the enhanced emotional and/or appreciation aspects of AE. If peripheral physiological responses reflect sensory processing, we would expect to replicate results from Vuoskoski et al. (2016) and find increased physiological activity in AO conditions. However, if physiological responses reflect the more emotional/aesthetic aspects, we would expect to replicate results from Chapados and Levitin (2008) and find increased physiological responses in AV conditions.

In summary, more research is needed to assess modality effects that enhance aesthetic experience in a more naturalistic setting. Further, the peripheral physiological correlates of aesthetic effects are so far inconsistent. The current study consists of two experiments that examine AE and physiology between AV and AO conditions in a concert hall setting. In both Experiments, we recorded behavioural responses and tested the hypothesis that AE will be higher in the AV condition. In Experiment 2, we additionally collected physiological responses and tested the hypothesis put forward by Vuoskoski et al. (2016) that peripheral physiological activity should be higher in AO conditions.

2. General method

2.1. Overview

We present two experiments, each consisting of two concerts. Experiment 1 (Concerts 1 and 2) measured behavioural ratings, while Experiment 2 (Concerts 3 and 4) measured both behavioural ratings and physiological responses. Both involve the same stimuli and the same within-subjects experimental design: participants listening to piano performances of Bach, Beethoven, and Messiaen, in AO and AV conditions. Modality order was counterbalanced across concerts.

2.2. Stimuli

Upon engaging a pianist, three musical pieces were selected from their repertoire in accordance with the pianist and musical experts to represent various emotional expressions (cheerful, sad, and ambiguous) and musical styles (Baroque, Classical-Romantic, and 20th century music): Johann Sebastian Bach: Prelude and Fugue in D major (Book Two from the Well-Tempered Clavier, BWV 874), Ludwig Van Beethoven: Sonata No. 7, Op. 10, No. 3, second movement (Largo e mesto), and Olivier Messiaen: *Regard de l'Esprit de joie* (Number 10 from Vingt Regards sur L'Enfant-Jésus). These pieces were presented to the participants during each concert twice in the two different modalities: in audiovisual (AV) and an audio only (AO) versions. We considered this repetition of pieces as a naturalistic part of the design as piece repetition is a practice (although not extremely common) in concert programming (Halpern, Chan, Müllensiefen, & Sloboda, 2017).

Both AV and AO presentations of the music pieces were performed by the same pianist, playing on the same piano (Steinway B-211), in the same concert hall. AV versions of the music pieces were performed live during the concerts and the audience could see and hear the pianist performing the music. AO versions of the music pieces were recorded in the same concert hall, on the same piano in advance of the concerts, without an audience. The AO versions were presented during the concerts via a stereo setup with high-quality full-range loudspeakers (Fohhn LX-150 + Fohhn XS-22), so that the audience could only hear the music. During this time, the pianist was backstage, so that the audience could only see the piano. The playback AO versions were the same in all concerts in both experiments. To ensure similarity of sound levels between AO and AV presentations, a trained sound engineer checked that the loudness across the modalities was equal.

Although modality conditions were controlled as much as possible, we would assume that repeated performances of the same musical piece might have slight deviations from each other, even when performed by a highly trained professional musician (Chaffin, Lemieux, & Chen, 2007). Therefore, we checked that the stimuli nonetheless were comparable enough to eliminate confounding variables of potential acoustic differences between AV versions (different for each concert) and AO versions (the same across all concerts). We differentiated between score-based features and performance-based features (Goodchild, Wild, & McAdams, 2019). The former refers to features that come from the notated scores (e.g., harmonies), which should remain the same across performances (assuming no errors in playing the scores). The latter refers to features that may also be notated in the scores (e.g., dynamic markings) but might deviate more depending on the performances, such as tempo, loudness, and timbre. Tempo was extracted using a combination of MIDI information for each note and manually locating the beat (using Sonic Visualiser, Cannam, Landone, & Sandler, 2010), where inter-beat intervals were obtained to calculate continuous beats per minute (bpm). Loudness and timbre were extracted from the audio signal using MIRToolbox (Lartillot & Toiviainen, 2007) in MATLAB (2019b, The Mathworks Inc., USA), with RMS (*mirrms*) and spectral centroid (*mircentroid*) representing loudness and timbre, respectively. In checking multicollinearity (Lange & Frieler, 2018), none of the features correlated highly, confirming that each feature represented an independent aspect of the music. Each of the features were averaged into average bins per bar (American: measure) to account for slight timing deviances between performances. The features over time were very similar (see Supplementary Figs. 1–6 in Supplementary Materials). This similarity was confirmed by significant correlations between concerts, all with r values >6 (see Supplementary Materials, Supplementary Table 1), suggesting that all performances were acoustically comparable.

2.3. Questionnaires

Questionnaires were presented after each musical piece to assess three types of questions. Firstly, we assessed the 'naturalness' of the concert by asking to what extent the experimental components of the setting (e.g., measurement of the behavioural responses) disturbed the concert experience, where 'disturbed by measurement' was rated from 1 (strongly disagree) to 7 (very much agree). We further assessed familiarity with the style of music as well as whether the participant knew the specific piece of music. This was rated from 1 (not at all familiar) to 7 (very familiar). Thirdly, we assessed the main dependent variable of interest: aesthetic experience (AE). As an AE is made up of several components (Brattico & Pearce, 2013; Schindler et al., 2017), we assessed the aesthetic experience with a set of eight individual items, consisting of how much they liked the piece, how much they liked the interpretation of the piece, and how absorbed they felt in the music (see Supplementary Materials for all questions).

2.4. Procedure

Participants were invited to attend piano concerts that took place at the ArtLab of the Max-Planck-Institute for Empirical Aesthetics in Frankfurt, a custom-built concert hall seating 46 audience members (https://www.aesthetics.mpg.de/en/artlab/information.html). Concerts were kept as identical as possible for factors such as lighting, temperature, and timing. Prior to the concert, participants were informed about the experiment and filled in consent forms before being seated in the ArtLab. During the concert and after each piece of music, participants answered the short questionnaire described above. All participants saw the three pieces in both conditions. For one concert per Experiment, the three music pieces were presented first in the AO modality, and then repeated in the AV modality. Modality order was counterbalanced so that in the other concert per Experiment, music pieces were first presented in the AV modality, and then again in the AO modality. An overview of the procedure and modality condition orders can be found in Fig. 1. Behavioural measures were recorded in both Experiment 1 and 2. In Experiment 2 only, physiological data were additionally collected (details in Section 4.1.2 Experiment 2 Procedure).

2.5. Analysis

Statistical analyses were conducted in R and R studio (R Core Team, 2021; RStudio Team, 2021).

Items chosen for the questionnaire (see Supplementary Materials) reflect elements of an aesthetic experience. Thus, it was assumed that the items might be related to each other. Indeed, in both Experiments, items in the self-reports capturing the aesthetic experience were highly correlated. Therefore, rather than comparing modality differences for each item, we reduced the questionnaire items to an overall, more interpretable factor - that retains important information from each item - using a factor analysis (Fabrigar, Wegener, MacCallum, & Strahan, 1999). This reduced factor yielded new factor scores that mixed scores from the original items together based on loadings, i.e., regression weights (using *fa* from the psych package, see accompanying code; Revelle, 2022). The more one item contributed to - or loaded onto - the reduced factor, the higher the 'item loading' was for that factor. Table 1 shows the item loadings of factors in both experiments. These factor scores were used as a new overall variable that represents a summary of the questionnaire items. Details about each factor analysis (FA) for each experiment are explained below in the experiment-specific methods.

Linear mixed models (LMMs) were run with the factor scores extracted from the factor analysis as the dependent variable, with modality (AV / AO) as a predictor (fixed effect). We also ran LMMs for each physiological measure, where modality was the predictor (fixed effect) as well as a LMM assessing relationship between factor scores and physiological measures. LMMs are more appropriate than repeated measures ANOVA, as they are more fitting for physiological data, can account for missing trials, and can model random sources of variance and non-independence in the observations (Barr, Levy, Scheepers, & Tily, 2013; Page-Gould, 2016; Winter, 2013). Ratings and physiological

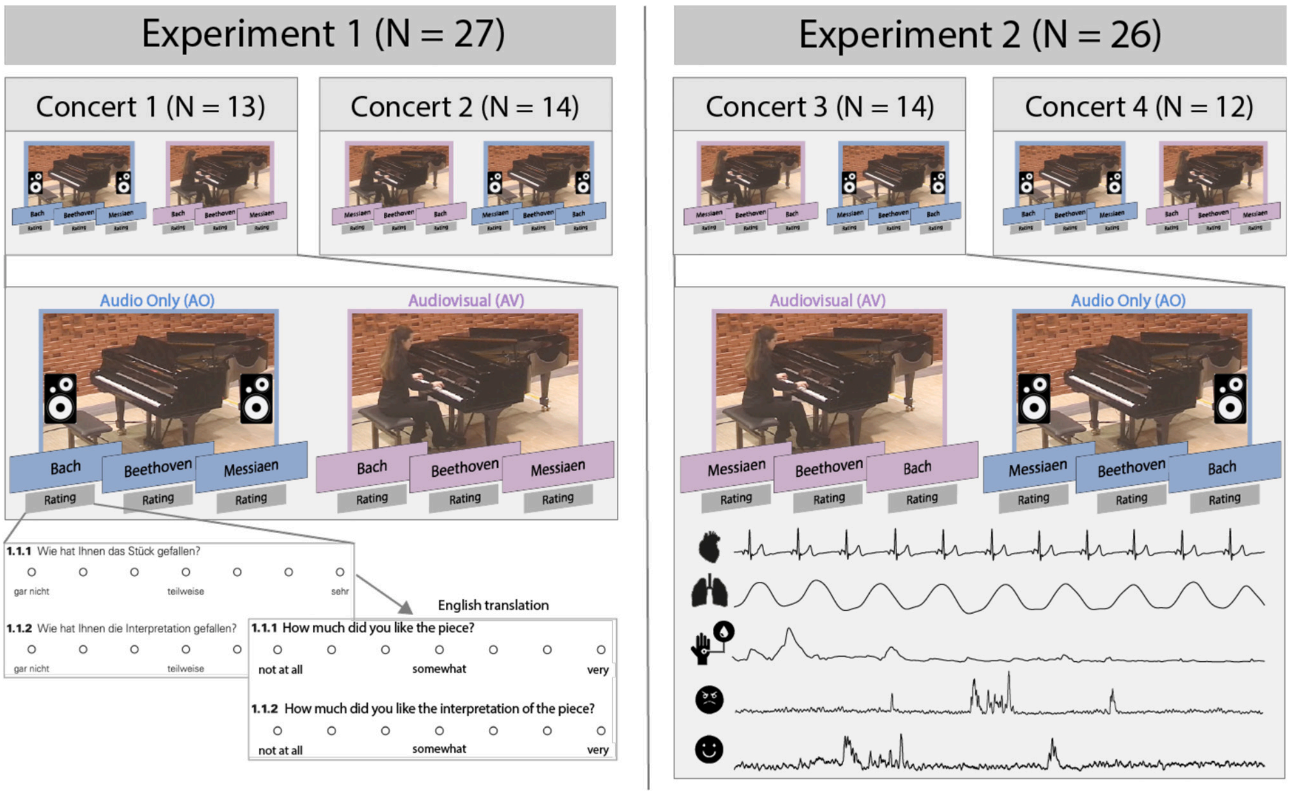

Fig. 1. Outline of the experimental procedure in Experiment 1 (behavioural audience ratings) and Experiment 2 (audience ratings and peripheral physiological measures). Pieces were presented both in an AV version (purple boxes) and an AO version (presented via speakers, blue boxes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
FA loadings from questionnaire items in both Experiment 1 and 2. Factor 1 for both Experiment 1 and 2 is interpreted as 'Aesthetic experience'.

