Aesthetic and physiological effects of naturalistic multimodal music listening

Anna Czepiel*[1,2], Lauren K. Fink[1,3], Christoph Seibert[4], Mathias Scharinger[5,6], & Sonja A. Kotz[2,7]

[1]*Department of Music, Max Planck Institute for Empirical Aesthetics, Germany*
[2]*Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands*
[3]*Max Planck-NYU Center for Language, Music, and Emotion, Frankfurt am Main, Germany & New York*
[4]*Institute for Music Informatics and Musicology, University of Music Karlsruhe, Germany*
[5]*Research Group Phonetics, Department of German Linguistics, University of Marburg, Germany*
[6]*Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Germany*
[7]*Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Germany*

*Corresponding author
Email: anna.czepiel@ae.mpg.de
Anna Czepiel
Max Planck Institute for Empirical Aesthetics,
Grüneburgweg 14, 60322,
Frankfurt am Main,
Germany
https://orcid.org/0000-0002-7101-945X

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 (cardio-respiratory measures, skin conductance, and facial muscle activity) were continuously measured (Experiment 2). AE was significantly higher in the AV condition in both experiments. Physiological arousal indices – skin conductance and LF/HF ratio, which represent activation of the sympathetic nervous system – were higher in the AO condition, suggesting increased arousal, perhaps because sound onsets in the AO condition were less predictable. However, 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., skin conductance and 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 more 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.

Keywords: audiovisual, physiology, naturalistic, neuroaesthetics, motor mimicry

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 et al., 2012; Sloboda & O'Neill, 2001; Wald-Fuhrmann et al., 2021). Although the aesthetic experience (AE) of music is enhanced in a concert through several factors (see Wald-Fuhrman et al., 2021 for an overview), one which the current paper explores 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 et al., 2014), but their findings are inconsistent. 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 is currently of immense interest in cognitive neuroscience and the neuroscientific subdiscipline of neuroaesthetics. Here, processes such as perception, emotion, and appreciation are considered as influencing AE (for comprehensive reviews, see Anglada-Tort & Skov, 2020; Brattico & Pearce, 2013; Juslin, 2013; Pelowski et al., 2016; Schindler et al., 2017). Specific to the dynamic nature of music, Brattico and colleagues (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. More recent work further suggests that certain brain oscillations – particularly frontal theta – are related to music-evoked pleasure (Ara & Marco-Pallarés, 2020; Chabin et al., 2022; Sammler et al., 2007; Tervaniemi et al., 2021).

However, measuring brain activity in more naturalistic settings comes with significant challenges. A more accessible approach involves using peripheral physiological measures. 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 et al., 2021). Such peripheral measures are likewise mentioned in the AE chronometry approach (Brattico et al., 2013) as reflecting tension and chill responses (Grewe et al.,

2009; Salimpoor et al., 2009). However, unlike brain regions (fMRI) and the latency/polarity of (EEG/MEG) responses that can be more easily attributed to a mental mechanism, 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., 2021, for an example in pupillometry).

Our 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. SNS and PNS activation can be related to cardiac rhythms: the high frequency (HF, 0.4-0.15Hz) component is attributed to PNS activity, while the 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, with the facial muscles of zygomaticus major ('smiling') and corrugator supercilii ('frowning') commonly measured. Although under voluntary control, certain facial muscle responses may be partly unconscious (i.e., occur without attention or conscious awareness, Dimberg et al., 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 (Luft & Bhattacharya, 2015; Shaffer & Ginsberg, 2017), while zygomatic and corrugator muscle activity seem to reflect positive and negative valence, respectively (Bradley & Lang, 2000; Cacioppo et al., 2000; Dimberg, 1982; Lang et al., 1993; Larsen et al., 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

[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).

studies as follows. 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 et al., 2009). Physiological changes occur in response to cognitive music processes such as recognising unexpected harmonic chords (Koelsch et al., 2008; Steinbeis et al., 2006) and deviant stimuli (in an MMN-like paradigm, (Chuen et al., 2016 , though see Lyytinen et al., 1992), which might be enhanced by attention (Frith & Allen, 1983). In more naturalistic music listening, many studies showed that arousing music (faster tempos and unpredictable harmony) increase SC, HR, and RR (Bernardi et al., 2006; Coutinho & Cangelosi, 2011; Czepiel et al., 2021; Dillman Carpentier & Potter, 2007; Egermann et al., 2015; Khalfa et al., 2002; Krumhansl, 1997; though we note this is not also 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 et al., 2008). However, other work showed it can increase during unpleasant (dissonant) music (Dellacherie et al., 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 et al., 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 part 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, for example, 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, if the sound appears after 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 et al., 2022; Luck et al., 2010; Morrison & Selvey, 2014; Vines et al., 2011; Vuoskoski et al., 2014) perception of emotional intention (Dahl & Friberg, 2007; Vines et al., 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), such 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 et al., 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; Czepiel et al., 2021; Merrill et al., 2021; Scherer et al., 2019; Swarbrick et al., 2019; Tervaniemi et al., 2021), where participants report to experience stronger emotions (Gabrielsson & Wik, 2003; Lamont, 2011); though note that Belfi et al. (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 & 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 et al., 2019). It is also worth extending Coutinho & 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 & 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 varying 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 language studies focusing on the N100: an EEG event-related potential component that reflects early sensory processing, which is larger in amplitude in response to less predictable stimuli. The N100 component has been shown to be enhanced in AO (compared to AV) conditions in speech (Klucharev et al., 2003; van Wassenhove et al., 2005), emotional expression (Jessen & Kotz, 2013), 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, the evidence suggests that peripheral responses might be 1) higher during AO conditions if they reflect sensory processing, or 2) higher during AV conditions if they reflect the higher level 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 & Levitin (2008) and find increased physiology in AV conditions.