	Experiment 1	Experiment 2
Item	Factor 1	Factor 1
Liking	0.78	0.87
Liking of interpretation	0.69	0.63
Absorption	0.90	0.67
Passive reception	0.73	0.09
Connection to musicians	0.56	0.43
Urge to move	0.17	0.21
Connection to co-listeners	0.28	0.14
Understanding	0.27	0.34

measures were recorded multiple times from each participant, who heard the same music piece more than once, in groups for each concert. To account for these random sources of non-independence, we added random intercepts for concert, piece, and participant. Participants were nested within concerts, while participant and piece were considered crossed effects. For the physiological data, piece sections were further nested within pieces to account for observations taken within pieces (see Methods for Experiment 2). We also included a random slope for participants. Thus, the models represent the maximal random effects structure justified by the design (Arnqvist, 2020; Barr, 2021; Barr et al., 2013). While LMMs do not rely on normally distributed data, we checked linearity, homoscedasticity, and normality of residuals of the models (Winter, 2013). We also checked for model errors. All maximal models generated singular fit errors, suggesting that the model might be too complicated and/or one or more random effects have (near to) zero variance or (near-)perfect correlations. Therefore, we followed the recommended procedure of simplifying models until error is removed (Barr, 2021), ultimately selecting a model with a random effect structure that is supported by the data (Barr et al., 2013; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). As error-free models are generally preferred (Barr et al., 2013), we report the models that generated no errors, but report all maximal and simplified models in the Supplementary Materials. LMMs were run using *lmer* from the *lme4* packages (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). Significance values, effect sizes, and Akaike information criterion (AIC) were obtained from the *tab_model* function from *sjPlot* package (Lüdecke, 2023). Pairwise comparisons were run with the *emmeans* function from *emmeans* package (Lenth, 2021) with Bonferroni corrections. As a sanity check for the linear mixed models, we also ran ANOVAs (Arnqvist, 2020). Corresponding code and required to run these analyses are available at Open Science Framework (OSF): https://osf.io/edzu9/

3. Experiment 1

3.1. Method

3.1.1. Participants

The study was approved by the Ethics Council of the Max Planck Society and in accordance with the Declaration of Helsinki. Participants gave their written informed consent. Twenty-seven participants attended the experimental concerts (13 and 14 participants in Concert 1 and 2, respectively), 18 females (9 males), with mean age of 57.96 years (SD = 20.09), who on average had 6.99 years of music lessons (SD = 7.87) and attended approximately 13 concerts in the last 12 months (M = 12.62; SD = 13.37). Participants also provided ratings on their perception being a musician (from 1 = does not apply, to 7 completely applies), most participants selected 1 (*N* = 13) or 2 (*N* = 4), and less selected 3 (N = 1), 4 (*N* = 2), 5 (*N* = 3), 6 (N = 2) and 7 (N = 2). Most had a college/university degree (*N* = 22), the others either vocational training (N = 2) or completed A-levels/high school (N = 3). Wilcoxon tests showed that participants did not differ in Concert 1 and 2 in terms of age ($p = .590$), musician level ($p = .877$), years of music lessons ($p = 1.00$), and number of concerts attended in the last 12 months ($p = .173$).

3.1.2. *Factor analysis and statistical analysis*

Questionnaire items were chosen to reflect elements of an aesthetic experience. As they were highly correlated (see accompanying code), we chose to reduce these variables to an interpretable factor using factor analysis. A Kaiser-Meyer-Olkin (KMO) measure verified sampling adequacy (KMO = 0.801, well over the 0.5 minimum required) and all KMO values for individual items were >0.670. Bartlett's test of sphericity was significant, revealing that correlations between items were large enough for a FA, $X_2(28) = 408.844$, $p < .001$. Kaiser's criterion of eigenvalues >1 and a scree plot indicated a solution with one factor. Thus, a maximum-likelihood factor analysis was conducted with one factor, which explained 37% of the variance. We took the scores of this factor and created a new variable. As items of liking, liking of interpretation, and absorption loaded highly onto this factor, and these aspects have been identified as critical aspects of an aesthetic experience (Brattico & Pearce, 2013; Orlandi, Cross, & Orgs, 2020), we referred to this new variable as the overall 'aesthetic experience' (AE). Nine trials with an outlier exceeding ±3 Median Absolute Deviations (MAD, Leys, Ley, Klein, Bernard, & Licata, 2013) was removed from further analyses. In total, we had 153 observations for the AE scores [(27 participants × 3 pieces × 2 modality conditions) - 9]. We compared AE factor scores between modality conditions using LMMs (see General Methods, corresponding code).

3.2. *Results*

3.2.1. *Assessing naturalistic situations*

Results of whether the measurements disturbed the concert are shown in Table 2. The mean rating was 1.537 (SD = 1.016) out of 7, with 88% of ratings at 1 or 2 on the scale (i.e., strongly disagree or disagree that measurements disrupted the concerts, respectively). Thus, behavioural measurements did not disrupt the concert, confirming the ecological validity of the experimental setting.

3.2.2. *Piece familiarity*

Ratings for familiarity of style were similarly high for Bach (M = 5.796, SD = 1.279) and Beethoven (M = 5.630, SD = 1.248), but lower for Messiaen (M = 3.333, SD = 1.981). Most participants did not know the pieces specifically, though 18%, 26%, and 11% of participants knew the Bach, Beethoven, and Messiaen pieces, respectively.

Table 2
Ratings of feeling disturbed by the measurement, and familiarity with style and specific piece in Experiment 1.

Ratings of feeling disturbed by the measurement

Rating	1	2	3	4	5	6	7
	69%	19%	5%	3%	3%	1%	0%

Familiarity with style of piece

Rating	1	2	3	4	5	6	7
Bach,	0%	2%	4%	11%	18%	26%	39%
Beethoven,	0%	2%	4%	15%	16%	35%	28%
Messiaen,	26%	17%	9%	20%	9%	11%	8%

Familiarity with specific piece

	0 (No)	1 (Yes)	Not sure
Bach,	78%	18%	4%
Beethoven,	70%	26%	4%
Messiaen,	85%	11%	4%

3.2.3. *Aesthetic experience: Modality differences*

LMMs showed modality was a significant predictor of AE (see Table 4). AV scores were significantly higher (M = 0.186, SE = 0.296, 95% CI [−0.962 1.33]) than AO scores (M = −0.102, SE = 0.297, 95% CI [−1.245, 1.04]), $t(124) = -0.240$, $p = .018$) (see Fig. 2). This effect was confirmed by the maximal model, despite generating a singular fit error: it yielded the same estimates and had similar effect sizes, AIC, and significance (see Supplementary Table 3). The modality effect was confirmed by an ANOVA, which yielded a significant main effect of modality ($F(1,26) = 5.564$, $p = .026$).

3.3. *Discussion*

Experiment 1 tested whether participants had higher AE in the audio-only (AO) or audiovisual (AV) piano performances in a naturalistic concert setting. We confirmed that the measurements did not disturb participants and the findings show that AE increased in the AV compared to AO condition. These results support prior experimental laboratory results that showed liking and appreciation of expressivity are increased in AV conditions (Platz & Kopiez, 2012). We confirm that these results can be extended in a more naturalistic setting. One study that compared emotional differences between modalities in a naturalistic context, found higher wonder ratings but lower boredom ratings in live AV performances of music (Coutinho & Scherer, 2017). Our results likewise fit and extend this work, showing that the preference (liking) and absorption of the AE is also higher in AV modality. As naturalistic environments allow less control, it is important that these findings are replicated.

4. Experiment 2

Previous studies aimed at gaining further insight into potential emotional differences between uni- and bimodal music performances by measuring physiological responses (Chapados & Levitin, 2008; Vuoskoski et al., 2016). However, so far results are inconsistent. In Experiment 2, we explored whether different modalities would affect peripheral physiological responses similarly to the behavioural responses of AE (Exp. 1), and whether peripheral signals might serve as an index of AE.

4.1. *Method*

4.1.1. *Participants*

The study was approved by the Ethics Council of the Max Planck Society and in accordance with the declarations of Helsinki. Participants gave their written informed consent. Twenty-six participants in total attended either Concert 3 ($N = 14$) or Concert 4 ($N = 12$). Experiment 2 in total included nine females (17 males), with a mean age of 51.64 years (SD = 15.41), who on average had 5.94 (SD = 8.13) years of music lessons and attended an average of 14 concerts per year (M = 13.62, SD = 19.70). Participant provided ratings on their perception being a musician (from 1 = does not apply, to 7 completely applies), and most participants selected 1 ($N = 15$) or 2 ($N = 3$), while less selected 3 ($N = 0$), 4 (N = 1), 5 ($N = 4$), 6 ($N = 2$), or 7 (N = 1). All had either vocational training ($N = 7$) or a college/university degree ($N = 19$). Wilcoxon tests showed no significant differences between participants in Concert 3 and Concert 4 in terms of age ($p = .72$), years of music lessons ($p = .14$), and number of concerts attended in the last 12 months ($p = 1.00$). There was a significance in musician level between concerts ($p = .039$).

In assessing differences between the participant samples of the two Experiments, Experiment 1 had a significantly older audience on average (mean age in Experiment 1 = 58, Experiment 2 = 52, $p = .041$), but no significant differences for number of music lessons ($p = .334$),

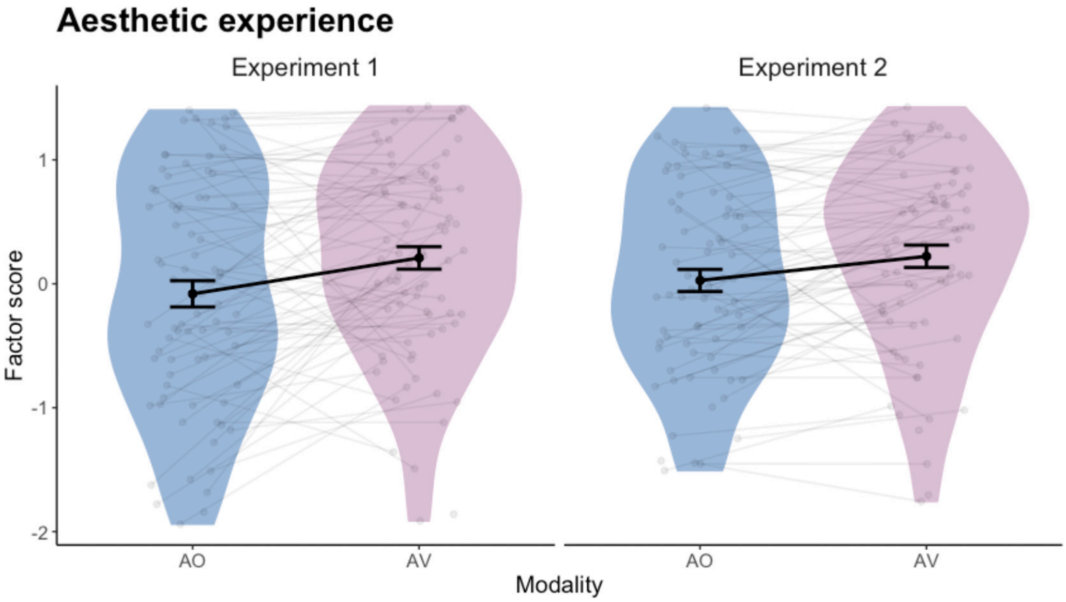

Fig. 2. Aesthetic experience factor scores (which had high item loadings of liking, liking interpretation, and absorption, see Table 1) as a function of modality (Audio Only (AO) is blue and Audiovisual (AV) is purple). The left panel shows results for Experiment 1, while the right panel shows results for Experiment 2. Each point represents factor scores for each participant and each piece. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

concert attendance in the last 12 months ($p = .755$), and musician level ($p = .575$).

Self-report data from all 26 participants were used in the analysis, while one physiological dataset from Concert 3 was lost due to technical problems (physiology: $N = 25$).

4.1.2. Procedure

Participants were invited to arrive an hour before the concert, during which they were fitted with physiological equipment. All signals were collected with a portable recording device, 'plux' (https://plux.info/12-biosignalsplux), that continuously measured physiology across the duration of the concert at a 1000 Hz sampling rate. Respiration was measured via two respiration belts: one respiration belt was placed around the upper chest of the participant, and one respiration belt was placed around the lower belly. ECG, EMG, and EEG were collected using gelled self-adhesive disposable Ag/AgCl electrodes. Locations for the EMG, EEG, and ECG were prepared with peeling gel (under the left eyebrow and on left cheek for EEG, on the chest for ECG, and on the forehead for EEG). Three ECG electrodes were placed on the chest in a triangular arrangement; two as channels and one as the ground. Two facial muscles were recorded on the left side of participants' faces; two electrodes were placed at the zygomaticus major ('smiling') muscle, and two electrodes were placed on the corrugator supercilii ('frowning') muscle, with a ground placed behind the left ear. EDA was collected via two electrodes placed on the middle phalanges of the non-dominant hand of participants. EEG activity from the frontal region was collected from three electrodes placed on the upper forehead, with a reference electrode placed in the middle of the forehead (in a similar location to an Fpz location in a conventional EEG cap), with additional two electrodes placed above the left and right eyebrows (in a similar position to Fp1 and Fp2 in a conventional EEG cap, respectively). EEG data are not reported in this paper.

4.1.3. Factor analysis

We used the same items as in Experiment 1. Again, these item ratings were highly correlated (see accompanying code) and we chose to reduce these variables with a factor analysis. A Kaiser-Meyer-Olkin measure verified sampling adequacy (KMO = 0.609). All but one item had KMO values >0.5; this one item ('connection with co-listeners') had a value close to 0.5 (0.416). Correlations between items were large enough for a FA (Bartlett's test of sphericity, $X_2(28) = 264.725$, $p < .001$. Kaiser's criterion of eigenvalues >1 and a scree plot indicated a solution with one factor. Thus, a maximum likelihood factor analysis was conducted with one factor, which explained 24% of the variance. We took the scores from this factor and created a new variable. As we had similar loadings to Experiment 1, we also refer to this factor as the overall aesthetic experience (AE). In this factor, eleven outlier values exceeding ±3 Median Absolute Deviations (MAD, Leys et al., 2013) were removed from further analyses. In total, we had a total of 145 observations [(26 participants × 3 pieces × 2 modality conditions) - 11].

4.1.4. Physiological pre-processing

Pre-processing of physiological signals (Experiment 2) was conducted in MATLAB (2019b, The Mathworks Inc., USA). Any missing data (gaps ranging from 5 to 53 ms long) were first linearly interpolated at the original sampling rate. Continuous data were then cut per piece. Using Ledalab (www.ledalab.de), skin conductance data were manually screened for artefacts (8% of data were rejected), downsampled to 20 Hz and separated into phasic (SCR) and tonic (SCL) components using Continuous Decomposition Analysis (Benedek & Kaernbach, 2010). Following previous literature, data were detrended to remove remaining long-term drifts (Omigie et al., 2021; cf. Salimpoor et al., 2009). Respiration, ECG, and EMG data were pre-processed using the Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and biosig toolboxes in MATLAB (http://biosig.sourceforge.net/help/index.html). Manual screening of respiration data showed that the respiration signals obtained from the lower belly were stronger than those obtained from the upper chest; only data from the respiration belt around the lower belly were therefore used for further analysis. Respiration data were low-pass filtered at 2 Hz, ECG data were band-pass filtered between 0.6 and 20 Hz (Butterworth, 4th order), and both demeaned. QRS peaks in the ECG signal were extracted using *nqrsdetect* function from biosignal, and peaks were found in respiration using custom functions. Computationally identified peaks were manually screened to ensure correct identification; any missing QRS peaks were manually added, while falsely identified QRS peaks were removed. Any ECG/respiration data that were too noisy for extraction of clear QRS/respiration peaks were rejected from further analysis (ECG = 14%, respiration = 7%). Differential timing of signal peaks – i.e., interbeat intervals (IBI, also known as RR-intervals) for ECG, and inter-breath intervals (IBrI) for respiration – were converted to beats per minute and interpolated at the original sampling rate to obtain a continuous respiration and heart rate. Heart rate variability measures were extracted using the *heartratevariability* function in biosig (http://biosig.sourceforge.net/). Normalised units of high frequency (HF, 0.15–0.4 Hz) power as well as the LF/HF ratio were taken into further analysis to reflect SNS and PNS activity (frequencies that adhere to the European Task Force recommendations (Malik, 1996). Electromyography (EMG) data for zygomaticus major (EMGZM) and corrugator supercilii (EMGCS) were band-pass filtered between 90 and 130 Hz and demeaned. We proceeded with the smoothed absolute value of the Hilbert transformed EMG signals.

Although there are questions as to what the most appropriate (central tendency) representation of physiological data is, we relied most closely on the methodology applied by Vuoskoski et al. (2016) to compare results. Therefore, the average of each (pre-processed) physiological measure was the main metric. As physiological responses change over time (i.e., they are non-stationary), and to gain a better representation (signal-to-noise ratio) of the responses across the course of each long piece, data for each piece were divided into piece sections that were driven by the musical structure (which were confirmed by a music theorist). Responses were averaged across these sections. Beethoven was split into nine, Messiaen into nine, and Bach into seven sections (see Supplementary Materials for more information). Overall, we were interested in eight physiological measures: averages of SCL, SCR, HR, HF power and LF/HF ratio, RR, as well as zygomaticus and corrugator activity, which we averaged per participant, modality, piece, and section. As with behavioural data, we removed outliers exceeding ±3 MAD. Total observations for each physiological measure after exclusion of noisy data and outliers were as follows: EMGCS = 1037, EMGZM = 1082, HR = 1073, HF = 1050, LF/HF ratio = 1041, RR = 1152, SCL = 1066, SCR = 910.

4.1.5. *Analysis*

Statistical analysis for the AE scores obtained in Experiment 2 were conducted as described in Experiment 1. We also compared physiology between AO and AV modalities using LMMs (see General Methods, accompanying code). To determine if behavioural results were related to peripheral responses, we ran a LMM with aesthetic experience as the dependent variable and the eight peripheral measures (all of which were averaged across piece sections to represent rating per piece and scaled to be included in the same model) and condition as predictors. Random effect represented design-driven maximal were included: random intercepts were added for concert, piece, modality condition, and participant. Participants were nested within concerts, while participant, condition, and piece were considered as crossed effects. Variance Inflation Factors (VIF) were checked using the *car* package (Fox & Weisberg, 2019), confirming that VIFs were below 3.

4.2. *Results*

4.2.1. *Assessing naturalistic situations*

We first assessed the extent to which the behavioural/physiological measurements disturbed the overall experience during the concert (i.e., for all pieces/conditions). Ratings suggested that measurements did not disrupt the concert experience, with a mean rating of 2.019 (SD = 1.416) and with 75% of ratings at 1 or 2 on the scale. Results are shown in Table 3. These results provide an important validation that physiological measurements can be used in the concert hall settings without impacting ecological validity.

4.2.2. *Piece familiarity*

Similar to Experiment 1, ratings for familiarity of style were high for Bach (M = 5.385, SD = 1.484) and Beethoven (M = 5.333, SD = 1.532), but lower for Messiaen (M = 4.135, SD = 1.879). Approximately a third knew the Beethoven and Bach pieces, whereas only 19% knew the Messiaen piece.