In summary, more research is required to assess modality effects in enhancing aesthetic experience in a more naturalistic setting. Further, the peripheral physiological correlates of aesthetic effects are mixed. The current study consists of two experiments that examine AE and physiology between AV and AO conditions in a concert hall setting. Both involve the same exact within-subjects experimental design: participants listening to piano performances of Bach, Beethoven, and Messiaen, in AO and AV conditions, in a concert hall. Modality order was counterbalanced across concerts. Both experiments record behavioural responses relating to AE, while the second experiment additionally included measurement of peripheral physiological signals. Thus, with Experiment 1 and 2, we tested the hypothesis that AE will be higher in the AV condition. In Experiment 2, we tested the hypothesis put forward by Vuoskoski et al. (2016) that peripheral physiological activity is 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 experiments used the same stimuli and similar procedure as described below.

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* (No. 10 from Vingt Regards sur L'Enfant-Jésus). These pieces were presented to the participants during the concert in two modalities: once as an audiovisual (AV) version that was played live by the pianist and an audio only (AO) version that was a live playback through loudspeakers. We consider this repetition of pieces as a naturalistic part of the design, since repetition of stimuli is a practice (although not extremely common) in concert programming (Halpern et al., 2017).

The different conditions were controlled as much as possible, with all stimuli performed by the same pianist playing the same instrument and in the same room. All performances were recorded both in audio and MIDI formats; audio recordings were presented to the participants via a stereo setup with high-quality full-range loudspeakers (Fohhn LX-150 + Fohhn XS-22). A trained sound engineer checked that loudness across modalities were equal. Since the AO and AV performances were different, we checked that the stimuli were comparable to eliminate confounding variables of potential acoustic differences between performances. We differentiated between score-based features and performance-based features (Goodchild et al., 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), while 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 et al., 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, with RMS (*mirrms*) and brightness (*mirbrightness*) representing loudness and timbre, respectively. In checking multicollinearity (Lange & Frieler, 2018), none of the features highly correlated, 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 combat for slight timing deviances between performances. All correlations between concerts were significant with r values > .4 (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 the extent to which the experimental components of the setting (e.g., measurement of the behavioural responses) disturbed the experience across the concert, where 'disturbed by measurement' was rated from 1 (strongly disagree) to 7 (very much agree). We further assessed familiarity of 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 our main dependent variable of interest: AE. Since an aesthetic experience 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 regarding 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, participants answered the short questionnaire described above. Modality order was counterbalanced across concerts. Concert 1 and 4 presented AO modality, while Concert 2 and 3 presented AV modality first. An overview can be seen in Figure 1.

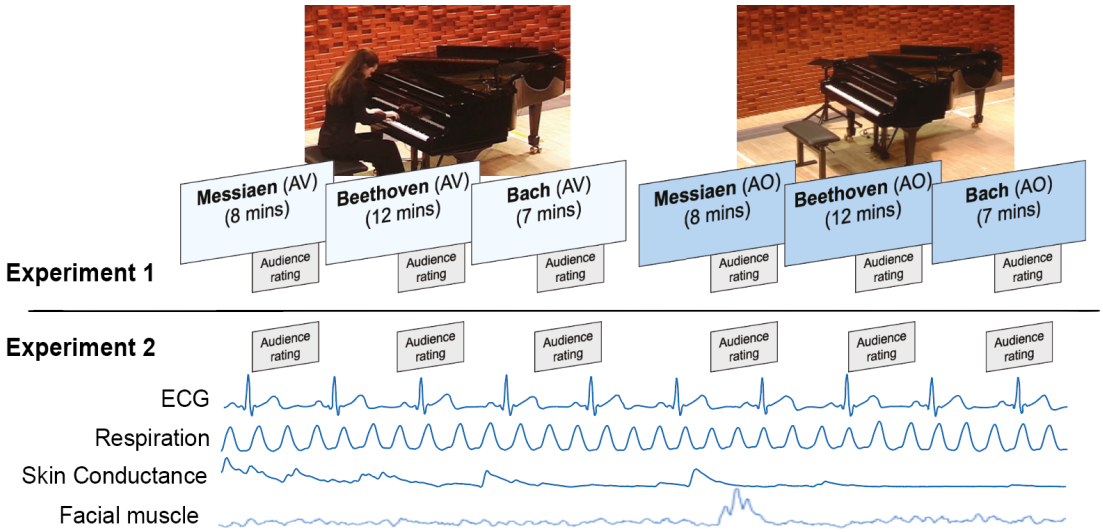

Figure 1. Outline of experimental procedure for Experiment 1 (behavioural audience ratings) and for Experiment 2 (audience ratings and peripheral physiological measures). Pieces were performed in an AO and AV version. Figure shows piece/modality order as presented for concerts 1 and 4. Modality order and piece order was counterbalanced in the concerts 2 and 3.