Table 3
Ratings of feeling disturbed by the measurement and familiarity with style and specific piece in Experiment 2.

Ratings of feeling disturbed by the measurement								na
Rating	1	2	3	4	5	6	7	
	51%	24%	9%	9%	4%	1%	2%	

Familiarity with style of piece								
Rating	1	2	3	4	5	6	7	
Bach	0%	8%	4%	11%	23%	27%	27%	
Beethoven	0%	8%	4%	17%	15%	27%	27%	2%
Messiaen	8%	15%	19%	14%	15%	15%	14%	

Familiarity with piece			
	0	1	Not sure
Bach	65%	35%	0%
Beethoven	67%	33%	0%
Messiaen	79%	19%	2%

4.2.3. *Aesthetic experience: Modality differences*

For the behavioural AE results, LMMs showed modality was a significant predictor of AE (see Table 4) with AV scores significantly higher (M = 0.222, SE = 0.229, 95% CI [−2.07 2.52]) than AO scores (M = 0.003, SE = 0.229, 95% CI [−2.28 2.29], $t(119) = -0.207$, $p = .041$) (see Fig. 2). Although the maximal model generated a singular fit error, it yielded the same estimate and significance, as well as a similar effect size and AIC to the simplified model that generated no error (see Supplementary Table 4). The modality effect was also confirmed by an ANOVA ($F(1,25) = 6.832$, $p = .015$). These results replicated the behavioural findings of Experiment 1.

4.2.4. *Physiological differences between modality*

LMM results are presented in Table 5 (see also Fig. 3). Modality condition was a significant predictor for LF/HF ratio, which represents SNS activation (higher arousal). Comparison of estimated marginal means indicated that this measure was higher in the AO than the AV condition (Table 6). This effect was consistent in the maximal models (see Supplementary Table 8) and confirmed by ANOVA ($F(1,21) = 5.393$, $p = .030$).

Modality was a significant predictor for respiration rate (RR) and corrugator muscle activity (EMGCS), with a significant increase in the AV compared to AO condition (see Tables 5 and 6). However, in the maximal models that generated errors, the modality effect was not significant for EMGCS nor RR (see Supplementary Tables 5 and 10). Corresponding ANOVAs yielded insignificant results for RR ($F(1,22) = 1.95$, $p = .177$), though EMGCS was almost significant ($F(1,21) = 3.679$, $p = .069$). Due to the inconsistency of results between maximal models that generate errors and models with a simplified random structure that is free of errors, findings of EMGCS and RR are only cautiously interpreted.

4.2.5. *Peripheral measures that predict behaviour*

In a model where AE was the dependent variable and all peripheral measures were predictors, zygomaticus activity (EMGZM) was a significant predictors of self-reported AE (see Table 7): increased smiling muscle activity was positively associated with AE.

4.3. *Discussion*

The main aims of Experiment 2 were to replicate the behavioural

Table 4
Linear mixed models for Aesthetic Experience factor scores between modality conditions.

Aesthetic Experience (AE)						
	Experiment 1			Experiment 2		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	−0.10	−0.69–0.48	0.733	0.00	−0.45–0.46	0.987
cond [AV]	0.29	0.05–0.52	**0.018**	0.22	0.01–0.43	**0.040**
Random Effects						
σ^2	0.55			0.40		
τ_{00}	$0.04_{\text{id\_n:concert}}$			$0.16_{\text{id\_n:concert}}$		
	0.24_{piece}			0.08_{concert}		
ICC	0.34			0.37		
N	3_{piece}			2_{concert}		
	$15_{\text{id\_n}}$			$16_{\text{id\_n}}$		
	2_{concert}					
Observations	153			145		
Marginal R2/Conditional R2	0.024 / 0.355			0.019 / 0.383		
AIC	371.967			324.565		

results of Experiment 1 and to gain further insight into peripheral physiological measures as a function of modality. Importantly, subjective ratings again showed that the measurement of physiological signals did not disturb participants.

As in Experiment 1, AE was significantly higher in the AV condition. We further tested whether peripheral responses between modality conditions. Compared to the AV condition, the AO condition evoked higher LF/HF ratio responses. These results support findings of (Vuoskoski et al., 2016), who reported higher physiological arousal in AO musical performances. On the other hand, respiration rate and corrugator muscle activity were higher in the AV condition. As both respiration and facial muscle activity are under voluntary muscle control, one interpretation is that viewing movements of the musician increased motor simulation. This is supported by research showing that viewing effortful movements increases respiration (Brown, Kemp, & Macefield, 2013; Mulder, de Vries, & Zijlstra, 2005; Paccalin & Jeannerod, 2000) and corrugator activity (de Morree & Marcora, 2010). However, inconsistencies occurred for RR and EMGCS in maximal LMMs compared to error-free LMMs. This model discrepancy suggests the modality effect in respiration and facial muscle activity needs to be complemented and confirmed by further studies with larger sample sizes.

When assessing if self-reported AE was predicted by physiological responses, AE was positively associated with zygomaticus activity. However, as increased zygomaticus activity has likewise been related to unpleasant experiences of (dissonant) music (Dellacherie et al., 2011; Merrill et al., 2021), we only cautiously attribute such facial muscle activity with positive AE.

5. General discussion

The current experiments aimed to broaden our understanding of naturalistic concert experiences by testing whether (1) AV information enhances aesthetic experience (AE) in a more ecological setting and (2) peripheral physiological responses are higher in AO or AV modality. We also (3) assess the relationship between AE and peripheral physiological responses. We confirm that in both experiments, the measurement of self-report and physiology did not disturb the audiences, supporting the idea that a semi-experimental setting with naturalistic stimulus presentation can yield results of high ecological validity.

As there are several aspects that can make up an AE (Brattico & Pearce, 2013; Juslin, 2013; Schindler et al., 2017), questionnaire items related to certain aspects of an aesthetic experience were used. In both experiments, these items could be reduced to one factor in a factor analysis. Although the factor had slightly different loadings in the two experiments, three main items consistently loaded highly: absorption, liking, and liking of interpretation. Indeed, liking is a strong element of aesthetic experience both in philosophy (as the *evaluative dimension* of AE, Shusterman, 1997) and in empirical work (Brattico & Pearce, 2013). Preference of interpretation (e.g., how fast or expressive) has likewise been shown to play a strong role in AE. For example, observers prefer an expressive – compared to a non-expressive – interpretation of dance (Christensen, Azevedo, & Tsakiris, 2021). Similarly, dance choreography performed with more varied velocities was rated as more aesthetically pleasing compared to when it is performed with a more uniform velocity (Orlandi et al., 2020). Absorption has also shown to be an important factor in mediating aesthetic experience (Brattico & Pearce, 2013) and can even be indexed by peripheral measures, such as microsaccades (Lange, Zweck, & Sinn, 2017). As these items have a strong connection to AE, it seemed appropriate to refer to this factor as such. Further, the fact that all of these items were correlated with each other and captured well by one factor, corroborates previous research that an aesthetic experience comprises many aspects (Brattico & Pearce, 2013; Merrill et al., 2021) and supports the use of dimensionality reduction techniques which trade specificity in favour of a more holistic AE measure.

Both Experiment 1 and 2 consistently showed that AE increases more in the AV than AO modality consistently across models and ANOVAs. Previous laboratory work has revealed that visual information carries several cues of musical expression (Davidson, 1993; Luck et al., 2010), quality (Tsay, 2013; Waddell & Williamon, 2017) and emotion (Dahl & Friberg, 2007; Van Zijl & Luck, 2013), which enhances aesthetic appreciation (Platz & Kopiez, 2012). Though these findings show that AE was significantly higher in AV than AO music performances, the effect size (just under 0.1) was relatively small (Cohen., 1988), likely due to the small sample size. Nonetheless, the overall model effect size (0.3–0.4) is considered medium (Cohen., 1988).

The current results extend the effect of modality influencing musical appreciation in a naturalistic performance setting. Similar work in a concert setting found that the live AV condition had increased wonder and decreased boredom (Coutinho & Scherer, 2017). However, their main focus was on emotion; we extend their findings to the preference (liking, liking of the interpretation) aspect of AE. We emphasise the importance of conducting AE research in a naturalistic performance setting, as it is more likely to elicit stronger and more realistic responses (Gabrielsson & Wik, 2003; Lamont, 2011). Of note is that results found in laboratory settings are not always replicated in more naturalistic settings. For example, previous laboratory studies have demonstrated that body movement (Platz & Kopiez, 2012) and familiarity (see North & Hargreaves, 2010) increase appreciation of music, even though the latter component has an inverted U-relationship. However, these findings were not replicated in a field study that was conducted in a more realistic situation (busking) and using a dependent variable of appreciation (i.e., money rather than ratings, Anglada-Tort et al., 2019), suggesting that components of music performance influence music

Table 5

Linear mixed models for physiological responses.

Physiological results

	EMGCS			EMGZM			HF		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.0025	0.0021–0.0029	**<0.001**	0.0022	0.0019–0.0024	**<0.001**	0.1384	0.1147–0.1621	**<0.001**
cond [AV]	0.0002	0.0001–0.0003	**<0.001**	0.0001	−0.0000–0.0002	0.180	0.0044	−0.0069–0.0157	0.447
Random Effects									
σ^2	0.00			0.00			0.00		
τ_{00}	$0.00_{\text{section:piece}}$			$0.00_{\text{section:piece}}$			$0.00_{\text{section:piece}}$		
	$0.00_{\text{id\_n:concert}}$			$0.00_{\text{id\_n:concert}}$			$0.00_{\text{id\_n:concert}}$		
τ_{11}				$0.00_{\text{id\_n.condAV}}$			$0.00_{\text{id\_n1.condAO}}$		
				$0.00_{\text{id\_n1.condAO}}$			$0.00_{\text{id\_n2.condAV}}$		
				$0.00_{\text{id\_n2.condAV}}$					
ρ_{01}									
ρ_{01}									
ICC	0.65			0.50			0.47		
N	9_{section}			9_{section}			9_{section}		
	3_{piece}			3_{piece}			3_{piece}		
	$15_{\text{id\_n}}$			$14_{\text{id\_n}}$			$15_{\text{id\_n}}$		
	2_{concert}			2_{concert}			2_{concert}		
Observations	1037			1082			1050		
Marginal R^2 / Conditional R^2	0.007 / 0.657			0.003 / 0.502			0.001 / 0.467		

Physiological results (continued 1)

	LF/HF ratio			HR			RR		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	2.19	1.78–2.61	**<0.001**	62.00	54.96–69.04	**<0.001**	18.31	17.17–19.45	**<0.001**
cond [AV]	−0.26	−0.41 to −0.11	**0.001**	−0.19	−0.44–0.05	0.123	0.26	0.09–0.43	**0.003**
Random Effects									
σ^2	1.52			4.16			2.11		
τ_{00}	$0.00_{\text{section:piece}}$			$0.11_{\text{section:piece}}$			$0.30_{\text{section:piece}}$		
	$0.95_{\text{id\_n:concert}}$			$65.70_{\text{id\_n:concert}}$			$6.71_{\text{id\_n:concert}}$		
				19.99_{concert}			0.08_{concert}		
ICC	0.38			0.95			0.77		
N	9_{section}			2_{concert}			2_{concert}		
	3_{piece}			9_{section}			9_{section}		
	$15_{\text{id\_n}}$			3_{piece}			3_{piece}		
	2_{concert}			$15_{\text{id\_n}}$			$15_{\text{id\_n}}$		
Observations	1041			1073			1152		
Marginal R^2 / Conditional R^2	0.007 / 0.389			0.000 / 0.954			0.002 / 0.771		

Physiological results (continued 2)

	SCR			SCL		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	−0.00	−0.00–0.00	0.156	0.00	−0.02–0.03	0.970
cond [AV]	−0.00	−0.00–0.00	0.754	0.00	−0.01–0.02	0.764
Random Effects						
σ^2	0.00			0.01		
τ_{00}	$0.00_{\text{section:piece}}$			$0.00_{\text{section:piece}}$		
	$0.00_{\text{id\_n:concert}}$					
ICC	0.25			0.21		
N	9_{section}			9_{section}		
	3_{piece}			3_{piece}		
	$14_{\text{id\_n}}$					
	2_{concert}					
Observations	855			910		
Marginal R^2 / Conditional R^2	0.000 / 0.252			0.000 / 0.213		

appreciation differently depending on the context. Overall, despite the fact that a naturalistic setting might allow less control, together with results from previous work (Coutinho & Scherer, 2017), we provide consistent support that audiovisual information enhances AE; a finding that likely generalises to more naturalistic human behaviour.

We further elucidated peripheral responses of AE in multimodal contexts (Experiment 2), as research to date is inconsistent (Chapados & Levitin, 2008; Vuoskoski et al., 2016). Based on the framework of Brattico et al., (2013), we assume AE is made up of perceptual, cognitive, affective, and aesthetic responses (e.g., liking). These components can be relatively well distinguished by self-reports and – to an extent – by different brain regions and event-related brain potentials, depending on their latency and polarity (e.g., early components are related to early sensory processes). However, changes in physiology/facial muscle activity have been related to all of these cognitive, affective, and aesthetic responses (e.g., Roy et al., 2009; Salimpoor et al., 2009; Steinbeis et al., 2006), depending on the design and control condition of the study. Some show physiological changes related to

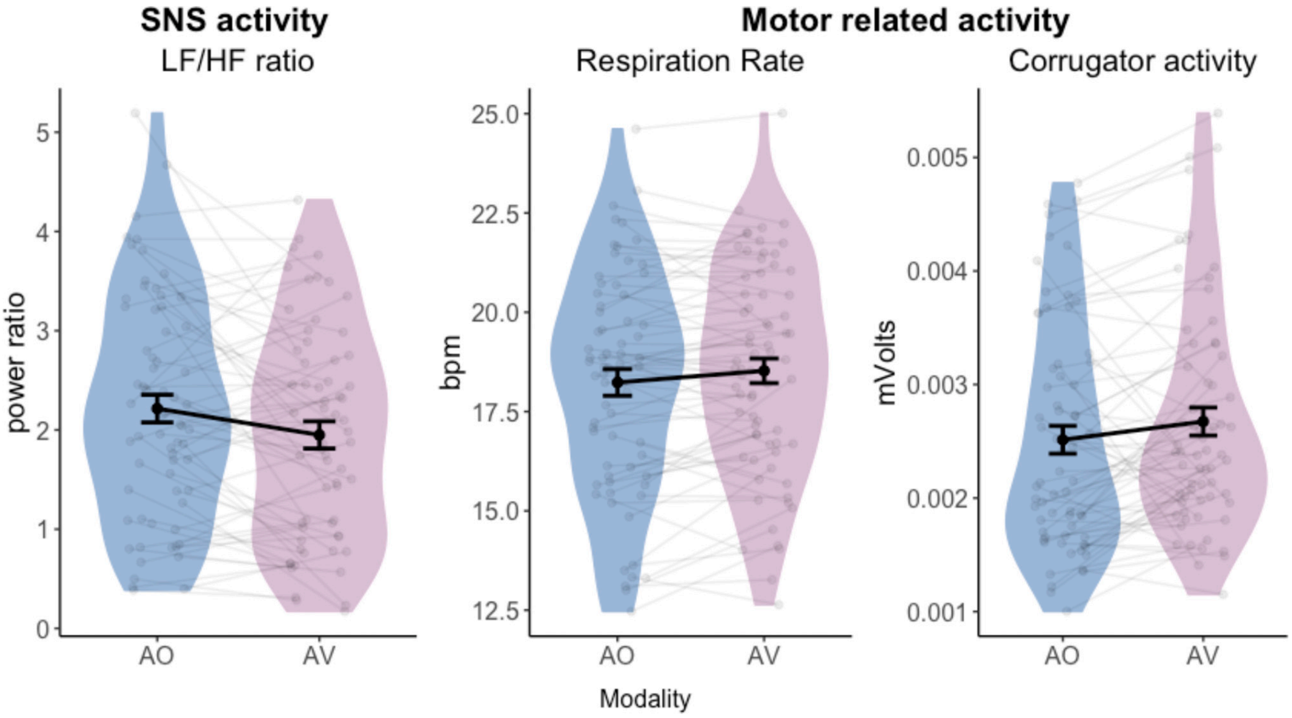

Fig. 3. Physiological responses in each modality condition (AO: blue; AV: purple). Different panels represent different physiological measures; from left to right: LF/HF ratio, respiration rate (RR), and EMG activity of corrugator supercilii (frowning) muscle (Corrugator activity). Each point represents the physiological response value for each participant and each piece. Bold text refers to overall activity, i.e., related to Sympathetic nervous system activity (SNS) or related to motor system. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6

Results of linear mixed models comparing aesthetic experience between AO and AV.

Phys	Estimated Marginal Means		Pairwise difference (AO-AV)			
	AO: M, (SE), [95% CI])	AV: M, (SE), [95% CI])	ß	*SE*	*T*	*p*
EMGCS	**0.0025 (0.0002), [0.0022, 0.0029]**	**0.0026 (0.00012), [0.0023, 0.0031]**	**−0.0002**	**0.000**	**−4.423**	**<0.0001**
EMGZM	0.0021 (0.0001), [0.0019, 0.0025]	0.0023, (0.0001), [0.0019, 0.0026]	−0.000	0.000	−1.334	0.207
HR	62.0 (0.359), [16.5, 107]	61.8 (0.359), [16.3, 107]	0.193	0.125	1.542	0.1234
HF	0.138 (0.0012), [0.112, 0.165]	0.143, (0.014), [0.113, 0.172]	−0.007	0.004	−1.927	0.054
LF/HF ratio	**2.219 (0.211), [1.76, 2.63]**	**1.93 (0.211), [1.50, 2.37]**	**0.26**	**0.077**	**3.380**	**< 0.001**
RR	**18.3 (0.587), [11.9, 24.8]**	**18.6, (0.587), [12.1, 25.0]**	**−0.257**	**0.087**	**−2.972**	**0.003**
SCL	0.0005 (0.013), [−0.026, 0.027]	0.0027 (0.0013), [−0.023, 0.029]	−0.002	0.007	−0.300	0.764
SCR	−0.002 (0.001), [−0.005, 0.001]	−0.002, (0.001), [−0.005, −0.001]	0.0002	0.001	0.313	0.755

sensory (orienting response, e.g., Barry & Sokolov, 1993) and acoustic changes (e.g., Chuen et al., 2016), while others show this activity is related to aesthetic preference (e.g., Grewe et al., 2009; Salimpoor et al., 2009). In further understanding physiological responses, we draw on neural and behavioural evidence that gives better insight into what kind of AE-related processing might take place. On the one hand, responses related to sensory processing should be greater in the AO condition, due to less predictable sound onsets (Jessen & Kotz, 2011), as also shown by Vuoskoski et al. (2016). On the other hand, AV information conveys more emotion (Dahl & Friberg, 2007; Van Zijl & Luck, 2013); therefore, responses could also be higher in the AV condition, as shown in Chapados and Levitin (2008). Thus, we tested again whether physiological responses are higher in AO or AV.

We consistently found that the LF/HF ratio increased in the AO condition. As this measure represents increased SNS activation, this suggests that AO conditions increase physiological arousal, likely reflecting an increase in uncertainty of sound onsets when visual information is absent (Jessen & Kotz, 2011; Klucharev et al., 2003; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). This is in line with results from Vuoskoski et al. (2016), who found that AO evoked more physiological arousal (as shown by skin conductance) compared to AV musical performances. We also support findings by Richardson et al. (2020) who likewise found higher physiological arousal in audio-only, compared to video versions of narratives (e.g., *Games of Thrones* and *Pride and Prejudice*).

We also found partial support for the hypothesis that AV music performances lead to higher peripheral physiological responses than in AO performances. We state partial evidence, as design-driven LMMs differed from error-free ones. Simplified, error-free models revealed a significant modality effect for RR and EMGCS. Maximal models, which generated errors, did not. These differences could be attributed to the fact that removing the slopes to avoid singularity fit errors could have increased degrees of freedom and the possibility of Type 1 errors (Arnqvist, 2019). However, a model with a complex random-effects structure can lead to increased Type II error and lack of power (Barr, 2021; Matuschek et al., 2017). Thus, future studies with larger sample sizes are required to confirm this modality effect. As there is general consensus that error-free models are preferable (Barr et al., 2013), these models are reported. Nonetheless, we aim to be transparent; the reader is pointed to not only the Supplementary Materials, but also the code showing the maximal models and how models are simplified step by step. While only cautiously interpreting the modality effects in RR and

Table 7
Model of physiology predicting AE self-reports.