2.5 Analysis

Statistical analyses were conducted in R and R studio (R Core Team, 2021; RStudio Team, 2021). In both Experiment 1 and 2, the self-report items capturing the aesthetic experience (scaled within participants, using *scale* function in R) were highly correlated and thus reduced with a principal component analysis (PCA). Table 1 shows the loadings for both experiments. Details about each PCA are explained below in the experiment-specific methods.

Linear mixed models were run with the component extracted from the PCA as the dependent variable, with modality (AV / AO) as a fixed factor. As familiarity affects AE, a familiarity within pieces random intercept was added. We did not include a random participant intercept, as participant effects were already controlled for by scaling responses. Linear mixed models were run using *lmer* from the *lme4* packages (Bates et al., 2015; Kuznetsova et al., 2017) and significance values were obtained from the *anova* function and *summ* function from *jtools* package (Long, 2020). Pairwise comparisons were run with the *emmeans* function from *emmeans* package (Lenth, 2021) with Bonferroni corrections. Corresponding code and required to run these analyses are available in OSF (please note that while paper is in review, the repository is private and only available to reviewers, but will be made public once the manuscript has been accepted).

3 Experiment 1

3.1 Method

3.1.1 Participants

The study was approved by the Ethics Committee of the Max-Planck-Society and in accordance with the declarations of Helsinki. Participants gave their written informed consent. Twenty-seven participants attended the experimental concerts (13 and 14 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). For ratings whether the participant thought themselves as 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 there were no significant differences between participants 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 PCA analysis

A Kaiser-Meyer-Olkin (KMO) measure verified sampling adequacy (KMO = .79, well over the .5 minimum required) and all KMO values for individual items were > .61. Bartlett's test of sphericity was significant, revealing that correlations between items were large enough for a PCA, $X_2(28) = 471.323$, $p > .001$. Based on Kaiser's criterion of eigenvalues < 1, parallel analyses suggested the use of one component. Thus, a PCA was conducted with one component using the principal function (from *psych* package in R). The component explained 44% of the variance, with a good estimate of internal reliability (Cronbach's alpha = .81). Since liking, liking of interpretation, and absorption have been shown in literature to be critical components of aesthetic experience (Brattico & Pearce, 2013; Orlandi et al., 2020), we refer to this component as the overall 'aesthetic experience' (AE). One trial with a value exceeding ±3 Median Absolute Deviations (MAD, Leys et al., 2013) was removed from further analyses.

Table 1. PCA loadings from questionnaire items in both Experiment 1 and 2.

	Experiment 1	**Experiment 2**	
Item	**PC1**	**TC1**	**TC2**
Absorption	0.88	0.40	
Liking	0.86	0.81	
Liking of interpretation	0.83	0.84	
Passive reception	0.74		0.69
Connection to musicians	0.62		0.43
Connection to co-listeners	0.50		0.41
Understanding			
Urge to move		0.58	-0.43

Note: component loadings < .4 have been omitted. PC1 for both Experiment 1 and 2 is interpreted as 'Aesthetic experience'.

3.2 Results

3.2.1 Assessing naturalistic situations.

Results from the 162 responses (27 participants × 3 pieces × 2 modalities) of whether the measurements disturbed the concert are shown in Table 2. The mean rating was 1.54 (SD = 1.02), 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 overall did not disrupt the concert, confirming the ecological validity of our setting and results.

3.2.2 Piece familiarity.

Ratings for familiarity of style were similarly high for Bach (M = 5.80, SD = 1.28) and Beethoven (M = 5.63, SD = 1.25), but lower for Messiaen (M = 3.33, SD = 1.98). Most participants did not know the pieces specifically, though 26%, 18%, and 11% of participants knew the Beethoven, Bach, and Messiaen pieces, respectively.

3.2.3 Aesthetic experience: Modality differences.

Aesthetic experience was higher in the AV modality (see Figure 2), as confirmed by a significant main effect of modality $F(1,145) = 13.09$, $p < .001$, pseudo-R2 (fixed = 0.06, pseudo-R2 (total) = 0.38) with AV scores significantly higher (M = 0.167, SE = .161, 95% CI [-.165 0.500]) than AO scores (M = -0.283, SE = .161, 95% CI = [-0.618 0.051], $t(144) = -3.609$, $p < .001$).