LMM comparison physiology predicting Aesthetic Experience (AE)			
	Maximal model		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.1940	−0.1535–0.5415	0.270
SCL	0.0982	−0.0737–0.2700	0.259
SCR	−0.0108	−0.1884–0.1668	0.904
HR	−0.0500	−0.2333–0.1333	0.589
RR	0.0628	−0.0984–0.2241	0.440
HFnu	−0.0007	−0.2729–0.2715	0.996
LFHFratio	−0.0351	−0.3110–0.2409	0.801
EMGCS	−0.0767	−0.2508–0.0973	0.383
EMGZM	0.1828	0.0196–0.3460	**0.029**
Random Effects			
σ^2			0.49
$\tau_{00\ \text{id\_n:concert}}$			0.06
$\tau_{00\ \text{piece}}$			0.00
$\tau_{00\ \text{cond}}$			0.01
$\tau_{00\ \text{concert}}$			0.03
ICC			0.18
N_{concert}			2
N_{piece}			3
N_{cond}			2
$N_{\text{id\_n}}$			13
Observations			94
Marginal R^2 / Conditional R^2			0.065 / 0.234
AIC			256.674

EMGCS, we believe it is worth briefly discussing the results from error-free models.

RR was faster in the AV condition. 'Frowning' muscle (EMGCS) activity, which typically reflects negative valence (Bradley & Lang, 2000), also increased in the AV condition. The discrepancy between the increase in both frowning muscle activity and (generally positive) AE in the AV condition could be explained by the fact that higher aesthetic pleasure can also derive from perceiving negatively valenced musical expression and/or affective states (Eerola et al., 2018), such as being moved (Eerola et al., 2016). However, some question whether facial expressions reflect valence (Wingenbach et al., 2020) or affective states at all (Lewis, 2011; Matsumo, 1987). Thus, another possible interpretation is that observing the musician increased mimicry in the observers. Indeed, participants mimic observed facial expressions (Dimberg, 1982; Magnee et al., 2007). Additionally, viewing effortful movements increases respiration (Brown et al., 2013; Mulder et al., 2005; Paccalin & Jeannerod, 2000) and corrugator activity (de Morree & Marcora, 2010). Such motor mimicry likely extends to music performance. Motor activity increases when listening to music (Bangert et al., 2006; Grahn & Brett, 2007; Janata et al., 2012), especially in audiovisual performances (Chan et al., 2013; Griffiths & Reay, 2018). Indeed, sensorimotor embodied mechanisms related to motor mimicry have been proposed and shown to enhance AE (Brattico & Pearce, 2013; Cross, 2011; Freedberg & Gallese, 2007; Gallese & Freedberg, 2007). Thus, faster breathing and increased facial muscle activity in AV conditions may be a reflection of motor mimicry that occurs when viewing musicians' movements. In sum, we provide partial evidence of a modality effect in RR and EMGCS, potentially reflecting motor mimicry.

Facial muscle activity was significantly associated with AE. The zygomaticus ('smiling') muscle activity was a significant predictor for AE scores. Increased zygomaticus activity was positively related to AE, supporting previous work showing that zygomaticus activity was higher for pleasant music (Fuentes-Sánchez, Pastor, Eerola, Escrig, & Pastor, 2022), liked positive music (Witvliet & Vrana, 2007), positively evaluated art (Gernot, Pelowski, & Leder, 2018), and liked dance movements (Kirsch, Snagg, Heerey, & Cross, 2016). This is further support for the embodied aesthetics theory, where sensorimotor embodied mechanisms might enhance AE (Brattico & Pearce, 2013; Cross, 2011; Freedberg & Gallese, 2007; Gallese & Freedberg, 2007). However, increased 'smiling' muscle activity has also been shown to increase in unpleasant (dissonant) music, suggesting that such activity might represent a grimace or ironic laughter (Dellacherie et al., 2011; Merrill et al., 2021). Therefore, it is vital to collect self-report data to support interpretations of physiological responses, rather than considering certain responses a direct index of a specific state, especially over a long period of time in such naturalistic settings.

LMMs show that LF/HF ratio were higher in AO, and tentative evidence suggests that respiration and muscle activity were higher in AV. These findings can be considered in conjunction with how much (in)voluntary control we have over them. As mentioned before, the heart is innervated by the ANS and made up of involuntary (cardiac) muscle. Voluntary skeletal muscles control EMG and (partly) respiration. On the one hand, due to the automatic nature of the heart, it seems plausible these might be more related to earlier (sensory) processes of an AE. On the other hand, the more voluntary peripheral measures seem to be related to the liking aspect of AE. Although we are cautious to attribute the increase of such measures as a direct index of aesthetic experience, the results point to the idea that the more voluntary the control of the peripheral measure, the more related it may be to later stages of the aesthetic processing, as outlined in Brattico and Pearce (2013).

One overall limitation of the current study is that although all versions were presented as part of a concert while participants were seated in the concert hall, AV was presented as a live version, while AO was presented as a playback. This was chosen to enhance ecological validity: people who listen to music in an AO version most likely listen to music as playback, while watching an AV version is more likely to be live (Sloboda et al., 2012). Indeed, this difference of visual information is also showed in Swarbrick et al. (2019), who similarly stated that AV performances are typically live. Although we do appreciate that tools and streaming platforms like YouTube, Digital Concert Hall of the Berliner Philharmoniker and MetOnDemand etc. have increased in popularity (especially with the COVID-19 pandemic) making audiovisual recording more popular, Belfi et al. (2021) found that felt pleasure did not differ between live and an audiovisual recording of that same performance. Therefore, it is likely that the live and playback differences do not play a strong role in influencing the current results. Future research might consider live audio-only playback of an offstage performer to fully mitigate this potential confound. Another limitation is that although the pieces were chosen to represent typical concert pieces (and a range of genres), they were not controlled for length. Nonetheless, length was a compromise when using naturalistic stimuli that heightened ecological validity. As we did not look at piece-specific differences, but rather average across sections of the pieces to examine the effect of condition, we did not consider this a confound in the current study. However, we note that effects driven by one piece may weigh our results more heavily than effects from the shorter pieces. Future research might consider choosing pieces of similar length, or at least similar lengths of sections. A further limitation is that we did not contrast visual only information with the other two conditions. This choice was a compromise to keep the within-in subject design time-manageable as well as to create a concert-like feel for the experiment.

6. Conclusion

Researchers are increasingly foregoing ultimate control for a more ecologically valid approach that enables participants to have more powerful aesthetic experiences. This study follows others that have moved more into the 'wild' to explore such naturalistic experiences (Chabin et al., 2022; Czepiel et al., 2021; Dotov & Trainor, 2021; Merrill et al., 2021; Swarbrick et al., 2019; Tervaniemi et al., 2021). The current findings show that a self-reported aesthetic experience significantly increases in audiovisual (compared to audio only) piano performances in the naturalistic setting of a concert hall.

Modality additionally influenced peripheral measures, revealing two main patterns. On the one hand, involuntary a physiological arousal response (heart rhythm reflecting SNS), was higher in the (less predictive) AO modality, likely reflecting more sensory processes. On the other hand, peripheral responses with more voluntary control (respiration, facial muscle activity) were higher in the AV modality, though due to inconsistencies in maximal/error-free models, these results should be interpreted with caution. The zygomaticus muscle was a significant predictor of self-reported AE. It could be that the involuntary-voluntary continuum of physiological responses is related to a sensory-affective continuum of AEs. We also suggest that visual information enhances motor mimicry (as shown by an increase in respiration and facial muscle activity), which is a mechanism that enhances AE (Cross et al., 2011; Freedberg & Gallese, 2007; Gallese & Freedberg, 2007; Kirsch et al., 2016). By exploring modality effects, we postulate that peripheral responses likely reflect sensory, sensorimotor, and affective responses that may culminate into an overall aesthetic experience (Brattico et al., 2013). However, we would like to emphasise that such peripheral responses alone cannot directly index AE; self-reports should support interpretations of peripheral physiological data. Nonetheless, the extent that physiological responses are simply sensory or reflect intertwined sensory and affective aspects of the aesthetic experience remains unclear. Further research, with larger sample sizes, should assess the robustness of the effects discussed here. To gain more insight, future research could bridge this gap by further exploring whether this involuntary-voluntary continuum reflects such sensory-aesthetic continuum and whether - and to what extent - there is an overlap of such systems.

CRediT authorship contribution statement

Anna Czepiel: Conceptualization, Investigation, Formal analysis, Visualization. **Lauren K. Fink:** Formal analysis, Visualization, Writing – review & editing. **Christoph Seibert:** Methodology, Writing – review & editing. **Mathias Scharinger:** Investigation, Formal analysis, Writing – review & editing. **Sonja A. Kotz:** Formal analysis, Writing – review & editing.

Declaration of Competing Interest

None.

Data availability

In the manuscript text, there is a link to the OSF repository where data and code have been made available.

Acknowledgements

The authors thank Lea T. Fink for helping with music theoretical analysis, as well as the ArtLab team and assistants for help for the concerts. We thank Klaus Frieler for advice on statistical analysis and the Music Department of MPI EA in Frankfurt and Melanie Wald-Fuhrmann for discussions about the selection of musical pieces. We also thank the anonymous reviewers who provided insightful feedback for improving the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105537.

References

Anglada-Tort, M., & Skov, M. (2022). What counts as aesthetics in science? A bibliometric analysis and visualization of the scientific literature from 1970 to 2018. *Psychology of Aesthetics, Creativity, and the Arts, 16*(3), 553–568.

Anglada-Tort, M., Thueringer, H., & Omigie, D. (2019). The busking experiment: A field study measuring behavioral responses to street music performances. *Psychomusicology: Music, Mind, and Brain, 29*(1), 46–55. https://doi.org/10.1037/pmu0000236

Ara, A., & Marco-Pallarés, J. (2020). Fronto-temporal theta phase-synchronization underlies music-evoked pleasantness. *NeuroImage, 212*, Article 116665. https://doi.org/10.1016/j.neuroimage.2020.116665

Ara, A., & Marco-Pallarés, J. (2021). Different theta connectivity patterns underlie pleasantness evoked by familiar and unfamiliar music. *Scientific Reports, 11*(1), 18523. https://doi.org/10.1038/s41598-021-98033-5

Ardizzi, M., Calbi, M., Tavaglione, S., Umiltà, M. A., & Gallese, V. (2020). Audience spontaneous entrainment during the collective enjoyment of live performances: Physiological and behavioral measurements. *Scientific Reports, 10*(1), 3813. https://doi.org/10.1038/s41598-020-60832-7

Arnqvist, G. (2020). Mixed models offer no freedom from degrees of freedom. *Trends in Ecology & Evolution, 35*(4), 329–335. https://doi.org/10.1016/j.tree.2019.12.004

Barr, D. J. (2021). Learning statistical models through simulation in R: An interactive textbook (Version 1.0.0). https://psyteachr.github.io/stat-models-v1.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F., & Russell, J. A. (2015). *The psychological construct of emotion*. The Guilford Press.