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%

3.3 Discussion

This experiment tested whether participants had higher AE in audio-only (AO) or audiovisual (AV) piano performances in a naturalistic concert setting. Once confirming that the measurements overall did not disturb participants, our findings show that AE increased in the AV compared to AO condition. Thus, we support previous lab results that show liking and appreciation of expressivity are increased in AV conditions (Platz & Kopiez, 2012), extending these results to a more naturalistic

setting. One study which 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 part of the AE is also higher in AV modality. As naturalistic environments have less control, it is notable, and particularly important for future research, that we show such findings are replicated.

4 Experiment 2

Previous studies have aimed to gain further insight in the emotional differences between uni- and bimodal music performances by measuring physiological responses (Chapados & Levitin, 2008; Vuoskoski et al., 2016). However, such research has been inconsistent. In Experiment 2, we were interested in whether modality would affect peripheral physiological responses in the way it did for behavioural responses of AE (Exp. 1), and which peripheral signals might even serve as an index of AE.

4.1 Method

4.1.1 Participants

The study was approved by the Ethics Committee of the Max-Planck-Society and in accordance with the declarations of Helsinki. Participants gave their written informed consent. Twenty-six participants in total (Concert 3 N = 14, Concert 4 N = 12) attended one of two concerts: 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 14 concerts per year (M = 13.62, SD = 19.70). For ratings whether the participant thought themselves as a musician (from 1 = does not apply, to 7 completely applies), most participants selected 1 (N = 15) or 2 (N = 3), and less selected 3 (N = 0), 4 (N= 1), 5 (N = 4), 6 (N = 2) and 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 participant samples of the two Experiments, Experiment 1 had 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 = .333), concert attendance in the last 12 months (p = .755), and musician level (p = .575).

Self-report data from all 26 participants was used in the analysis, while one physiological dataset from Concert 3 was lost due to technical issues (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 supercilli ('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 PCA analysis

A Kaiser-Meyer-Olkin measure verified sampling adequacy (KMO = .61). All but one item had KMO values > .5; this one item ('connection with colisteners') had a value of close to .5 (0.48). Correlations between items were large enough for PCA (Bartlett's test of sphericity, $X_2(28) = 199.64$, $p > .001$. A PCA with oblimin rotation (i.e., oblique rotation, as we expected the components to be related) was conducted with two components, as suggested by a parallel analysis. The two components explained 45% of the variance (first component = 25%, second component = 20%). PC1 achieved an adequate internal reliability (Cronbach's alpha = .58), while PC2 had lower interval reliability, alpha = .39, thus only the first component was used. Since we have similar loadings to Experiment 1, we also refer to these components as the overall aesthetic experience (AE). In this component, one value exceeding ±3 Median Absolute Deviations (MAD, Leys et al., 2013) was removed from further analyses.

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 - 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 (3% of data were rejected), downsampled to 20Hz and separated into phasic (SCR) and tonic (SCL) components using Continuous Decomposition Analysis (Benedek & Kaernbach, 2010). Respiration, ECG, and EMG data were pre-processed using the Fieldtrip (Oostenveld et al., 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 – 0.4Hz) power as well as the LF/HF ratio were taken for 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 used the method of the paper that had the most similar methodology to ours and with which we wanted to compare our results: Vuoskoski et al. (2016). Therefore, the average of each (pre-processed) physiological measure was our main metric. Since 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 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 ten, 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.

4.1.5 Analysis

In addition to statistics for behavioural results, for each of the physiological dependent variables, linear mixed models were run with modality (AV / AO) as a fixed factor, and familiarity among pieces as a random intercept. Since physiology was not normalised (following Vuoskoski et al., 2016), an additional subject random intercept was added to the linear models. To determine if behavioural results were related to peripheral responses, we ran a linear model with aesthetic experience as the dependent variable and the eight peripheral measures as predictors (all of which were scaled in order to be included in the same model); modality, concert (i.e., the order modalities where presented), piece, and familiarity were included as random intercepts. All model predictor variance inflation factors were under 2.47, confirming their suitability for inclusion in the model.

4.2 Results

4.2.1 Assessing naturalistic situations.

We first assessed the extent that the behavioural/physiological measurements disturbed the overall experience during the concert (i.e., for all pieces / conditions). Out of 150 responses (26 participants x 3 pieces x 2 conditions), ratings suggested that measurements did not disrupt the concert experience, with a mean rating of 2.02 (SD = 1.42) and with 74% 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 without severely impacting ecological validity.

4.2.2 Piece familiarity.

Similar to Experiment 1, ratings for familiarity of style were high for Bach (M = 5.38, SD = 1.48) and Beethoven (M = 5.33, SD = 1.53), but lower for Messiaen (M = 4.13, SD = 1.88). Approximately a third knew the Beethoven and Bach pieces, whereas only 19% knew the Messiaen piece.

4.2.3 Aesthetic experience: Modality differences.

For behavioural results of AE, linear models showed a main effect of modality (F (1,133) = 9.255, p = .003, pseudo-R2 (fixed) = 0.05, pseudo-R2 (total) = 0.25) with AV scores significantly higher (M = 0.233 SE = 0.148, 95% CI [-0.070 0.537]) than AO (M = -0.202, SE = 0.149, 95% CI [-0.507 0.103], t(138) = -3.035, p = .003) (see Figure 2). These results replicate our findings from Experiment 1.

4.2.4 *Physiological differences between modality*

Statistical results are presented in Table 4 (see also Figure 3). There was a main effect of modality in SCL and LF/HF ratio, which both represent SNS activation (higher arousal). Comparisons of means indicated that these measures were higher in the AO compared to AV condition. There was also a significant main effect of modality for RR and corrugator muscle activity, with a significant increase in the AV compared to AO condition.

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 was the only significant predictor of self-reported AE (estimate beta = 0.16, SE = 0.05, $t(572) = 3.53$, $p < .001$, see Table 6 and Figure 4).

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%		

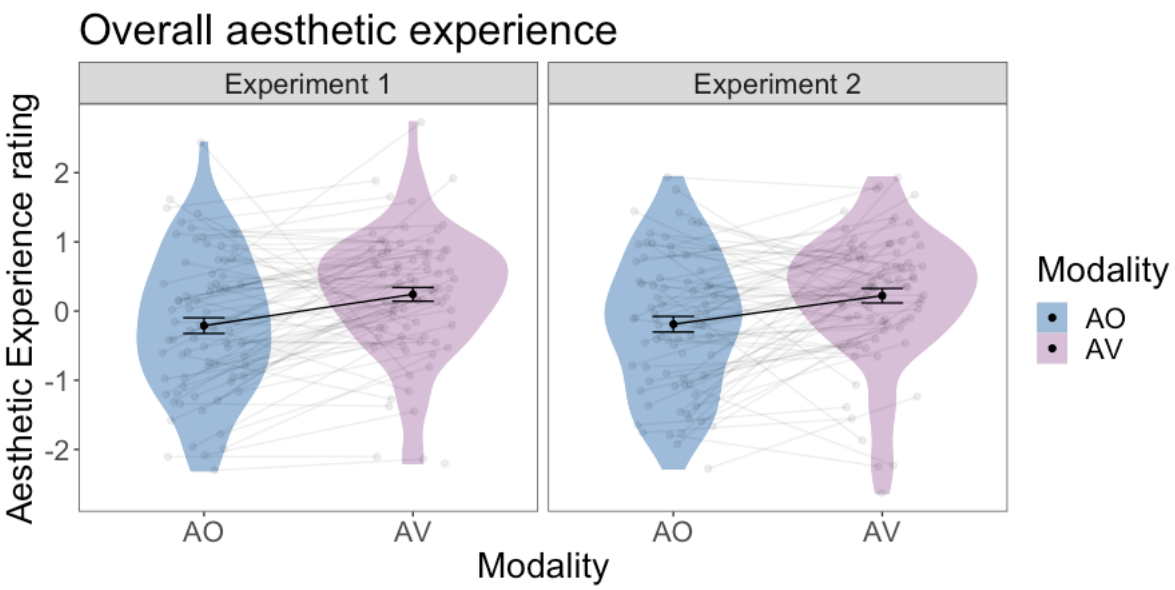

Figure 2. Aesthetic experience component scores (which had high item loadings of liking, liking interpretation and absorption, see Table 1) as a function of modality (Audio Only (AO) is blue and Audiovisual (AV) is purple). The left panel shows results for Experiment 1, while the right panel shows results for Experiment 2.

Table 4. Linear model results for main effect of Modality (AV / AO) on each physiological measure

Phys	**d.f.**	**F**	***p***	**R^2 fixed (total)**	**AO: M, (SE), [95% CI])**	**AV: M, (SE), [95% CI])**	***t*-test (AO-AV)**		
							df	***T***	***P***
LF/HF ratio	1, 1066	4.865	0.028	0.01 (0.42)	2.21 (0.126) [1.96, 2.46]	2.01 (0.127) [1.76, 2.26]	1067	2.202	0.023
RR	1, 1164	3.886	0.049	0.000 (0.80)	18.5 (0.291) [17.9 19.1]	18.7 (0.290) [18.1 19.3]	1126	-1.971	0.049
SCL	1, 1101	6.496	0.011	0.00 (0.81)	3.59 (0.197) [3.20, 3.98]	3.42 (0.197) [3.03, 3.81]	1078	2.548	0.011
EMG CS	1, 1066	10.01	.002	0.01 (0.66)	0.003 (0.000) [0.0023 0.0027]	0.003 (0.000) [0.0025 0.0029]	1066	-3.162	.002

Table 5. Model of physiology predicting AE self-reports

	β	S.E.	T	d.f.	*p*
SCL	-0.049	0.032	-1.710	568	0.134
SCR	0.003	0.024	0.548	555	0.897
HR	-0.037	0.026	-1.462	562	0.151
RR	-0.007	0.024	-0.307	556	0.7774
HF	0.042	0.033	1.434	549	0.209
LF/HF ratio	-0.018	0.032	-0.500	550	0.572
CS	0.030	0.024	1.135	558	0.202
ZM	0.065	0.022	2.849	553	0.003