Barry, R. J. (1975). Low-intensity auditory stimulation and the GSR orienting response. *Physiological Psychology, 3*(1), 98–100. https://doi.org/10.3758/BF03326832

Barry, R. J., & Sokolov, E. N. (1993). Habituation of phasic and tonic components of the orienting reflex. *International Journal of Psychophysiology, 15*(1), 39–42. https://doi.org/10.1016/0167-8760(93)90093-5

Bartlett, D. L. (1996). Physiological responses to music and sound stimuli. In *Handbook of music psychology* (pp. 343–385).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Behne, K.-E., & Wöllner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae, 15*(3), 324–342. https://doi.org/10.1177/1029864911410955

Belfi, A. M., Samson, D. W., Crane, J., & Schmidt, N. L. (2021). Aesthetic judgments of live and recorded music: Effects of congruence between musical artist and piece. *Frontiers in Psychology, 12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.618025.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Bernardi, L., Porta, C., & Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: The importance of silence. *Heart, 92*(4), 445–452. https://doi.org/10.1136/hrt.2005.064600

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. https://doi.org/10.3389/fphys.2017.00785

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences, 98*(20), 11818–11823. https://doi.org/10.1073/pnas.191355898

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37*(2), 204–215. https://doi.org/10.1111/1469-8986.3720204

Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts, 7*(1), 48–61. https://doi.org/10.1037/a0031624

Brattico, E., Bogert, B., & Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00206

Broughton, M., & Stevens, C. (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music, 37*(2), 137–153. https://doi.org/10.1177/0305735608094511

Brown, R., Kemp, U., & Macefield, V. (2013). Increases in muscle sympathetic nerve activity, heart rate, respiration, and skin blood flow during passive viewing of exercise. *Frontiers in Neuroscience, 7*. https://www.frontiersin.org/article/10.3389/fnins.2013.00102.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In *The handbook of emotion* (2nd ed., pp. 173–191). Guildford Press.

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings* *of the 18th ACM international conference on multimedia* (pp. 1467–1468).

Chabin, T., Gabriel, D., Chansophonkul, T., Michelant, L., Joucla, C., Haffen, E., … Pazart, L. (2020). Cortical patterns of pleasurable musical chills revealed by high-density EEG. *Frontiers in Neuroscience, 14*, Article 565815. https://doi.org/10.3389/fnins.2020.565815

Chabin, T., Gabriel, D., Comte, A., Haffen, E., Moulin, T., & Pazart, L. (2022). Interbrain emotional connection during music performances is driven by physical proximity and individual traits. *Annals of the New York Academy of Sciences, 1508*(1), 178–195. https://doi.org/10.1111/nyas.14711

Chaffin, R., Lemieux, A. F., & Chen, C. (2007). 'It is different each time I play': Variability in highly prepared musical performance. *Music Perception, 24*(5), 455–472. https://doi.org/10.1525/mp.2007.24.5.455

Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates. *Cognition, 108*(3), 639–651. https://doi.org/10.1016/j.cognition.2008.05.008

Christensen, J. F., Azevedo, R. T., & Tsakiris, M. (2021). Emotion matters: Different psychophysiological responses to expressive and non-expressive full-body movements. *Acta Psychologica, 212*, Article 103215. https://doi.org/10.1016/j.actpsy.2020.103215

Chuen, L., Sears, D., & McAdams, S. (2016). Psychophysiological responses to auditory change: Psychophysiological responses to auditory change. *Psychophysiology, 53*(6), 891–904. https://doi.org/10.1111/psyp.12633

Cohen.. (1988). Set correlation and contingency tables. *Applied Psychological Measurement, 12*(4), 425–434. https://doi.org/10.1177/014662168801200410

Coutinho, E., & Cangelosi, A. (2011). Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion, 11*(4), 921–937. https://doi.org/10.1037/a0024700

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. https://doi.org/10.1177/0305735616670496

Craig, D. G. (2005). An exploratory study of physiological changes during "chills" induced by music. *Musicae Scientiae, 9*(2), 273–287. https://doi.org/10.1177/102986490500900207

Cross, E. (2011). The impact of aesthetic evaluation and physical ability on dance perception. *Frontiers in Human Neuroscience, 5*. https://doi.org/10.3389/fnhum.2011.00102

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. https://doi.org/10.1038/s41598-021-00492-3

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in Musicians' body movements. *Music Perception, 24*(5), 433–454. https://doi.org/10.1525/mp.2007.24.5.433

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music, 21*(2), 103–113. https://doi.org/10.1177/030573569302100201

de Morree, H. M., & Marcora, S. M. (2010). The face of effort: Frowning muscle activity reflects effort during a physical task. *Biological Psychology, 85*(3), 377–382. https://doi.org/10.1016/j.biopsycho.2010.08.009

Dellacherie, D., Roy, M., Hugueville, L., Peretz, I., & Samson, S. (2011). The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance: Psychophysiology of musical emotion. *Psychophysiology, 48*(3), 337–349. https://doi.org/10.1111/j.1469-8986.2010.01075.x

Di Bernardi Luft, C., & Bhattacharya, J. (2015). Aroused with heart: Modulation of heartbeat evoked potential by arousal induction and its oscillatory correlates. *Scientific Reports, 5*(1), 15717. https://doi.org/10.1038/srep15717

Dillman Carpentier, F. R., & Potter, R. F. (2007). Effects of music on physiological arousal: Explorations into tempo and genre. *Media Psychology, 10*(3), 339–363. https://doi.org/10.1080/15213260701533045

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*(1), 86–89. https://doi.org/10.1111/1467-9280.00221

Dotov, D., & Trainor, L. J. (2021). Cross-frequency coupling explains the preference for simple ratios in rhythmic behaviour and the relative stability across non-synchronous patterns. *Philosophical Transactions of the Royal Society B: Biological Sciences, 376*(1835), 20200333. https://doi.org/10.1098/rstb.2020.0333

Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience, 13*(3), 533–553. https://doi.org/10.3758/s13415-013-0161-y

Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2015). Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese pygmies. *Frontiers in Psychology, 5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.01341.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fink, L. K., Simola, J., Tavano, A., Lange, E., Wallot, S., & Laeng, B. (2023). From pre-processing to advanced dynamic modeling of pupil data. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02098-1.

Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences, 11*(5), 197–203. https://doi.org/10.1016/j.tics.2007.02.003

Frith, C. D., & Allen, H. A. (1983). The skin conductance orienting response as an index of attention. *Biological Psychology, 17*(1), 27–39. https://doi.org/10.1016/0301-0511(83)90064-9

Fuentes-Sánchez, N., Pastor, R., Eerola, T., Escrig, M. A., & Pastor, M. C. (2022). Musical preference but not familiarity influences subjective ratings and psychophysiological correlates of music-induced emotions. *Personality and Individual Differences, 198*, Article 111828. https://doi.org/10.1016/j.paid.2022.111828

Gabrielsson, A., & Wik, S. L. (2003). Strong experiences related to music: A descriptive system. *Musicae Scientiae, 7*(2), 157–217. https://doi.org/10.1177/102986490300700201

Gallese, V., & Freedberg, D. (2007). Mirror and canonical neurons are crucial elements in esthetic response. *Trends in Cognitive Sciences, 11*(10), 411. https://doi.org/10.1016/j.tics.2007.07.006

Gernot, G., Pelowski, M., & Leder, H. (2018). Empathy, Einfühlung, and aesthetic experience: The effect of emotion contagion on appreciation of representational and abstract art using fEMG and SCR. *Cognitive Processing, 19*(2), 147–165. https://doi.org/10.1007/s10339-017-0800-2

Goodchild, M., Wild, J., & McAdams, S. (2019). Exploring emotional responses to orchestral gestures. *Musicae Scientiae, 23*(1), 25–49. https://doi.org/10.1177/1029864917704033

Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological Bulletin, 65*(5), 305–320. https://doi.org/10.1037/h0023258

Grewe, O., Kopiez, R., & Altenmüüller, E. (2009). The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception, 27*(1), 61–74. https://doi.org/10.1525/mp.2009.27.1.61

Griffiths, N. K., & Reay, J. L. (2018). The relative importance of aural and visual information in the evaluation of Western canon music performance by musicians and nonmusicians. *Music Perception, 35*(3), 364–375. https://doi.org/10.1525/mp.2018.35.3.364

Halpern, A., Chan, C., Müllensiefen, D., & Sloboda, J. (2017). Audience reactions to repeating a piece on a concert programme. *Participations, 14*(2), 135–152.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in Cognitive Sciences, 16*(9), 458–466. https://doi.org/10.1016/j.tics.2012.07.006

Hodges, D. (2009). Bodily responses to music. In *The Oxford handbook of music psychology* (pp. 121–130).

Huron, D. (2006). *Sweet anticipation*. Cambridge, MA: MIT Press.

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage, 58*(2), 665–674. https://doi.org/10.1016/j.neuroimage.2011.06.035

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. https://doi.org/10.1016/j.plrev.2013.05.008

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*(5), 559–575. https://doi.org/10.1017/S0140525X08005293

Kappenman, E. S., & Luck, S. J. (2011). *ERP components: The ups and downs of brainwave recordings*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0014

Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters, 328*(2), 145–149. https://doi.org/10.1016/S0304-3940(02)00462-7

Kirsch, L. P., Snagg, A., Heerey, E., & Cross, E. S. (2016). The impact of experience on affective responses during action observation. *PLoS One, 11*(5), Article e0154681. https://doi.org/10.1371/journal.pone.0154681

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research, 18*(1), 65–75. https://doi.org/10.1016/j.cogbrainres.2003.09.004

Koelsch, S., & Jäncke, L. (2015). Music and the heart. *European Heart Journal, 36*(44), 3043–3049. https://doi.org/10.1093/eurheartj/ehv430

Koelsch, S., Kilches, S., Steinbeis, N., & Schelinski, S. (2008). Effects of unexpected chords and of Performer's expression on brain responses and Electrodermal activity. *PLoS One, 3*(7), Article e2631. https://doi.org/10.1371/journal.pone.0002631

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 51*(4), 336–353. https://doi.org/10.1037/1196-1961.51.4.336

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). {lmerTest} Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lamont, A. (2011). University students' strong experiences of music: Pleasure, engagement, and meaning. *Musicae Scientiae, 15*(2), 229–249. https://doi.org/10.1177/1029864911403368

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Lange, E. B., & Frieler, K. (2018). Challenges and opportunities of predicting musical emotions with perceptual and automatized features. *Music Perception, 36*(2), 217–242. https://doi.org/10.1525/mp.2018.36.2.217

Lange, E. B., Zweck, F., & Sinn, P. (2017). Microsaccade-rate indicates absorption by music listening. *Consciousness and Cognition, 55*, 59–78. https://doi.org/10.1016/j.concog.2017.07.009

Lange, E. B., Fünderich, J., & Grimm, H. (2022). Multisensory integration of musical emotion perception in singing. *Psychological Research*. https://doi.org/10.1007/s00426-021-01637-9

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology, 40*(5), 776–785. https://doi.org/10.1111/1469-8986.00078

Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects*. Bordeuax.

Lenth, R. V. (2021). Emmeans: Estimated marginal means, aka least-squares means. https://CRAN.R-project.org/package=emmeans.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Luck, G., Toiviainen, P., & Thompson, M. R. (2010). Perception of expression in Conductors' gestures: A continuous response study. *Music Perception, 28*(1), 47–57. https://doi.org/10.1525/mp.2010.28.1.47

Lundqvist, L.-O., Carlsson, F., Hilmersson, P., & Juslin, P. N. (2008). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music, 37*(1). https://doi.org/10.1177/0305735607086048

Lyytinen, H., Blomberg, A.-P., & Näätänen, R. (1992). Event-related potentials and autonomic responses to a change in unattended auditory stimuli. *Psychophysiology, 29*(5), 523–534. https://doi.org/10.1111/j.1469-8986.1992.tb02025.x

Lüdecke, D. (2023). *sjPlot: Data Visualization for Statistics in Social Science* (R package version 2.8.14).

Malik, M. (1996). Heart rate variability.: Standards of measurement, physiological interpretation, and clinical use: Task force of the European Society of Cardiology and the north American Society for Pacing and Electrophysiology. *Annals of Noninvasive Electrocardiology, 1*(2), 151–181. https://doi.org/10.1111/j.1542-474X.1996.tb00275.x

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Merrill, J., Czepiel, A., Fink, L. T., Toelle, J., & Wald-Fuhrmann, M. (2021). The aesthetic experience of live concerts: Self-reports and psychophysiology. *Psychology of Aesthetics, Creativity, and the Arts, 17*(2), 134–151. https://doi.org/10.1037/aca0000390

Morrison & Selvey. (2014). The effect of conductor expressivity on choral ensemble evaluation. *Bulletin of the Council for Research in Music Education, 199*, 7. https://doi.org/10.5406/bulcouresmusedu.199.0007

Mulder, T., de Vries, S., & Zijlstra, S. (2005). Observation, imagination and execution of an effortful movement: More evidence for a central explanation of motor imagery. *Experimental Brain Research, 163*(3), 344–351. https://doi.org/10.1007/s00221-004-2179-4

Nemati, S., Akrami, H., Salehi, S., Esteky, H., & Moghimi, S. (2019). Lost in music: Neural signature of pleasure and its role in modulating attentional resources. *Brain Research, 1711*, 7–15. https://doi.org/10.1016/j.brainres.2019.01.011

Omigie, D., Frieler, K., Bär, C., Muralikrishnan, R., Wald-Fuhrmann, M., & Fischinger, T. (2021). Experiencing musical beauty: Emotional subtypes and their physiological and musico-acoustic correlates. *Psychology of Aesthetics, Creativity, and the Arts, 15*(2), 197–215. https://doi.org/10.1037/aca0000271

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 1–9. https://doi.org/10.1155/2011/156869

Orlandi, A., Cross, E. S., & Orgs, G. (2020). Timing is everything: Dance aesthetics depend on the complexity of movement kinematics. *Cognition, 205*, Article 104446. https://doi.org/10.1016/j.cognition.2020.104446

Paccalin, C., & Jeannerod, M. (2000). Changes in breathing during observation of effortful actions. *Brain Research, 862*(1–2), 194–200. https://doi.org/10.1016/S0006-8993(00)02145-4

Page-Gould, E. (2016). Multilevel Modeling. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (4th ed., pp. 662–678). Cambridge University Press. https://doi.org/10.1017/9781107415782.030.

Pelowski, M., Markey, P. S., Lauring, J. O., & Leder, H. (2016). Visualizing the impact of art: An update and comparison of current psychological models of art experience. *Frontiers in Human Neuroscience, 10*. https://www.frontiersin.org/article/10.3389/fnhum.2016.00160.

Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception, 30*(1), 71–83. https://doi.org/10.1525/mp.2012.30.1.71

Purves, D., & Williams, S. M. (Eds.). (2001). *Neuroscience* (2nd ed.). Sunderland (MA): Sinauer Associates https://www.ncbi.nlm.nih.gov/books/NBK10799/.

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/.

Revelle, W. (2022). Package 'psych'. https://personality-project.org/r/psych-manual.pdf.

Richardson, D. C., Griffin, N. K., Zaki, L., Stephenson, A., Yan, J., Curry, T., … Devlin, J. T. (2020). Engagement in video and audio narratives: Contrasting self-report and physiological measures. *Scientific Reports, 10*(1), 11298. https://doi.org/10.1038/s41598-020-68253-2

Roy, M., Mailhot, J.-P., Gosselin, N., Paquette, S., & Peretz, I. (2009). Modulation of the startle reflex by pleasant and unpleasant music. *International Journal of Psychophysiology, 71*(1), 37–42. https://doi.org/10.1016/j.ijpsycho.2008.07.010

RStudio Team. (2021). *RStudio: Integrated development environment for R*. RStudio, PBC. http://www.rstudio.com/.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS One, 4*(10), Article e7487. https://doi.org/10.1371/journal.pone.0007487

Sammler, D., Grigutsch, M., Fritz, T., & Koelsch, S. (2007). Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology, 44*(2), 293–304. https://doi.org/10.1111/j.1469-8986.2007.00497.x

Scherer, K. R., Trznadel, S., Fantini, B., & Coutinho, E. (2019). Comments on comments by Cupchik (2019) and Jacobsen (2019). *Psychology of Aesthetics, Creativity, and the Arts, 13*(3), 264–265. https://doi.org/10.1037/aca0000246

Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., & Scherer, K. R. (2017). Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLoS One, 12*(6), Article e0178899. https://doi.org/10.1371/journal.pone.0178899

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1791–1810. https://doi.org/10.1037/a0016455

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36*(6), 888–897. https://doi.org/10.1068/p5635

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health, 5*. https://www.frontiersin.org/article/10.3389/fpubh.2017.00258.

Shusterman, R. (1997). The end of aesthetic experience. *The Journal of Aesthetics and Art Criticism, 55*(1), 29. https://doi.org/10.2307/431602

Sloboda, J. A., & O'Neill, S. A. (2001). Emotions in everyday listening to music. In *Music and emotion: Theory and research* (pp. 415–429). Oxford University Press.

Sloboda, J. A., Lamont, A., & Greasley, A. (2012). In S. Hallam, I. Cross, & M. Thaut (Eds.)*, Vol. 1. Choosing to hear music*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199298457.013.0040.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience, 18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Stekelenburg, J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid Audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Swarbrick, D., Bosnyak, D., Livingstone, S. R., Bansal, J., Marsh-Rollo, S., Woolhouse, M. H., & Trainor, L. J. (2019). How live music moves us: Head movement differences in audiences to live versus recorded music. *Frontiers in Psychology, 9*, 2682. https://doi.org/10.3389/fpsyg.2018.02682

Tervaniemi, M., Pousi, S., Seppälä, M., & Makkonen, T. (2021). Brain oscillation recordings of the audience in a live concert-like setting. *Cognitive Processing, 23*(2). https://doi.org/10.1007/s10339-021-01072-z

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences, 110*(36), 14580–14585. https://doi.org/10.1073/pnas.1221454110

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, 102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Van Zijl, A. G. W., & Luck, G. (2013). Moved through music: The effect of experienced emotions on performers' movement characteristics. *Psychology of Music, 41*(2), 175–197. https://doi.org/10.1177/0305735612458334

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. https://doi.org/10.1016/j.cognition.2005.09.003

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. https://doi.org/10.1016/j.cognition.2010.11.010

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*(2), 591–604. https://doi.org/10.3758/s13414-013-0582-2

Vuoskoski, J. K., Gatti, E., Spence, C., & Clarke, E. F. (2016). Do visual cues intensify the emotional responses evoked by musical performance? A psychophysiological investigation. *Psychomusicology: Music, Mind, and Brain, 26*(2), 179–188. https://doi.org/10.1037/pmu0000142

Waddell, G., & Williamon, A. (2017). Eye of the beholder: Stage entrance behavior and facial expression affect continuous quality ratings in music performance. *Frontiers in Psychology, 8*, 513. https://doi.org/10.3389/fpsyg.2017.00513

Wald-Fuhrmann, M., Egermann, H., Czepiel, A., O'Neill, K., Weining, C., Meier, D., … Tröndle, M. (2021). Music listening in classical concerts: Theory, literature review, and research program. *Frontiers in Psychology, 12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.638783.

Wingenbach, T. S. H., Brosnan, M., Pfaltz, M. C., Peyk, P., & Ashwin, C. (2020). Perception of discrete emotions in others: Evidence for distinct facial mimicry patterns. *Scientific Reports, 10*(1), 4692. https://doi.org/10.1038/s41598-020-61563-5

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications (arXiv:1308.5499). *Online Tutorial. arXiv*. http://arxiv.org/abs/1308.5499.

Witvliet, C. V. O., & Vrana, S. R. (2007). Play it again Sam: Repeated exposure to emotionally evocative music polarises liking and smiling responses, and influences other affective reports, facial EMG, and heart rate. *Cognition & Emotion, 21*(1), 3–25. https://doi.org/10.1080/02699930601000672