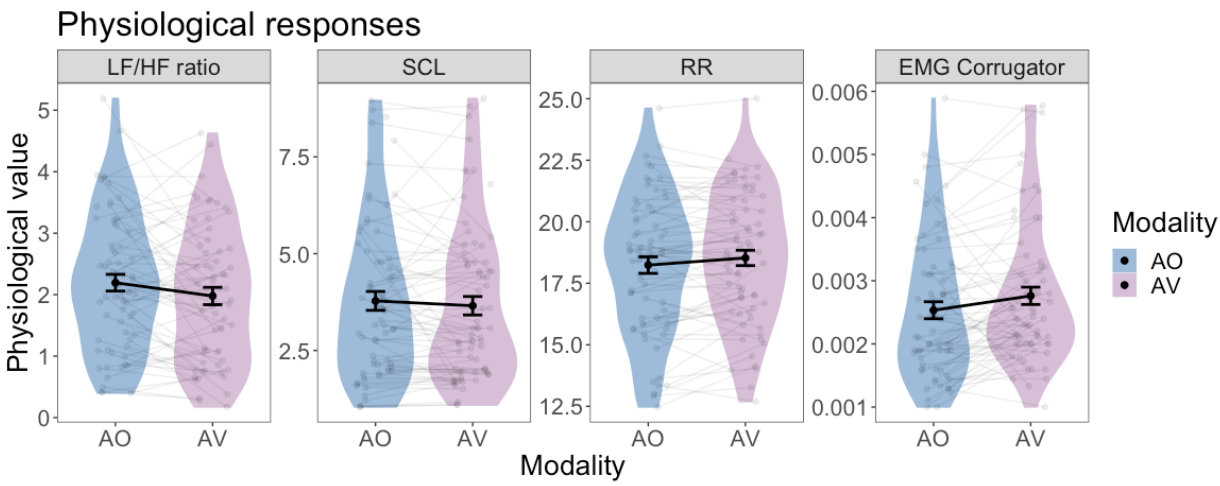

Figure 3. Physiological responses in each modality condition (AO: blue; AV: purple). Different panels represent different physiological measures; from left to right: LF/HF ratio, skin conductance level (SCL), respiration rate (RR), EMG activity of corrugator supercilii (EMG Corrugator).

4.3 Discussion

The main aims of Experiment 2 were to replicate the behavioural results from Experiment 1 and gain further insight into peripheral physiological measures as a function of modality. Importantly,

subjective ratings showed again that the measurement of physiological signals did not disturb participants.

AE was again significantly higher in the AV condition. We additionally tested whether peripheral responses are higher in AO compared to AV modalities. The current study showed that, compared to the AV condition, the AO condition evoked higher SCL and LF/HF ratio. These results support findings of (Vuoskoski et al., 2016), who also found higher physiological arousal in AO musical performances. On the other hand, respiration rate and corrugator activity were higher in the AV condition. As both of these measures have 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 has been shown to increase respiration (Brown et al., 2013; Mulder et al., 2005; Paccalin & Jeannerod, 2000) and corrugator activity (de Morree & Marcora, 2010).

In relating physiological responses to self-reported AE, zygomaticus activity was the only significant predictor. However, as zygomaticus activity increases have likewise been related to unpleasant experiences of (dissonant) music (Dellacherie et al., 2011; Merrill et al., 2021), we only cautiously attributed 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 confirm that in both experiments, the measurement of self-report and physiology did not disturb the audiences, supporting the idea that a semi-experimental context with naturalistic stimuli can yield results with high ecological validity.

As there are several components that can make up an AE (Brattico & Pearce, 2013; Juslin, 2013; Schindler et al., 2017), questionnaire items related to certain components of an aesthetic experience were used here. In both experiments, these items reduced to one reliable component. Although they had slightly different loadings between 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 et al., 2021). Similarly, dance choreography performed with more varied velocities was rated as more aesthetically pleasing compared to when it was performed with a more uniform velocity profile (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 et al., 2017). As these items have strong connections 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 principal component corroborates previous research that the aesthetic experience comprises many subcomponents (Brattico & Pearce, 2013; Merrill et al., 2021) and supports the use of dimensionality reduction techniques which trade specificity for the ability to more holistically measure AE.

Both Experiment 1 and 2 consistently showed that AE increases more in the AV than AO modality. 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). The current results extend this evidence 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 emotions; we extend their findings to the preference (liking, liking of the interpretation) part of the AE. We emphasise the importance of conducting AE research in a naturalistic performance setting, since it is more likely to elicit stronger and more realistic responses (Gabrielsson & Wik, 2003; Lamont, 2011). Of note is that results found in lab settings are not always replicated in more naturalistic settings. For example, previous lab studies have demonstrated that body movement (Platz & Kopiez, 2012) and familiarity (see North & Hargreaves, 2010) increase appreciation of music; 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 more realistic dependent variable of appreciation (i.e., money rather than ratings, Anglada-Tort et al., 2019). Thus, 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 is likely generalisable to more naturalistic human behaviour.

We further aimed to understand the peripheral responses of AE in multimodal contexts (Experiment 2), especially since research to date is inconsistent (Chapados & Levitin, 2008; Vuoskoski et al., 2016). Based on the framework from Brattico et al., (2013), we acknowledge AE as made up of perceptual, cognitive, affective, and aesthetic responses (e.g., liking). These components can be relatively well distinguished by self-report and – to an extent – different brain regions and neural ERPs, depending on their latency and polarity (e.g., early components should be related to early sensory processes). However, changes in physiology/facial muscle activity have been related to all of these components, depending on the design and control condition of the study. Some show physiological changes related to 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 greater insight into what kind of AE-related processing is taking 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, 2013), as shown by Vuoskoski et al. (2016). On the other hand, AV information conveys more emotion (Dahl & Friberg, 2007; Van Zijl & Luck, 2013); therefore, responses should perhaps 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 found that SCL and LF/HF ratio increased in the AO condition. As these two measures represent increased SNS activation, this suggests that AO conditions increase physiological arousal, likely reflecting the increased unpredictability of sound onsets when visual information is absent. This is in line with results from Vuoskoski et al. (2016), who found that AO evokes more SCL and SCR compared to AV musical performances. Although we do not replicate this effect for the phasic (SCR) component of the skin conductance, this is likely due to the overall length of the stimuli we used. The current experiment had longer (5-12 minute) musical pieces than Vuoskoski et al. (2016) (just over 4 minute). 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*).

However, we also found support for the alternative hypothesis, that AV has higher peripheral physiological responses. First, RR was faster in the AV condition. As respiration is a slower signal than HR and skin conductance, it might not be as sensitive to sensory stimuli. Indeed, the majority of studies showing physiology responses are influenced by low-level sensory stimuli (i.e., orienting responses/ startle reflex) have been measured with faster moving signals, such as blinking, skin conductance (SCR and SCL), heart rate (Barry & Sokolov, 1993). Second, corrugator activity (which typically reflects negative valence, Bradley & Lang, 2000) increased in the AV condition. Although AE (generally positive) was also higher in the AV condition, this discrepancy could be explained by the fact that high aesthetic pleasure could also come with the perception of negatively valenced musical expression and/or affective states (Eerola et al., 2018) for example via being moved (Eerola et al., 2016). Such a state might be reflected via facial expression. The enhanced activity in AV could also be because the presence of a performer creates a social situation that increases emotional communication via facial expression (Griffiths & Reay, 2018). However, there is growing evidence suggesting that corrugator and zygomaticus are not just related to valence (Wingenbach et al., 2020), as well as other work that questions whether facial expression reflects affect states at all (Lewis, 2011; Matsumo, 1987). Thus, another possible interpretation is that observing the musician increased mimicry in the observers. This idea is supported by past work that has shown that participants mimic the facial expression of observed faces (Dimberg, 1982; Magnee et al., 2007) in AV musical performances (Chan et al., 2013). Mimicry may also be an explanation for why there is faster breathing in the AV condition: viewing effortful

movements increases respiration (Brown et al., 2013; Mulder et al., 2005; Paccalin & Jeannerod, 2000) and corrugator activity (de Morree & Marcora, 2010). Indeed, sensorimotor, embodied mechanisms related to motor mimicry have been proposed and shown to be a mechanism for enhancing AE (Brattico & Pearce, 2013; Cross et al., 2011; Freedberg & Gallese, 2007; Gallese & Freedberg, 2007). Thus, faster breathing and increased facial muscle activity in AV may be a reflection of motor mimicry occurring when viewing musicians' movements.

The zygomaticus ('smiling') muscle activity was a predictor for AE ratings. In other words, increased zygomaticus muscle activity was positively related to AE, supporting previous work showing that zygomaticus activity was higher for liked dance movements (Kirsch et al., 2016). 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.

That SC and LF/HF ratio were higher in AO, but respiration and muscle activity were higher in AV, is relatively striking, once a more thorough biological make-up of these different peripheral measures is considered in conjunction with how much (in)voluntary control we have over them. As mentioned before, SC and heart are innervated exclusively by the ANS and made up of involuntary (cardiac) muscle, respectively. Voluntary skeletal muscles control EMG and (partly) respiration. On the one hand, due to the automatic nature of SC and 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 study is that although all versions were presented as part of a concert while participants were seated in the concert hall throughout, AV was presented as a live version, while AO was presented as a live 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 also occurs in Swarbrick et al. (2019) who state a similar motivation 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 our results.

Future research may consider live audio-only playback of an offstage performer, if they wish 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 in terms of length. Nonetheless, length was a compromise when using naturalistic stimuli that heightened ecological validity. As we do not look at piece-specific differences, but rather average across sections of the pieces to examine the effect of condition, we do not consider this a confound for the current study, though we do note that effects driven by one piece may weigh our results more heavily than effects from the shorter pieces. Future researchers 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 in order to keep the within-in subject design manageable in terms of time, 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 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 physiological arousal responses (skin conductance and heart rhythms reflecting SNS), were higher in the (less predictive) AO modality, likely reflecting more sensory processes. On the other hand, peripheral responses with more voluntary control (respiration, facial muscles) were higher in the AV modality, with the zygomaticus muscle even serving as a significant predictor of self-reported AE. Thus, it could be that the involuntary-voluntary continuum of physiological responses is related to the 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). Thus, by exploring modality effects, we postulate that peripheral responses likely reflect sensory, sensorimotor, and affective responses which may culminate to an overall aesthetic experience (Brattico et al., 2013). However, we stress 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 reflecting intertwined

sensory and affective aspects of the aesthetic part of the experience remain unclear. There is also a gap in research as to which musical features may affect physiological but not explicit behavioural measures. 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.

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.

Declaration of Interest

None.

Author contribution: Using CRediT

Conceptualisation: CS
Methodology: CS
Investigation: CS, MS
Formal analysis: AC, SAK, MS, LKF
Writing - Original draft: AC
Visualisation: AC, LKF
Writing - review and editing: LKF, SAK, MS, CS

References

Anglada-Tort, M., & Skov, M. (2020). 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*.

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*, 116665. https://doi.org/10.1016/j.neuroimage.2020.116665

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

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., 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

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

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. *Proceedings of the 18th ACM International Conference on Multimedia*, 1467–1468.

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

Chan, L. P., Livingstone, S. R., & Russo, F. A. (2013). Facial Mimicry in Response to Song. *Music Perception*, *30*(4), 361–367. https://doi.org/10.1525/mp.2013.30.4.361

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*, 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

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. S., Kirsch, L., Ticini, L., & Schütz-Bosbach, S. (2011). The Impact of Aesthetic Evaluation and Physical Ability on Dance Perception. *Frontiers in Human Neuroscience*, *5*. https://www.frontiersin.org/article/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

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. (1982). Facial Reactions to Facial Expressions. *Psychophysiology*, *19*(6), 643–647. https://doi.org/10.1111/j.1469-8986.1982.tb02516.x

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

Eerola, T., Vuoskoski, J. K., & Kautiainen, H. (2016). Being Moved by Unfamiliar Sad Music Is Associated with High Empathy. *Frontiers in Psychology*, *7*, 1176. https://doi.org/10.3389/fpsyg.2016.01176

Eerola, T., Vuoskoski, J. K., Peltola, H.-R., Putkinen, V., & Schäfer, K. (2018). An integrative review of the enjoyment of sadness associated with music. *Physics of Life Reviews*, *25*, 100–121. https://doi.org/10.1016/j.plrev.2017.11.016

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

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

Fink, L., simolaJaana, J., Tavano, A., Lange, E. B., Wallot, S., & Laeng, B. (2021). From preprocessing to advanced dynamic modeling of pupil data. *PsyArXiv*.

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

Gabrielsson, A., & Wik, S. L. (2003). Strong Experiences Related to Music: Adescriptive 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

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*. MIT Press.

Jessen, S., & Kotz, S. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Frontiers in Human Neuroscience*, *7*, 369. https://doi.org/10.3389/fnhum.2013.00369

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

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

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

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*, *237*.

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Lewis, M. (2011). Inside and Outside: The Relation between Emotional States and Expressions. *Emotion Review*, *3*(2), 189–196. https://doi.org/10.1177/1754073910387947

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

Long, J. A. (2020). *jtools: Analysis and Presentation of Social Scientific Data* (R package version 2.1.0) [Computer software]. https://cran.r-project.org/package=jtools

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

Luft, C. D. B., & 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

Lundqvist, L.-O., Carlsson, F., Hilmersson, P., & Juslin, P. N. (2008). Emotional responses to music: Experience, expression, and physiology: *Psychology of Music*. 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

Magnee, M. J. C. M., Stekelenburg, J. J., Kemner, C., & de Gelder, B. (2007). Similar facial electromyographic responses to faces, voices, and body expressions: *NeuroReport*, *18*(4), 369–372. https://doi.org/10.1097/WNR.0b013e32801776e6

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

MATLAB. (2019b). *Version 9.7 (R2019b)*. The MathWorks Inc.

Matsumo, D. (1987). The Role of Facial Response in the Experience of Emotion: More Methodological Problems and a Meta-Analysis. *Journal of Personality and Social Psychology*, *52*(4).

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*.

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

North, A. C., & Hargreaves, D. J. (2010). Music and marketing. In *Handbook of music and emotion: Theory, research, applications* (pp. 909–930). Oxford University Press.

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*, 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

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://doi.org/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). Sinauer Associates.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Richardson, D. C., Griffin, N. K., Zaki, L., Stephenson, A., Yan, J., Curry, T., Noble, R., Hogan, J., Skipper, J. I., & 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/

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), 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). Assessing emotional experiences of opera spectators in situ. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(3), 244–258. https://doi.org/10.1037/aca0000163

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), 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., Lamont, A., & Greasley, A. (2012). *Choosing to hear music* (S. Hallam, I. Cross, & M. Thaut, Eds.; Vol. 1). 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*. 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., 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

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

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. http://dx.doi.org/10.1037/pmu0000142

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

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., Tschacher, W., Uhde, F., Toelle, J., & 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